

---

# AdaPKC: PeakConv with Adaptive Peak Receptive Field for Radar Semantic Segmentation

---

Teng Li<sup>2\*</sup> Liwen Zhang<sup>1\*†</sup> Youcheng Zhang<sup>1</sup> Zijun Hu<sup>1</sup>  
Pengcheng Pi<sup>1</sup> Zongqing Lu<sup>2</sup> Qingmin Liao<sup>2†</sup> Zhe Ma<sup>1</sup>

<sup>1</sup>Intelligent Science and Technology Academy of CASIC

<sup>2</sup>Shenzhen International Graduate School, Tsinghua University

liteng21@mails.tsinghua.edu.cn\* lwzhang9161@126.com\*†

liaoqm@tsinghua.edu.cn†

## Abstract

Deep learning-based radar detection technology is receiving increasing attention in areas such as autonomous driving, UAV surveillance, and marine monitoring. Among recent efforts, PeakConv (PKC) provides a solution that can retain the peak response characteristics of radar signals and play the characteristics of deep convolution, thereby improving the effect of radar semantic segmentation (RSS). However, due to the use of a pre-set fixed peak receptive field sampling rule, PKC still has limitations in dealing with problems such as inconsistency of target frequency domain response broadening, non-homogeneous and time-varying characteristic of noise/clutter distribution. Therefore, this paper proposes an idea of **adaptive** peak receptive field, and upgrades **PKC** to **AdaPKC** based on this idea. Beyond that, a novel fine-tuning technology to further boost the performance of AdaPKC-based RSS networks is presented. Through experimental verification using various real-measured radar data (including publicly available low-cost millimeter-wave radar dataset for autonomous driving and self-collected Ku-band surveillance radar dataset), we found that the performance of AdaPKC-based models surpasses other SoTA methods in RSS tasks. The code is available at <https://github.com/lihua199710/AdaPKC>.

## 1 Introduction

As a common remote sensing device, radar exhibits superior robustness in complex environments (*e.g.*, varying weather and lighting conditions) compared to cameras, and it is more cost-effective and resilient in extreme weather scenarios compared to LiDARs. Benefiting from the physical advantages of radar sensors and the powerful capabilities of deep learning techniques, modern deep learning-based radar signal interpretation has become a hot research topic in the field of radio frequency detection technology. It has been extensively explored in autonomous driving [30, 19, 34, 6], UAV surveillance [9, 17], sea monitoring [27, 21, 28], etc. Considering the similar dense representations between radar frequency maps and optical images, most of these works directly transfer convolution networks or modules developed for optical signals to radar perception tasks, such as radar object detection (ROD) and radar semantic segmentation (RSS), and they have achieved impressive performance. Nevertheless, without specific design for the inherent characteristics of radar signals, these approaches fail to fully liberate the potential of deep learning techniques.

Recently, PKCIn-Net [32] introduced an innovative convolution operator named PeakConv (PKC), tailored for the efficient analysis of radar signals, and this operator seamlessly integrates the advan-

---

\*Equal contribution. †Corresponding author. This research is supported by Young Science Foundation of National Natural Science Foundation of China (No.62206258).

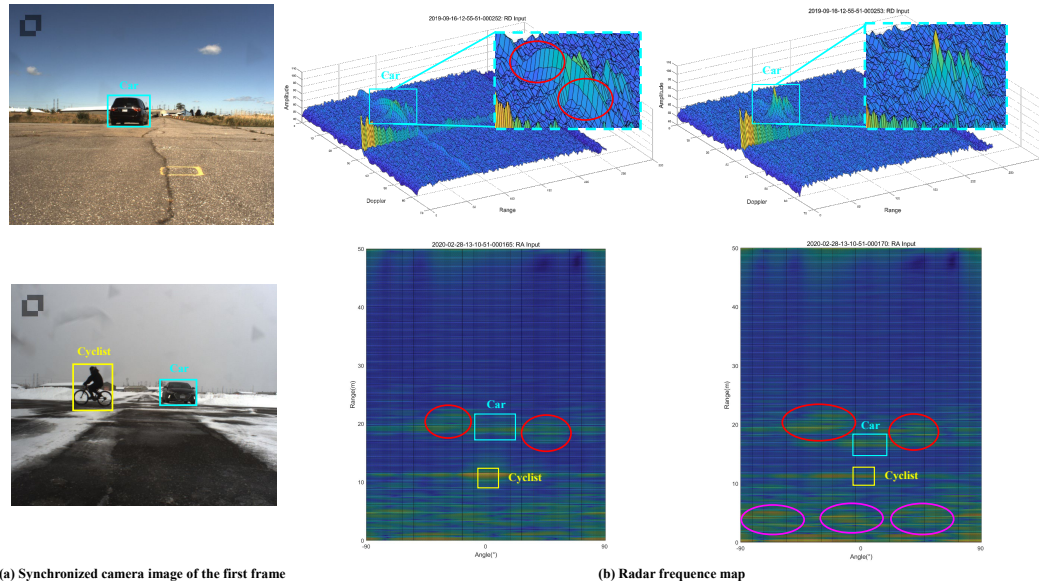


Figure 1: The illustration of variations in target signature and interfering signals in radar frequency map. The first row illustrates the variations of the same target across temporally consecutive frames in the range-Doppler (RD)-amplitude 3D representation. The second row demonstrates the disparities of different targets in the same frame, as well as the same target across different frames, in the range-angle (RA) 2D representation. Cyan and yellow rectangles represent target areas, illustrating the variations of target signature with dimensions, categories, and time, etc. Red ellipses indicate prominent interfering clutter, while purple ellipses represent clutter undergoing significant changes.

tages of classic radar detectors [22] with common convolution networks [19, 6]. For radar signals, the frequency responses of objects comprise target echoes and interference, and share a distinct peak-shaped pattern, thus most classic radar detection methods [22, 10, 25, 23] build peak detection algorithms upon constant false alarm rate (CFAR) criteria. Extending from cell averaging-CFAR (CA-CFAR) [22], PKC explicitly embeds a similar band-pass peak enhancement mechanism in a standard convolution operator for better characterising target signatures in radar signals. Concretely, following the guard-reference policy of CA-CFAR, it first estimates interfering signals with the center unit/cell and reference units outside predefined guard bands. Then, with estimated interference, it finishes noise suppression for each cell under test (CUT) in feature space and enhances peak frequency response associated with objects of interest.

Despite its superior suitability for radar data than alternative convolution operators [31, 5, 33], there exists even greater potential for PKC to learn peak frequency response of radar signals. Research on CFAR detectors [10, 25, 23, 11] reveals that there exist significant variations in target signature and associated interference within radar signals, rendering the predefined reference cells in CA-CFAR inadequate for precisely locating interfering signals, and this limitation can also be observed in PKC. To provide a clearer depiction, let us delve deeper into these variations present within radar signals, as illustrated in Fig. 1. On the **target side**, since multi-dimensional radar tensors are generated through a sequence of cascading fast Fourier transformations (FFTs), target signatures along different dimensions exhibit distinct degrees of frequency response tailing (broadening). Additionally, the broadening degrees of different instances would also be influenced by target categories or states, *i.e.*, the relative distances, azimuth or velocity from the radar. On the **interference side**, the noise or clutter distribution commonly exhibits non-homogeneous and time-varying characteristics. However, since the PKC kernel always gathers the reference units at fixed locations for noise estimation, *i.e.*, the predefined peak receptive field (PRF), the dynamic variations in both targets and interference degrade its performance. In short, **the fixed PRF essentially limits the learning ability of PKC**, thus, it hinders the RSS model from obtaining better performance.

Motivated by the adaptive selection of reference cells in classic CFAR detectors [10, 25, 23], in this work we introduce two novel data-adaptive band-pass filtering mechanisms aimed at **upgrading PKC to adaptively adjust its PRF for each CUT** in a data-driven manner, namely **adaptive PeakConv (AdaPKC)**. Concretely, both versions of AdaPKC first measure the correlation between CUT and its

alternative reference units in high-dimensional feature representations, then select proper reference units and integrate them seamlessly with PKC to effectively take care of the fluctuating dynamics of radar signals. The main contributions of our work are:

- We present the first attempt specially tailored for radar signal processing to dynamically adjust the receptive field for convolution operators. Concretely, we propose a novel updated version of PKC, termed as AdaPKC, which can adaptively adjust the PRF (or reference units) at cell-level granularity. And two different implementation versions are provided, which both exhibit enhanced flexibility and robustness in handling fluctuating radar signals compared to original PKC.
- To better release the learning ability of AdaPKC, a fine-tuning technology with a thresholding on-line switch is presented. With such technology, the same AdaPKC-based model can even achieve better performance with less computational cost.
- To verify the effectiveness of AdaPKCs, quantitative and qualitative experiments are conducted on various real-measured large scale radar datasets including CARRADA [20] collected from a low-cost FMCW ( $\approx 77\text{GHz}$ ) radar in autonomous driving scenario and self-collected dataset recorded from a Kurz-under (Ku) band ( $\approx 17\text{GHz}$ ) radar for UAV surveillance and sea monitoring. Results show that AdaPKC-based models achieve SoTA RSS performance and our fine-tuning strategy further brings visible improvements, verifying the scope of application of AdaPKCs.

## 2 Related Work

**Receptive field (RF) adjustment.** RF is crucial for modern deep convolution models, affecting the granularity of modeling primitives, computation architecture and their representation capabilities, etc. The rational use of RF can directly improve the representation ability of the models, *e.g.* enlarging the scope [31], multi-scale modeling [2], and dynamically changing shapes [5, 33]. Beyond that, the concept of RF has also been applied to Transformers [7, 16]. These methods are proposed for vision tasks and have achieved significant results. However, compared with conventional convolution, the improvement is not satisfactory enough on radar signals. Recently, by fully considering the characteristics of radar signals, the concept of PRF has been proposed, which makes the original convolution have the ability of band-pass filtering and noise suppression [32]. However, the bandwidth of filtering is pre-set, which hinders the adaptive ability of PKC to radar data. To this end, this paper attempts to study a data-driven PRF adjustment method, and introduces the concept of adaptive PRF (AdaPRF), so as to further improve the RSS performance. Considering that the adaptive adjustment of the suppression (guard) bandwidth and the non-suppression (reference) units is essentially a dynamic adjustment of RF, and from this point of view, the content studied in this paper is related to deformable convolution (DefConv) [5, 33]. Unfortunately, the dynamic RF technology of DefConvs cannot solve the problem in hand, for the following reasons: i) *Differences of signaling mechanism.* DefConv uses the visual prior information that the target is visually deformed geometrically, which cannot be directly corresponded to the radar signal. ii) *Mismatched prediction method.* DefConv generates new RF through prior prediction, *i.e.*, the regular RF is still used to infer the sampling point outside regular RF. However, in radar signal, the interference (noise/clutter) with large entropy, is often difficult or even impossible to predict. At present, a better way is under the premise of observation, *i.e.*, posterior measurement or statistics. iii) *Different mechanisms of representation.* Noise suppression is not required to be considered by DefConv, thus it does not need to distinguish between the center unit and surroundings during calculation. To this end, a novel adaptive RF adjustment method is required.

**Radar semantic segmentation.** Benefiting from the reliable perceptual capabilities, convolutional neural networks (CNNs) play an indispensable role in existing RSS networks for radar frequency maps processing, whether in pure CNN models [13, 8, 19, 32] or transformer-assisted CNN models [34, 12, 6]. RSS-Net [13] utilizes a fully convolutional neural network with encoder-decoder structure to recognize targets in radar scans, and it incorporates an atrous spatial pyramid pooling (ASPP) [2] module to gather multi-scale spatial information. RAMP-CNN [8] employs parallel branches to extract features from multiple views and adopts 3D convolutions to better capture temporal information. TMVA-Net [19] leverages these techniques to develop a multi-view RSS model, which is capable of making semantic predictions across multiple views simultaneously. T-RODNet [12] integrates Swin Transformer [16] modules into a CNN-based RSS model to strengthen its modeling capability. TransRSS [34] and TransRadar [6] introduce attention blocks into the multi-view feature fusion stage to enhance the fusion effectiveness. Recently, as the first fundamental convolution operator tailored for radar signal processing, PKC [32] is proposed. Compared to Dilated Convolution [31]

and Deformable ones [5, 33], PKC demonstrates superior RSS performance. However, it is inherently constrained by its fixed peak receptive field, posing challenges in achieving consistent interference (noise/clutter) suppression under the dynamic and time-varying nature of radar signals. By contrast, our AdaPKCs overcome this limitation well by implementing novel data-adaptive band-pass filtering mechanisms.

### 3 Method

In this section, the proposed two versions of AdaPKC with different AdaPRF mechanisms are first introduced in § 3.1. Then, the proposed fine-tuning strategy to further uncover the potential of AdaPKC-based RSS models is presented in § 3.2. We design these models using both multi-view [19, 32] and single-view frameworks, which are comprehensively described in Appendix A.2 and A.3 for saving space.

#### 3.1 AdaPKC

##### 3.1.1 AdaPKC<sup>ξ</sup>: PKC w/ Metric-based AdaPRF

To ensure the reliability of estimating the proper unit-level PRF, *i.e.*, AdaPRF, for AdaPKC in radar signals, we establish the estimation process in a posterior way: we first define a set of candidate PRFs within the neighbourhood of the center unit, aligning with the local peak response of targets and the local scanning process of convolution, and then design a measuring criterion to evaluate these PRFs, estimating AdaPRF primarily occupied by interfering signals. Motivated by classic CFAR detectors [10, 25, 23], we first demonstrate how to estimate AdaPRF in an explicitly measuring way, referred to as metric-based AdaPRF (AdaPKC<sup>ξ</sup>). To illustrate the mechanism of AdaPKC<sup>ξ</sup>, we begin by defining the search space of alternative PRFs. Reviewing the PRF definition in previous work [32] we can see that, the PRF for center unit  $\mathbf{x}_c$  encompasses  $\mathbf{x}_c$  itself and a set of sampled reference units  $\{\mathbf{x}_r^{(i)}\}_{i=1}^{N_r}$ , and the area of reference units is governed by horizontal- and vertical-symmetry guard bandwidth  $\mathbf{b}^G \triangleq \{b_x^G, b_y^G\}$  and reference bandwidth  $\mathbf{b}^R \triangleq \{b_x^R, b_y^R\}$ , as illustrated in Fig. 2-(a). Following this definition, the PRF adjustment corresponds precisely to the adjustment of the reference unit set, thus we can define the PRF search space by defining the candidate sets of reference units with the adjustment ranges for the guard bandwidth and reference bandwidth. Given that adjusting the reference bandwidth leads to a drastic change in the number of sampled reference units compared to adjusting the guard bandwidth, in AdaPKC<sup>ξ</sup> we keep anytime-fixed  $\mathbf{b}^R \triangleq \{b_x^R = 1, b_y^R = 1\}$  and denote the set of  $K$  guard bandwidth candidates as  $\Omega^G \triangleq \{\mathbf{b}_k^G\}_{k=1}^K = \{\mathbf{b}^G \mid b_{\min|x}^G \leq b_x^G \leq b_{\max|x}^G, b_{\min|y}^G \leq b_y^G \leq b_{\max|y}^G\}$ , generating  $K$  sets of reference units as  $\{\{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^{N_k}\}_{k=1}^K$  correspondingly. As a result, AdaPRF estimation in AdaPKC<sup>ξ</sup> is equivalent to selecting an appropriate reference unit set from these candidate sets for each CUT (or center unit), as illustrated in Fig. 2-(b).

For better explanation, we divide the observed radar signals into three subsets: i) signals reflected directly from a target,  $\mathcal{S}_t$ ; ii) the target-interfering noise, *i.e.*, the noise coupled with the signal that partially leaks out of the target,  $\mathcal{S}_{t-n}$ ; and iii) the target-independent noise,  $\mathcal{S}_n$ . In practice, it is the part that from  $\mathcal{S}_{t-n}$  really causes misjudgment. Therefore, classic CFAR detectors focus on filtering out such noise either by the extreme value [10, 25] or the median value [23] in amplitude domain of signals. Motivated by this idea, our AdaPKC<sup>ξ</sup> centers its attention on collecting reference units predominantly occupied by such noise.

However, as a learnable module, AdaPKC<sup>ξ</sup> needs to process the representation tensors of radar signals, implying that the measurement of the reference units should be conducted on feature space. For some CUT  $\mathbf{x}_c = \psi(s; \mathbf{W})$  and its candidate reference unit  $\mathbf{x}_r = \psi(s'; \mathbf{W})$ , where  $s \in \mathcal{S}_t$ ,  $s' \in \mathcal{S}_t \cup \mathcal{S}_{t-n} \cup \mathcal{S}_n$ , and  $\psi(\cdot; \mathbf{W}) \in \mathbb{R}^C$  denotes a convolution layer with shared weights,  $\mathbf{W}$ . Then, AdaPKC<sup>ξ</sup> should be responsible for transforming these features into a metric space that explicitly delineates their correlation with the target. This transformation is achieved by utilizing the inner product of representations between CUT and its candidate reference unit, *i.e.*,  $\mathbf{x}_c \mathbf{x}_r^T$ , sharing a similar spirit with the matched filter concept in conventional radar signal processing [22] and attention in [26].

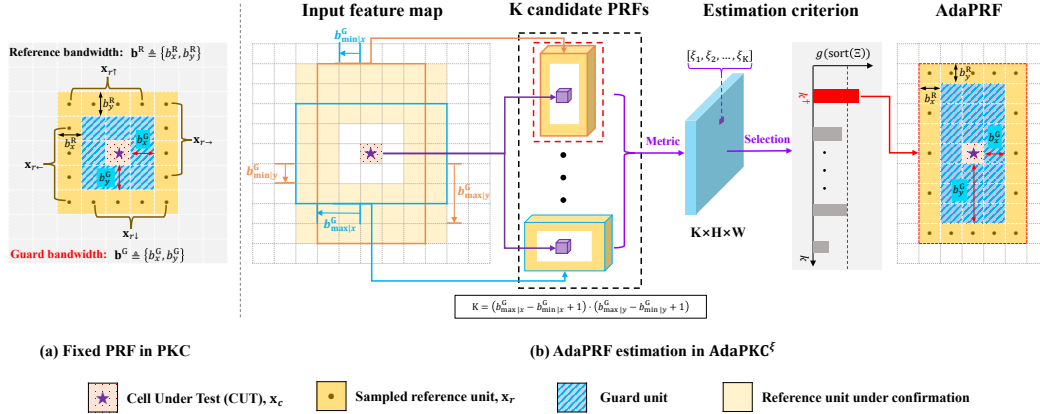


Figure 2: The illustration of AdaPRF in AdaPKC<sup>ξ</sup>. (a) illustrates the definition of PRF in PKC, whose area is governed by the reference bandwidth  $\mathbf{b}^R$  and guard bandwidth  $\mathbf{b}^G$ ; (b) describes the estimation process of AdaPRF in AdaPKC<sup>ξ</sup>, including denoting  $K$  candidate PRFs for each CUT, translating these PRFs into metric scores  $\{\xi_k\}_{k=1}^K$ , and finally selecting an appropriate PRF as the AdaPRF with these metric scores.

Under this definition, these measures exhibit the following statistical properties,

$$\mathbb{E}(\mathbf{x}_c \mathbf{x}_r^\top) = \begin{cases} \mathbb{E}(\|\psi(s; \mathbf{W})\|_2^2), & \text{if } s' \in \mathcal{S}_t \\ \mathbb{E}(\|\psi(s'; \mathbf{W})\|_2^2), & \text{if } s' \in \mathcal{S}_{t-n} \\ 0, & \text{if } s' \in \mathcal{S}_n \end{cases}, \quad (1)$$

where,  $\mathbb{E}(\cdot)$  is the expectation and  $\|\cdot\|_2$  denotes the  $L_2$  norm. From Eq. (1), we can see that the inner product transformation assigns three statistical boundaries to  $\mathbf{x}_r$  from  $\mathcal{S}_t$ ,  $\mathcal{S}_{t-n}$  and  $\mathcal{S}_n$ : for  $s' \in \mathcal{S}_{t-n}$ , the expectation  $\mathbb{E}(\mathbf{x}_c \mathbf{x}_r^\top)$  consistently exhibits smaller value than case for  $s' \in \mathcal{S}_t$  and larger value than  $s' \in \mathcal{S}_n$ , with a notable separation between their respective magnitudes. This attribute significantly serves to facilitate the subsequent localization of reference units from target-interfering noise.

Then we elucidate the process of translating the  $K$  available sets of reference units (or PRFs) into the previously discussed metric space. Since different sets may comprise varying numbers of units, we uniformly sample  $N$  (16 by default) units as representatives, as illustrated in Fig. 2-(b). For the center unit  $\mathbf{x}_c$  and its  $k^{th}$  reference unit set  $\{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^N$ , let  $\mathcal{R}_k = \{\mathbf{x}_c, \{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^N\}$  denote its corresponding PRF. Then the correlation value (or metric score),  $\xi_k$  for the  $k^{th}$  PRF w.r.t.  $\mathbf{x}_c$  is formulated as

$$\xi_k = \frac{1}{N} \sum_{i=1}^N \sigma(\mathbf{x}_c \mathbf{x}_{r|k}^{(i)\top} / C), \quad (2)$$

where,  $C$  is the feature dimension;  $\sigma(\cdot)$  is the sigmoid function which normalizes  $\xi_k$  to (0, 1).

With the correlation values  $\Xi = \{\xi_k\}_{k=1}^K$  for all alternative PRFs, we can select the appropriate PRF  $\mathcal{R}^\dagger$ , which effectively encompasses target-interfering noise. In view of the attribute presented in Eq. 1, we employ the maximum value of the first-order gradient of  $\Xi$  as the selection criterion and please refer to Appendix B.1 for the detailed analysis. Then, we have final selection strategy as follows,

$$\mathcal{R}^\dagger \triangleq \mathcal{R}_{k^\dagger} \xleftarrow{k=k^\dagger} \{\mathbf{x}_c\} \cup \{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^N, \text{ s.t., } k^\dagger = \arg \max_k \{g[\text{sort}(\Xi)]\}, \quad (3)$$

where  $\arg \max$  operator retrieves the index corresponding to the maximum value in  $g[\text{sort}(\Xi)]$ ;  $g$  is the difference function;  $\text{sort}$  is the descending sort operator. After obtaining  $\mathcal{R}^\dagger$ , AdaPKC performs a convolution operation similar to PeakConv, detailed in Appendix A.1.

### 3.1.2 AdaPKC<sup>θ</sup>: PKC w/ Learning-based AdaPRF

In AdaPKC<sup>ξ</sup>, a measuring criterion is established in Eq. 3 based on prior knowledge of radar signals, providing a non-parametric way to achieve AdaPRF. Differing from AdaPKC<sup>ξ</sup>, AdaPKC<sup>θ</sup> learns to

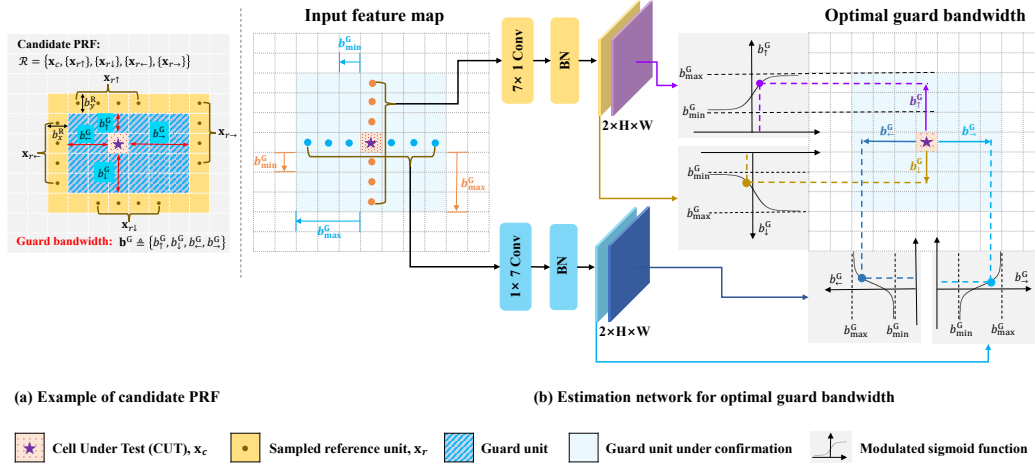


Figure 3: The illustration of AdaPRF in AdaPKC<sup>θ</sup>. (a) illustrates an example of candidate PRFs in AdaPKC<sup>θ</sup>, where the guard bandwidth  $\mathbf{b}^G$  is in a quadruple form; (b) describes the flowchart of the optimal guard bandwidth estimation network, which consists of two parallel branches that sample representative points in their corresponding directions and then automatically measure and select the optimal guard bandwidth.

build the criterion in a task-driven manner, employing a small network to estimate AdaPRF, *i.e.*, a parametric way. Thus, given center unit  $\mathbf{x}_c$  and its  $K$  candidate PRFs,  $\{\mathcal{R}_k\}_{k=1}^K$ , the natural way to locate AdaPRF  $\mathcal{R}^\dagger$ , is to define a function  $f(\cdot; \theta)$  with learnable parameters  $\theta$ , which is used to produce the likelihood of each  $\mathcal{R}_k$  being  $\mathcal{R}^\dagger$ , then we can have

$$\mathcal{R}^\dagger \triangleq \mathcal{R}_{k^\dagger}, \text{ where } k^\dagger = \arg \max_k \{f(\mathcal{R}_k; \theta)\}_{k=1}^K. \quad (4)$$

At a rough glance, there is no obvious difference between Eq. 4 and Eq. 3. However, AdaPKC<sup>θ</sup> involves joint optimization of estimation network parameters (*i.e.*,  $\theta$ ) and segmentation model parameters. The use of  $\arg \max$  operation will lead to the loss of gradient information of  $\theta$ , as a result, the optimization of the estimation network cannot be driven by the segmentation task. To this end, we transform the discrete estimation problem into a continuous form. Firstly, following the setup of anytime-fixed  $\mathbf{b}^R \triangleq \{b_x^R = 1, b_y^R = 1\}$  in AdaPKC<sup>ε</sup>, estimating  $\mathcal{R}^\dagger$  can be equivalently translated into estimating the optimal guard bandwidth  $\mathbf{b}^{G^\dagger}$ . Subsequently, to ensure that the estimation of  $\mathbf{b}^{G^\dagger}$  retains gradient information, the estimation network is designed to generate the continuous-valued  $\mathbf{b}^{G^\dagger}$ , instead of the likelihoods for alternative guard bandwidths. Finally, the derivable linear interpolation is used to associate the continuous-valued  $\mathbf{b}^{G^\dagger}$  with discrete spatial coordinates.

Concretely, given an input feature map  $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ , our AdaPKC<sup>θ</sup> is responsible for obtaining an appropriate guard bandwidth  $\mathbf{B}^{G^\dagger} = \{\mathbf{b}_{h,w}^{G^\dagger}\}_{h=1,w=1}^{H,W} \in \mathbb{R}^{4 \times H \times W}$ . For enhancing the expressive capability of AdaPKC, the horizontal- and vertical-symmetry  $\mathbf{b}^G$  design in the original PRF is extended to a quadruple form,  $\mathbf{b}^G = \{b_{\uparrow}^G, b_{\downarrow}^G, b_{\leftarrow}^G, b_{\rightarrow}^G\}$ , so that the shape of PRF would enjoy free change in four directions, *i.e.*, *top*, *bottom*, *left*, *right*, as illustrated in Fig. 3-(a), resulting in more diverse band-pass filters. As shown in Fig. 3-(b), we use a small network with two paralleled conv blocks,  $g^{\text{Hrz}}(\cdot) : \mathbb{R}^{C \times H \times W} \rightarrow \mathbb{R}^{2 \times H \times W}$  and  $g^{\text{Vtc}}(\cdot) : \mathbb{R}^{C \times H \times W} \rightarrow \mathbb{R}^{2 \times H \times W}$  to estimate  $\mathbf{b}^G \in \mathbb{R}^4$  for each unit of  $\mathbf{X}$  in horizontal ( $\leftarrow \rightarrow$ ) and vertical ( $\uparrow \downarrow$ ) directions, respectively. To ensure the directional consistency, the horizontal branch possesses kernels with a size of  $1 \times (2b_{\max}^G + 1)$ , and the kernel size of the vertical branch is  $(2b_{\max}^G + 1) \times 1$ , correspondingly. All four directions have the same lower bound,  $b_{\min}^G = 1$  and the same upper bound,  $b_{\max}^G = 3$  by default.

Then, for each  $\mathbf{x}_c$ , its AdaPRF  $\mathcal{R}^\dagger = \{\mathbf{x}_c, \{\mathbf{x}_r^{(i)\dagger}\}_{i=1}^N\}$  ( $N = 16$  by default) can be obtained by the following two steps:

**Step 1.** Guard bandwidth estimation:

$$\mathbf{B}^{G^\dagger} = (b_{\max}^G - b_{\min}^G) \cdot \sigma(\beta(g^{\text{Vtc}}(\mathbf{X}) \oplus g^{\text{Hrz}}(\mathbf{X}))) + b_{\min}^G \cdot \mathbf{1}, \quad (5)$$



where, each  $\mathbf{b}_{h,w}^{G^\dagger}$  in  $\mathbf{B}^{G^\dagger}$  is the learned guard bandwidth for  $\mathbf{x}_c$  of the spatial coordinate  $(h, w)$ ; together with a constant  $b_{\max}^G - b_{\min}^G$ , the sigmoid activation  $\sigma(\cdot)$  can modulate the input values to  $(0, b_{\max}^G - b_{\min}^G)$ ;  $\beta(\cdot)$  and  $\oplus$  denotes the BatchNorm and Concat operator, respectively;  $\mathbf{1} \in \mathbb{Z}^{4 \times H \times W}$  is an all-one cube.

**Step 2.** Reference unit sampling: for better illustration, we first define the default PRF with guard bandwidths equal to 1 in [32] as  $\mathcal{R}_0 = \{\mathbf{x}_c, \{\mathbf{x}_r^{(i)}\}_{i=1}^N\}$ , which is shown in Fig. 2-(a). let  $\mathbf{p}_c = (h_c, w_c)$  and  $\mathbf{p}_r^{(i)} = (h_r^{(i)}, w_r^{(i)})$  denote the spatial coordinate of the current center unit  $\mathbf{x}_c$  and one of its reference unit  $\mathbf{x}_r^{(i)}$  in  $\mathcal{R}_0$ , respectively. Then we split  $\{\mathbf{x}_r^{(i)}\}_{i=1}^N$  into four subsets as  $\{\mathbf{x}_{r\uparrow}^{(j)}\}_{j=1}^M$ ,  $\{\mathbf{x}_{r\downarrow}^{(j)}\}_{j=1}^M$ ,  $\{\mathbf{x}_{r\leftarrow}^{(j)}\}_{j=1}^M$  and  $\{\mathbf{x}_{r\rightarrow}^{(j)}\}_{j=1}^M$ , according to the four directions, where  $M = N/4$ , as indicated by the braces drawn in Fig. 2-(a). By using  $\mathbf{b}^{G^\dagger}$  obtained from **Step 1**, we can sample  $\mathcal{R}^\dagger = \{\mathbf{x}_c, \{\mathbf{x}_r^{(i)\dagger}\}_{i=1}^N\}$  through linear interpolation, as exemplified in Fig. 3-(a). Taking the top direction as an example, each  $\mathbf{x}_{r\uparrow}^{(j)\dagger}$  can be obtained by

$$\mathbf{x}_{r\uparrow}^{(j)\dagger} = \left[1 - \left(b_{\uparrow}^{G^\dagger} - \lfloor b_{\uparrow}^{G^\dagger} \rfloor\right)\right] \cdot \mathbf{x}_{h_{r\uparrow}^{(j)} - \lfloor b_{\uparrow}^{G^\dagger} - 1 \rfloor, w_{r\uparrow}^{(j)}} + \left(b_{\uparrow}^{G^\dagger} - \lfloor b_{\uparrow}^{G^\dagger} \rfloor\right) \cdot \mathbf{x}_{h_{r\uparrow}^{(j)} - \lceil b_{\uparrow}^{G^\dagger} - 1 \rceil, w_{r\uparrow}^{(j)}}. \quad (6)$$

Sampling values for the rest directions and back-propagation of gradients in linear interpolation are demonstrated in Appendix B.2.

### 3.2 Fine-tuning AdaPKC with Thresholding On-line Switch

To further release the learning ability of AdaPKC, a fine-tuning strategy is proposed. To improve interpretability, we focus mainly on optimizations for the explicit AdaPRF estimation version, *i.e.*, AdaPKC<sup>ξ</sup>. The motivation comes from two perspectives. i) **Model confidence**: the representation ability and confidence of the model gradually improve with training, hence, during early training phase the less representative features may result in unreliable metric scores calculated in Eq. 2, consequently affecting the selection of AdaPRF and misleading the subsequent learning process. ii) **Data sparsity**: AdaPKC is responsible for the PRF adaptation for both background units ( $\mathbf{x}_c \notin \mathcal{S}_t$ ) and target units ( $\mathbf{x}_c \in \mathcal{S}_t$ ). However, due to the sparsity of target occupation *w.r.t.* the radar detection range, most units in the input feature map are background units, thus the heavy class-imbalance hinders the model's concentration on target points and their AdaPRF optimization.

To address the above issues, we propose to fine-tune AdaPKC<sup>ξ</sup> with a thresholding on-line switch (FiTOS), and please refer to Appendix C for a visual illustration. Firstly, a pre-trained PKC model with pre-defined PRFs is used to initialize the AdaPKC-based model to be optimized, so that AdaPKC can have a **warm-start before PRF adjustment**. Thus, the risk of obtaining unreliable metric scores is greatly reduced. Secondly, considering that the majority of background units are occupied by locally similar noise, *i.e.*, their monotonic correlation/metric curves,  $\text{sort}(\Xi)$ , are relatively flat, FiTOS introduces a **confidence threshold  $\tau$  to filter out these background units in spatial dimension on-the-fly**. Specifically, if the steepness of the metric curve, *i.e.*,  $\max\{g[\text{sort}(\Xi)]\}$ , is below the threshold  $\tau$ , then the corresponding unit is considered a background unit, and we retain the initial PRF, *i.e.*, switch off PRF adjustment. Otherwise, we adopt the newly estimated AdaPRF, *i.e.*, switch on PRF adjustment. As a result, the PRF selection criterion in Eq. (3) is modified as follows,

$$\begin{aligned} \mathcal{R}^\dagger &\triangleq \mathcal{R}_{k^\dagger} \xleftarrow{k=k^\dagger} \{\mathbf{x}_c\} \cup \{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^N, \\ \text{s.t., } k^\dagger &= \begin{cases} \arg \max_k \{g[\text{sort}(\Xi)]\} & , \text{ if } \max\{g[\text{sort}(\Xi)]\} > \tau \\ k_0 & , \text{ otherwise} \end{cases}, \end{aligned} \quad (7)$$

where  $k_0$  denotes the index of pre-defined PRF in the pre-trained model.

## 4 Experiments

To verify the effectiveness of our methods, we conduct quantitative and qualitative experiments on two public multi-view radar datasets and our self-collected single-view radar dataset. For simplicity of notations, AdaPKC<sup>ξ</sup>-based and AdaPKC<sup>θ</sup>-based multi-view RSS model is denoted as AdaPKC<sup>ξ</sup>-Net and AdaPKC<sup>θ</sup>-Net, respectively. The proposed single-view baseline model is named KuRALS-Net. Additionally, we append “*FiT*” in the upper right corner of the models to indicate the use of FiTOS.

#### 4.1 Datasets and Training Setups

**CARRADA** [20] dataset is recorded by a low-cost FMCW radar in millimeter wave band ( $\approx 77\text{GHz}$ ). It comprises camera-radar synchronised multi-view radar recordings in various scenarios, and contains four categories of objects: *pedestrian*, *cyclist*, *car* and *background*. The dimensions of provided range-angle-Doppler (RAD) tensors are  $256 \times 256 \times 64$  and support multi-view (RA and RD views) RSS task. The dataset splits are the same as in [32, 19]. **CARRADA-RAC** [32] dataset is derived from CARRADA and mainly calibrates the original RA annotations, see Appendix D.1 for details.

**KuRALS** dataset is self-collected by a **K**urz-under band ( $\approx 17\text{GHz}$ ) surveillance **R**adar, which is recorded in multiple scenarios, including **A**erial vehicles, **L**and targets and ships on the **S**ea surface. Different from other public datasets, *e.g.*, CARRADA [20] and CRUW [29], KuRALS aims at exploring the performance of deep models in the field of monitoring radar, hence offering a greater range field ( $\leq 6.4\text{km}$ ) and higher Doppler resolution ( $\approx 0.198\text{m/s}$ ). The comprised RD tensors are stored as 2D matrices of size 2048 (range) by 128 (Doppler). This dataset contains 9 sequences of radar recordings and there exist four moving object categories: *UAV*, *pedestrian*, *vehicle* and *ship*.

**Training Setups.** Following previous works, all models are evaluated with Intersection over Union (IoU) and Dice scores. These metrics are averaged across all classes on the test subset for model performance comparison, yielding mean IoU (mIoU) and mean Dice (mDice). **Implementation details** are presented in Appendix D.2.

#### 4.2 Investigation of AdaPKC Mechanism

To investigate the working mechanism of AdaPKCs, a series of comparison analysis is conducted on CARRADA benchmark. We first compare AdaPKCs with PeakConv and a manual PRF adjustment method. Results in Tab. 1 demonstrate that, both versions of AdaPKC exhibit significant enhancements in RSS performance compared to PeakConv, especially in the RA view, and they incur affordable additional computational complexity and inference speed overhead. Considering the severer signal tailing effect in RA view, this suggests that AdaPKC can better handle situations with frequency response ambiguity. For better illustration, we analyze the distribution of guard bandwidths in Appendix E.1. Additionally, since manually adjusting PRFs directly affects the RSS performance of PeakConv-based models, as discussed in [32], in this work we undertake a similar exploration experiment within a broader range of guard bandwidths. Specifically, different guard bandwidths in range dimension,  $b_R^G \in \{1, 2, 3, 4, 5\}$ , are tested, while the guard bandwidths in angle and Doppler dimensions are fixed at 1 for controlling variables. To ensure consistent parameter counts under different bandwidth settings, we adopt the same strategy of uniform sampling as in AdaPKC<sup>ξ</sup>. As shown in Tab. 2, presetting a proper PRF globally in a hyper-parameter way can indeed help the PeakConv-based model achieve better performance. However, the manual adjustment way cannot cater to each unit, and the computational cost of traversing the guard bandwidth in all dimensions and directions is quite large. In contrast, AdaPKC completely automates the PRF adjustment, and the adjustment granularity reaches the unit-level. With this unit-level PRF adaptation capability, AdaPKCs demonstrate superior RSS performance compared to the manual adjustment way.

Table 1: Comparison between AdaPKCs and PeakConv. The best and secondary results are marked with **bold** and underline, correspondingly. Frame rate is calculated on a workstation with an Intel(R) Xeon(R) Platinum 8255C CPU and a Tesla V100-SXM2 GPU.

Conv Type	#Params @Frames	RD View		RA View		GMACs	FPS
		mIoU	mDice	mIoU	mDice		
<b>PeakConv</b>	6.3M@5	60.7%	72.6%	43.1%	53.7%	<b>109.8</b>	<b>21.1</b>
<b>AdaPKC<sup>ξ</sup></b>	6.3M@5	<u>61.2%</u>	<u>73.1%</u>	<b>44.1%</b>	<b>55.1%</b>	<b>109.8</b>	<u>20.7</u>
<b>AdaPKC<sup>θ</sup></b>	6.3M@5	<b>61.5%</b>	<b>73.6%</b>	<u>43.6%</u>	<u>54.5%</u>	<u>110.1</u>	18.8

Furthermore, a comparative analysis between AdaPKCs and DefConvs (DefConv and DefConvV2) [5, 33] is conducted under the same RSS framework. Results in Tab. 3 show that, sharing the similar unit-level dynamic RF adjustment spirit with AdaPKC, DefConvs demonstrate better RSS performance than regular convolution (Conv). However, due to the three task-mismatched reasons discussed in § 2, DefConvs exhibit inferior applicability compared to AdaPKC, highlighting AdaPKC’s suitability for



Table 2: The effectiveness of manual PRF adjustment in PKC.  $b_R^G$  represents guard bandwidth in range dimension.

$b_R^G$	RD View		RA View	
	mIoU	mDice	mIoU	mDice
1	<u>60.7%</u>	72.6%	<u>43.1%</u>	53.7%
2	<b>61.0%</b>	<b>73.0%</b>	<b>43.3%</b>	<b>54.1%</b>
3	<u>60.7%</u>	<u>72.7%</u>	42.5%	53.2%
4	59.7%	<u>71.5%</u>	43.0%	<u>53.8%</u>
5	60.4%	72.5%	42.7%	53.3%

Table 3: Comparison between AdaPKCs and DefConvs (DefConv and DefConvV2).

Conv Type	#Params @Frames	RD View		RA View	
		mIoU	mDice	mIoU	mDice
Conv	5.6M@5	56.1%	68.0%	37.7%	46.2%
DefConv	5.7M@5	58.0%	69.8%	39.1%	48.1%
DefConvV2	5.8M@5	58.8%	70.6%	39.3%	48.6%
AdaPKC $^\xi$	6.3M@5	<u>61.2%</u>	<u>73.1%</u>	<b>44.1%</b>	<b>55.1%</b>
AdaPKC $^\theta$	6.3M@5	<b>61.5%</b>	<b>73.6%</b>	<u>43.6%</u>	<u>54.5%</u>

achieving adaptive receptive fields in radar signals. For supplementary purposes, we also show the comparison results between AdaPKC and dynamic CFAR detectors [10, 25, 23] in Appendix E.2.

As further explorations, we present an investigation into adaptive sampling strategies for reference band in Appendix E.3. Additionally, in Appendix E.4 we demonstrate a training strategy to enlarge the search space of alternative PRFs under fixed resource constraints for AdaPKC $^\xi$ .

### 4.3 Comparison with State-of-The-Art (SoTA)

Our methods are further compared with fashionable visual segmentation models and existing SoTA RSS solutions. The **quantitative results** on the CARRADA benchmark are illustrated in Tab. 4 and the **qualitative comparisons** are presented in Appendix F.3. Both AdaPKC $^\xi$ -Net and AdaPKC $^\theta$ -Net outperform previous RSS models, including pure CNN models [13, 8, 19, 32] and transformer-assisted CNN models [12, 6]. Compared to the baseline model PKCIn-Net, AdaPKC $^\xi$ -Net exhibits improvements in both RD and RA views, with a particularly notable enhancement in the RA view. AdaPKC $^\theta$ -Net purely relies on task-driven PRF adjustment, achieving a better performance balance between two views. Additionally, without consuming extra training resources, our proposed FiTOS strategy further enhances RSS performance of AdaPKC $^\xi$ -Net, achieving an overall superiority over AdaPKC $^\theta$ -Net.

Table 4: SoTA comparison on CARRADA benchmark. '-' denotes an unreported and not replicable value. Detailed results by category are presented in Appendix F.1, and the trade-off between performance and complexity of these models is illustrated in Appendix F.2.

Frameworks	#Params @Frames	RD View		RA View	
		mIoU	mDice	mIoU	mDice
FCN[18]	134.3M@3	54.7%	66.3%	34.5%	40.9%
U-Net[24]	17.3M@3	55.4%	68.0%	32.8%	38.2%
DeepLabv3+[1]	59.3M@3	50.8%	61.6%	32.7%	38.3%
RSS-Net[13]	10.1M@3	32.1%	36.9%	32.1%	37.8%
RAMP-CNN[8]	106.4M@9	56.6%	68.5%	27.9%	30.5%
TMVA-Net[19]	5.6M@5	56.1%	68.0%	37.7%	46.2%
TransRadar[6]	4.9M@5	57.2%	69.1%	39.9%	49.5%
T-RODNet[12]	162.0M@4	-	-	43.5%	53.6%
PKCIn-Net[32]	6.3M@5	60.7%	72.6%	43.1%	53.7%
AdaPKC $^\xi$ -Net	6.3M@5	61.2%	73.1%	<u>44.1%</u>	<u>55.1%</u>
AdaPKC $^\theta$ -Net	6.3M@5	<u>61.5%</u>	<u>73.6%</u>	43.6%	54.5%
AdaPKC $^\xi$ -Net <sup>FiT</sup>	6.3M@5	<b>62.1%</b>	<b>74.0%</b>	<b>44.3%</b>	<b>55.5%</b>

Table 5: Performance comparison on CARRADA-RAC.

Frameworks	#Params @Frames	RD View		RA View	
		mIoU	mDice	mIoU	mDice
TMVA-Net	5.6M@5	59.7%	69.9%	46.6%	57.9%
PKCIn-Net	6.3M@5	60.6%	72.4%	47.3%	58.7%
AdaPKC $^\xi$ -Net	6.3M@5	<b>61.6%</b>	<b>73.6%</b>	47.9%	59.3%
AdaPKC $^\theta$ -Net	6.3M@5	60.7%	<u>72.7%</u>	<u>48.1%</u>	<u>59.6%</u>
AdaPKC $^\xi$ -Net <sup>FiT</sup>	6.3M@5	<u>60.8%</u>	<u>72.7%</u>	<b>48.8%</b>	<b>60.5%</b>

Table 6: RSS performance comparison on KuRALS. Detailed results by category are presented in Appendix F.4.

Frameworks	#Params@Frames	mIoU	mDice
FCN	134.3M@5	50.4%	59.4%
U-Net	17.3M@5	52.4%	60.1%
DeepLabv3+	59.3M@5	52.6%	61.8%
KuRALS-Net	1.2M@5	56.0%	65.5%
KuRALS-Net w/ PKC	1.2M@5	56.7%	65.9%
KuRALS-Net w/ AdaPKC $^\xi$	1.2M@5	57.3%	<u>67.2%</u>
KuRALS-Net w/ AdaPKC $^\theta$	1.2M@5	57.8%	<b>67.6%</b>
KuRALS-Net w/ AdaPKC $^\xi$ <sup>FiT</sup>	1.2M@5	<b>58.2%</b>	<b>67.6%</b>

Results on CARRADA-RAC dataset are shown in Tab. 5. Similar to the trend on CARRADA, the evaluation results exhibit the superior performance of our methods than these RSS baseline models. AdaPKC $^\theta$ -Net still shows a relatively balanced performance improvement in both views. AdaPKC $^\xi$ -Net demonstrates a more pronounced improvement in the RD view, while

AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> shows a greater enhancement in the RA view. We speculate that this might be attributed to the collaborative training of different views in the multi-view segmentation task.

#### 4.4 RSS Performance on KuRALS

To verify the effectiveness of AdaPKCs in other application scenarios, we conduct comparative experiments on KuRALS dataset. **Quantitative results** are shown in Tab. 6 and the **qualitative comparisons** are presented in Appendix F.5. Compared to conventional visual segmentation models, our proposed KuRALS-Net offers a stronger baseline and AdaPKCs further boost the RSS performance of KuRALS-Net, validating their application potential in surveillance radar detection scenarios.

#### 4.5 Ablation Study for FiTOS

Since the threshold  $\tau$  plays a vital role in FiTOS, we study the impact by testing different values of  $\tau$  from 0.1 to 0.9. The summary results on mDice are shown in Fig. 4 and the corresponding mIoU results are presented in Appendix F.6. The optimal outcome for AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> in RD and RA view is obtained with  $\tau = 0.6$  and  $0.7$ , respectively. It is observed that the performance of AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> declines when  $\tau$  is either too large or too small. When  $\tau$  goes larger, fewer target units adjust their PRFs during the fine-tuning stage, rendering AdaPKC<sup>ξ</sup> less effective. Conversely, when  $\tau$  is small, AdaPKC<sup>ξ</sup> tends to focus on background units with random fluctuations, resulting in training confusion. Compared to PKCIn-Net, AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> consistently demonstrates superiority under different  $\tau$ s, highlighting the essentiality of PRF adaptation for the original PKC. When compared to AdaPKC<sup>ξ</sup>-Net, AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> exhibits superior performance with most values of  $\tau$ , demonstrating the efficacy of the fine-tuning strategy. Especially, in RD view, the RSS performance shows significant improvement within a broad range of  $\tau$ . Taking mDice as an example, it consistently surpasses 73.5% for  $\tau \in [0.5, 0.8]$ , indicating a strong level of robustness *w.r.t.*  $\tau$ .

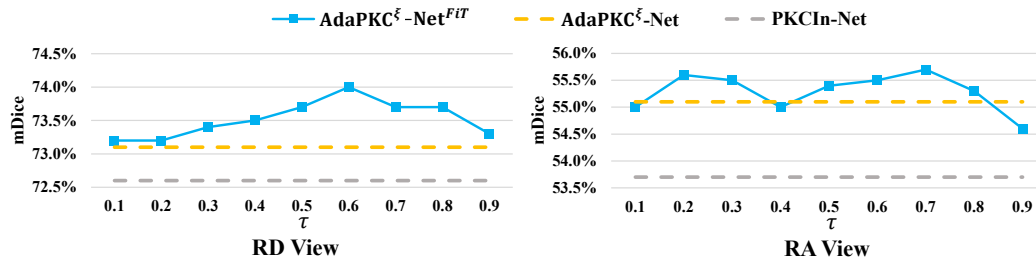


Figure 4: RSS performance of AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> with different values of  $\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ . AdaPKC<sup>ξ</sup>-Net and PKCIn-Net actually corresponds to the case where  $\tau$  of AdaPKC<sup>ξ</sup>-Net<sup>FiT</sup> equals to 0 in whole training process and 1 in the fine-tuning stage, respectively.

## 5 Conclusion

This work delves deeply into the convolution operator for radar signals, PKC, and improves upon it. Due to the design of PRF, PKC filters out the reference units corresponding to interfering signals in a band-pass filtering manner. Compared to other convolution operators in deep learning, PKC obtains a more robust representation by using the center unit and the reference unit to cancel each other out and then weighted fusion. However, the fixed suppression bandwidth (*i.e.*, guard bandwidth) setting limits the adaptability of PKC to signal diversity. Based on this, we propose a method for adaptive adjustment of the PRF, and provide two effective solutions based on metrics and learning, *i.e.*, AdaPKC<sup>ξ</sup> and AdaPKC<sup>θ</sup>. In addition, to further boost the learning ability of AdaPKC, a novel fine-tuning strategy is presented. To fully verify the effectiveness of AdaPKC, different real-measured radar datasets are used for experimental analysis. The results show the superior performance of AdaPKC in RSS tasks.

## References

- [1] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.
- [2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 833–851, Cham, 2018. Springer International Publishing.
- [3] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [4] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. 2022.
- [5] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [6] Yahia Dalbah, Jean Lahoud, and Hisham Cholakkal. Transradar: Adaptive-directional transformer for real-time multi-view radar semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 353–362, 2024.
- [7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [8] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu. Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. *IEEE Sensors Journal*, 21(4):5119–5132, 2021.
- [9] Daniel Gusland, Sigmund Rolfsjord, and Børge Torvik. Deep temporal detection - a machine learning approach to multiple-dwell target detection. In *2020 IEEE International Radar Conference (RADAR)*, pages 203–207, 2020.
- [10] V Gregers Hansen. Constant false alarm rate processing in search radars. In *IEE Conf. Publ. no. 105, "Radar-Present and Future"*, pages 325–332, 1973.
- [11] Ahsan Jalil, Hassan Yousaf, and Muhammad Iram Baig. Analysis of cfar techniques. In *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pages 654–659. IEEE, 2016.
- [12] Tiezhen Jiang, Long Zhuang, Qi An, Jianhua Wang, Kai Xiao, and Anqi Wang. T-rodnet: Transformer for vehicular millimeter-wave radar object detection. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2022.
- [13] Prannay Kaul, Daniele de Martini, Matthew Gadd, and Paul Newman. Rss-net: Weakly-supervised multi-class semantic segmentation with fmcw radar. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 431–436, 2020.
- [14] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.
- [15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- [16] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [17] Zhang Liwen, Pan Jian, Zhang Youcheng, Chen Yuanpei, Ma Zhe, Huang Xuhui, and Sun Kewu. Capturing temporal-dependence in radar echo for spatial-temporal sparse target detection. *Journal of Radars*, 12(R22228):356, 2023.
- [18] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

- [19] Arthur Ouaknine, Alasdair Newson, Patrick Pérez, Florence Tupin, and Julien Rebut. Multi-view radar semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15671–15680, October 2021.
- [20] Arthur Ouaknine, Alasdair Newson, Julien Rebut, Florence Tupin, and Patrick Pérez. Carrada dataset: Camera and automotive radar with range- angle- doppler annotations. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5068–5075, 2021.
- [21] Qizhe Qu, Weijian Liu, Jiaxin Wang, Binbin Li, Ningbo Liu, and Yong-Liang Wang. Enhanced cnn-based small target detection in sea clutter with controllable false alarm. *IEEE Sensors Journal*, 23(9):10193–10205, 2023.
- [22] M. A. Richards. *Fundamentals of Radar Signal Processing, Second Edition*. Fundamentals of Radar Signal Processing, Second Edition, 2005.
- [23] Hermann Rohling. Radar cfar thresholding in clutter and multiple target situations. *IEEE Transactions on Aerospace and Electronic Systems*, AES-19(4):608–621, 1983.
- [24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [25] Gerard V Trunk. Range resolution of targets using automatic detectors. *IEEE Transactions on Aerospace and Electronic Systems*, (5):750–755, 1978.
- [26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [27] Hao Wan, Xiaoqing Tian, Jing Liang, and Xiaofeng Shen. Sequence-feature detection of small targets in sea clutter based on bi-lstm. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.
- [28] Jingang Wang and Songbin Li. Maritime radar target detection in sea clutter based on cnn with dual-perspective attention. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023.
- [29] Yizhou Wang, Zhongyu Jiang, Xiangyu Gao, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu. Rodnet: Radar object detection using cross-modal supervision. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 504–513, 2021.
- [30] Yizhou Wang, Zhongyu Jiang, Yudong Li, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu. Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization. *IEEE Journal of Selected Topics in Signal Processing*, 15(4):954–967, 2021.
- [31] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2016.
- [32] Liwen Zhang, Xinyan Zhang, Youcheng Zhang, Yufei Guo, Yuanpei Chen, Xuhui Huang, and Zhe Ma. Peakconv: Learning peak receptive field for radar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17577–17586, 2023.
- [33] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [34] Hao Zou, Zhen Xie, Jiarong Ou, and Yutao Gao. Transrss: Transformer-based radar semantic segmentation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6965–6972, 2023.

## Supplementary Material

### A AdaPKC-based RSS Frameworks

#### A.1 AdaPKC

We formulate the detailed process of how AdaPKC conducts PeakConv on  $\mathcal{R}^\dagger$ . For the simplicity of notation, we uniformly represent the  $\mathcal{R}^\dagger$  acquired by the two AdaPKCs as  $\mathcal{R}^\dagger = \{\mathbf{x}_c, \{\mathbf{x}_r^{(i)\dagger}\}_{i=1}^N\}$ .  $\mathcal{R}^\dagger$  can be readily incorporated into both PeakConv operators proposed in [32], including vanilla-PKC and ReDA-PKC. Given ReDA-PKC's enhanced capability in utilizing interference and target information for noise suppression compared to vanilla-PKC, we choose ReDA-PKC as the baseline operator. Therefore, both AdaPKCs share the same structure as ReDA-PKC and differ only in the generation of  $\mathcal{R}^\dagger$ . Analogous to the formalization of ReDA-PKC, we can formulate AdaPKC ( $\mathcal{R}^\dagger; \mathbf{W} \in \mathbb{R}^{C_{in} \times N \times C_{out}}$ ) :  $\mathbb{R}^{C_{in}} \rightarrow \mathbb{R}^{C_{out}}$  for each  $x_c$  as follows,

$$\text{AdaPKC}(\mathcal{R}^\dagger; \mathbf{W}) = \text{Vec} \left( \left\{ \sum_{i=1}^N \mathbf{w}_j^{(i)} \cdot (\mathbf{x}_c - \mathbf{x}_r^{(i)\dagger})^\top \right\}_{j=1}^{C_{out}} \right), \quad (\text{S1})$$

$$\text{where, } \mathbf{w}_j^{(i)} \in \mathbb{R}^{C_{in}} (j = 1, \dots, C_{out}).$$

where,  $\mathbf{W}$  is the general learning weights of AdaPKCs;  $\text{Vec}(\cdot)$  is the vectorization operator.

#### A.2 Multi-view RSS Framework

For multi-view RSS tasks on public radar datasets, in alignment with the design of PKCIn-Net [32], we replace the ReDA-PKC in PKCIn-Net with AdaPKC $^\xi$  or AdaPKC $^\theta$ . And this brings two multi-view RSS models, denoted as AdaPKC $^\xi$ -Net and AdaPKC $^\theta$ -Net correspondingly, whose overall frameworks are depicted in Fig. S1. The remaining components of both AdaPKC-Nets keep consistent with PKCIn-Net and TMVA-Net [19]: i) the designed networks take RD, RA and AD tensors as inputs and conduct semantic segmentation on both RD and RA views concurrently; ii) each encoding branch for a single view comprises two 3D convolution blocks, allowing for the comprehensive utilization of temporal information; iii) an ASPP [2] module is used as a component of the encoding branch to aggregate multi-scale spatial information; iv) after the encoding phase, a Latent Space Encoder (LSE) aids in integrating the encoding features from multiple views and transmits them to the decoding section for the final prediction.

#### A.3 Single-view RSS Framework

For single-view RSS tasks on our self-collected radar dataset, we establish a lightweight baseline model called KuRALS-Net, the architecture of which is illustrated in Fig. S2. This model takes temporally contiguous multi-frame RD tensors as inputs, utilizing an Encoder with 3D convolutions to simultaneously gather temporal and spatial information. Then, the encoding features pass through 2D convolution blocks followed by an ASPP module for further information fusion. Finally, the fused features are fed into the Decoder to obtain pixel-wise semantic predictions in the RD view. For the validation of AdaPKC, we replace the 2D convolution blocks in KuRALS-Net with different AdaPKC blocks.

## B Supplementary Details of AdaPKC

#### B.1 Analysis of Selection Criterion in AdaPKC $^\xi$

We provide a detailed analysis of the selection criterion adopted in Eq. 3. In Eq. 2, the metric score  $\xi_k$  for the  $k^{\text{th}}$  PRF is calculated by averaging the metric scores of related reference units  $\{\mathbf{x}_{r|k}^{(i)}\}_{i=1}^N$ . For reference units mainly occupied by  $s' \in \mathcal{S}_{t-n}$ , the corresponding  $\xi_k$  is statistically smaller than those primarily influenced by  $s' \in \mathcal{S}_t$  and larger than those controlled by  $s' \in \mathcal{S}_n$ . When it is reflected on the curve of  $\text{sort}(\Xi)$ ,  $\xi_k$  is likely to lie near the quickly descending position. In view of this

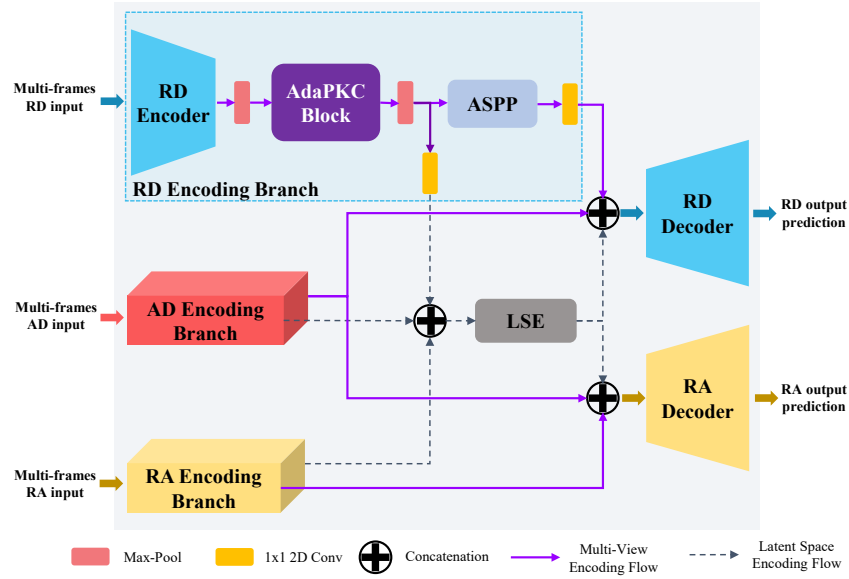


Figure S1: The overall frameworks of  $\text{AdaPKC}^{\xi}$ -Net and  $\text{AdaPKC}^{\theta}$ -Net, which only differ in the **AdaPKC Block**. Please note that the **AdaPKC Block** consists of two AdaPKC layers.

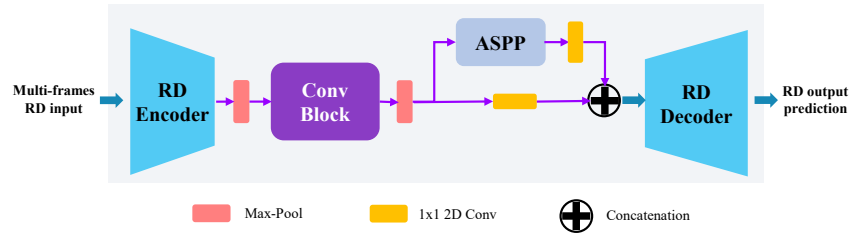


Figure S2: The overall framework of KuRALS-Net. The **Conv Block** can be readily replaced with different AdaPKC blocks to validate the effectiveness of AdaPKC.

attribute, we might adopt the median value of  $\Xi$  as the selection criterion. However, while the use of median value can locate  $\xi_k$  where it is desired in some case, it is clear that the maximum value of the first order gradient is more robust for a variety of scenarios, as illustrated in Fig. S3. Therefore, to promote the robustness of this selection, we finally employ the maximum value of the first-order gradient.

## B.2 Reference Units Sampling and Back-propagation in $\text{AdaPKC}^{\theta}$

The sampling values of reference units in down, left and right directions are obtained by the following forms of linear interpolation, accordingly:

$$\begin{aligned} \mathbf{x}_{r\downarrow}^{(j)\dagger} = & \left[1 - \left(b_{\downarrow}^{G\dagger} - \lfloor b_{\downarrow}^{G\dagger} \rfloor\right)\right] \cdot \mathbf{x}_{h_{r\downarrow}^{(j)} + \lfloor b_{\downarrow}^{G\dagger} - 1 \rfloor, w_{r\downarrow}^{(j)}} \\ & + \left(b_{\downarrow}^{G\dagger} - \lfloor b_{\downarrow}^{G\dagger} \rfloor\right) \cdot \mathbf{x}_{h_{r\downarrow}^{(j)} + \lceil b_{\downarrow}^{G\dagger} - 1 \rceil, w_{r\downarrow}^{(j)}}, \end{aligned} \quad (\text{S2})$$

$$\begin{aligned} \mathbf{x}_{r\leftarrow}^{(j)\dagger} = & \left[1 - \left(b_{\leftarrow}^{G\dagger} - \lfloor b_{\leftarrow}^{G\dagger} \rfloor\right)\right] \cdot \mathbf{x}_{h_{r\leftarrow}^{(j)}, w_{r\leftarrow}^{(j)} - \lfloor b_{\leftarrow}^{G\dagger} - 1 \rfloor} \\ & + \left(b_{\leftarrow}^{G\dagger} - \lfloor b_{\leftarrow}^{G\dagger} \rfloor\right) \cdot \mathbf{x}_{h_{r\leftarrow}^{(j)}, w_{r\leftarrow}^{(j)} - \lceil b_{\leftarrow}^{G\dagger} - 1 \rceil}, \end{aligned} \quad (\text{S3})$$

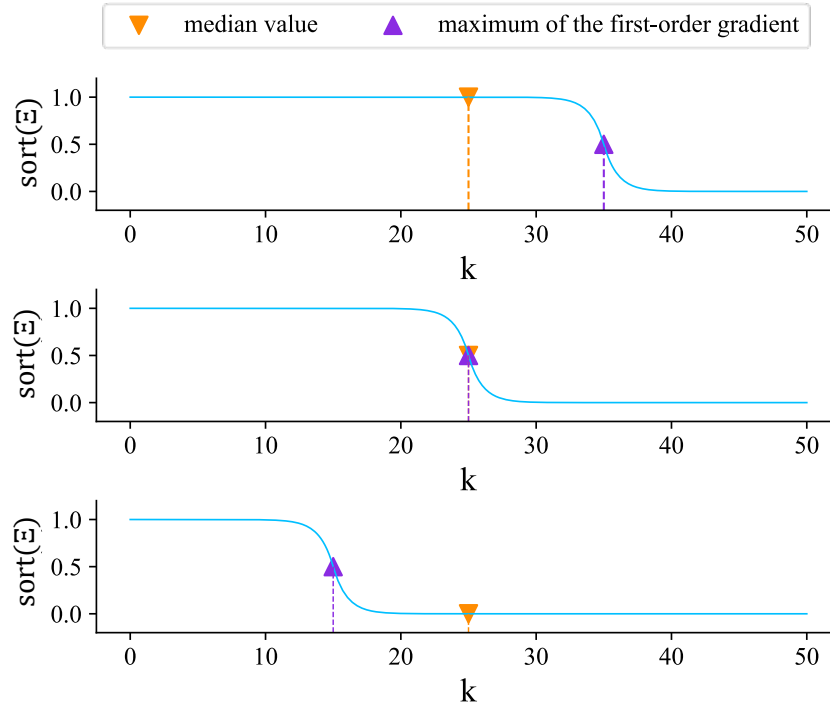


Figure S3: The comparison between different selection criteria across three representative scenarios. The median value fails to capture target-interfering noise in the first and last scenarios, whereas our introduced selection criterion demonstrates robustness in locating it.

$$\begin{aligned} \mathbf{x}_{r \rightarrow}^{(j)\dagger} &= [1 - (b_{\rightarrow}^{G\dagger} - \lfloor b_{\rightarrow}^{G\dagger} \rfloor)] \cdot \mathbf{x}_{h_{r \rightarrow}, w_{r \rightarrow}^{(j)} + \lfloor b_{\rightarrow}^{G\dagger} - 1 \rfloor} \\ &\quad + (b_{\rightarrow}^{G\dagger} - \lfloor b_{\rightarrow}^{G\dagger} \rfloor) \cdot \mathbf{x}_{h_{r \rightarrow}, w_{r \rightarrow}^{(j)} + \lceil b_{\rightarrow}^{G\dagger} - 1 \rceil}. \end{aligned} \quad (\text{S4})$$

For back-propagation of gradients in these linear interpolations, we compute the gradient of Eq. S1 w.r.t.  $\mathbf{b}^{G\dagger} = \{b_{\uparrow}^{G\dagger}, b_{\downarrow}^{G\dagger}, b_{\leftarrow}^{G\dagger}, b_{\rightarrow}^{G\dagger}\}$  as

$$\begin{aligned} &\frac{\partial \text{Vec} \left( \left\{ \sum_{i=1}^N \mathbf{w}_j^{(i)} \cdot (\mathbf{x}_c - \mathbf{x}_r^{(i)\dagger})^\top \right\}_{j=1}^{C_{\text{out}}} \right)}{\partial \mathbf{b}^{G\dagger}} \\ &= -\text{Vec} \left( \left\{ \sum_{i=1}^N \mathbf{w}_j^{(i)} \left( \frac{\partial \mathbf{x}_r^{(i)\dagger}}{\partial \mathbf{b}^{G\dagger}} \right)^\top \right\}_{j=1}^{C_{\text{out}}} \right). \end{aligned} \quad (\text{S5})$$

Since  $\{\mathbf{x}_r^{(i)\dagger}\}_{i=1}^N$  have been split into four subsets based on their directions, and each  $\mathbf{x}_r^{(i)\dagger}$  is solely determined by the guard bandwidth offset in its corresponding direction, there only remain four non-zero items in  $\left\{ \frac{\partial \mathbf{x}_r^{(i)\dagger}}{\partial \mathbf{b}^{G\dagger}} \right\}_{i=1}^N$ , i.e.,  $\frac{\partial \mathbf{x}_r^{(j)\dagger}}{\partial b_{\uparrow}^{G\dagger}}, \frac{\partial \mathbf{x}_r^{(j)\dagger}}{\partial b_{\downarrow}^{G\dagger}}, \frac{\partial \mathbf{x}_r^{(j)\dagger}}{\partial b_{\leftarrow}^{G\dagger}}$  and  $\frac{\partial \mathbf{x}_r^{(j)\dagger}}{\partial b_{\rightarrow}^{G\dagger}}$ , where  $j \in [1, M]$ . For simplicity, we demonstrate the calculation procedure with  $\frac{\partial \mathbf{x}_{r \rightarrow}^{(j)\dagger}}{\partial b_{\rightarrow}^{G\dagger}}$  as an example, and the remaining items follow a similar approach. From Eq. S4, we can get

$$\frac{\partial \mathbf{x}_{r \rightarrow}^{(j)\dagger}}{\partial b_{\rightarrow}^{G\dagger}} = -\mathbf{x}_{h_{r \rightarrow}, w_{r \rightarrow}^{(j)} + \lfloor b_{\rightarrow}^{G\dagger} - 1 \rfloor} + \mathbf{x}_{h_{r \rightarrow}, w_{r \rightarrow}^{(j)} + \lceil b_{\rightarrow}^{G\dagger} - 1 \rceil}. \quad (\text{S6})$$

With the combination of Eq. S5 and S6, we can obtain the desired gradients.



## C Visual Illustration of FiTOS

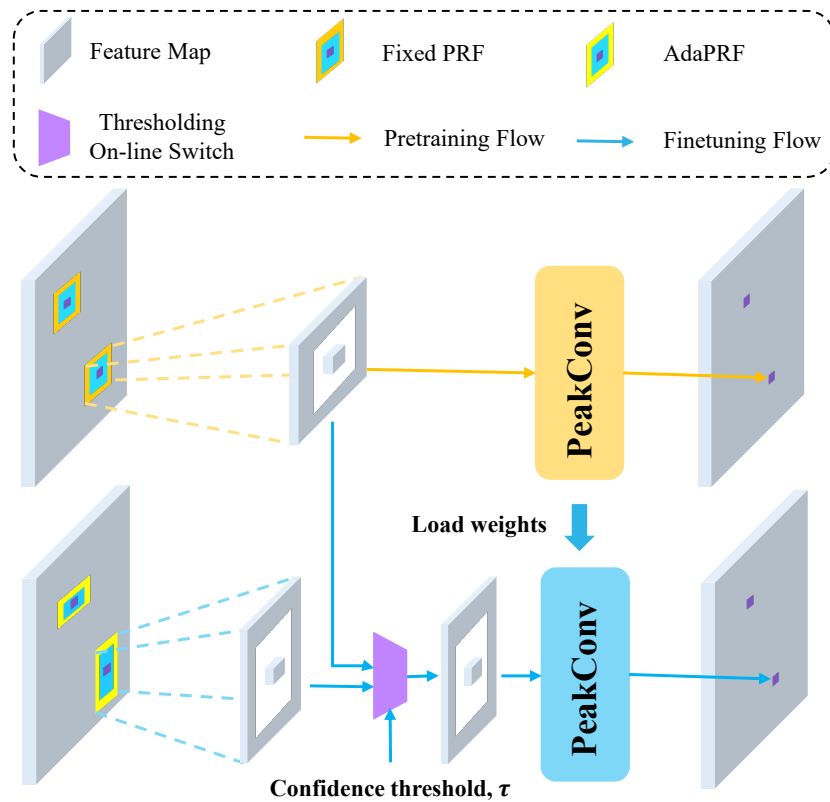


Figure S4: The illustration of FiTOS strategy. The **Thresholding On-line Switch** is detailed in Eq. 7.

The detailed process of FiTOS strategy is illustrated in Fig. S4.

## D Supplementary Details about Datasets and Training Setups

### D.1 Detailed Description of CARRADA-RAC Dataset

The CARRADA-RAC [32] dataset is derived from CARRADA and mainly calibrates the original RA annotations. For the generation of both RD and RA annotations, CARRADA adopts a semi-automatic method relying on the Mean-Shift clustering [3]. However, in CARRADA, the clustering performance is seriously degraded by unreliable centroid initialization from optical images and inaccurate candidate search space in RA representation. To alleviate these issues, CARRADA-RAC proposes an RD association strategy to refine the initial centroid, and introduces a regionalized CFAR for readjusting the search space.

### D.2 Implementation Details

**Multi-view RSS.** The input sizes of RA, AD and RD views are  $256 \times 256$ ,  $256 \times 64$  and  $256 \times 64$ , respectively. Both AdaPKC-Nets leverage a sequence of 5 input frames for temporal information aggregation, consistent with PKCIn-Net [32]. We train all these models on two NVIDIA-3090 GPUs and use the Adam optimizer [14] for training. The initial learning rate is  $1e-4$ , and decays in a cosine manner by default. We train these models for 300 epochs with a batch size of 6. For FiTOS, the 300 epochs are evenly distributed between the pre-training and fine-tuning stages, and we set  $\tau = 0.6$  for the fine-tuning stage by default. We train these models using a combination of weighted cross-entropy loss, Dice loss and coherence loss, configured with the recommended parameters outlined in [19].

**Single-view RSS.** In contrast to the multi-view RSS task, the experimental configurations for KuRALS dataset maintain consistency with the following two exceptions: i) the input includes only a single view of RD tensor after 0-Doppler frequency elimination, with shape of  $2048 \times 124$ ; ii) the weighted cross-entropy loss in multi-view RSS is substituted with weighted focal loss [15] to tackle the severer target-background imbalance in KuRALS.

## E More Exploration of AdaPKC mechanism

### E.1 Analysis of Guard Bandwidth Distribution in AdaPKC<sup>ξ</sup>-Net

We analyze the guard bandwidth distribution in AdaPKC<sup>ξ</sup>-Net from two perspectives: different categories and different views. The distribution histogram of guard bandwidths in AdaPKC<sup>ξ</sup>-Net is illustrated in Fig. S5. **From the perspective of different categories**, we observe that the guard bandwidth of the *background* class tends toward a uniform distribution, while that of the foreground class exhibits a more distinctly concentrated trend. For a better illustration, we present the variance values of different categories in Tab. S1, and the remarkable difference in variance between foreground and *background* classes confirms our observation. This observation aligns with the characteristics of our proposed metric score, indicating that AdaPKC<sup>ξ</sup> consistently enhances features of foreground-class targets, thus improving their discriminability from the *background* class. **From the perspective of different views**, this concentrated trend appears more apparent in the RA view compared to the RD view. Concretely, RA view shows a clear concentration on the guard bandwidth of (2, 1), while RD view exhibits considerable proportions on all guard bandwidths except (1, 1). Moreover, this concentrated guard bandwidth in the RA view remains consistent across different AdaPKC layers, while that of the RD view transitions from (2, 3) in the first layer to (1, 2) in the second layer. This comparative analysis of different views indicates that the guard bandwidth selection in RA view demonstrates more confidence and ease for AdaPKC, which explains the better performance improvement of AdaPKC on the RA view.

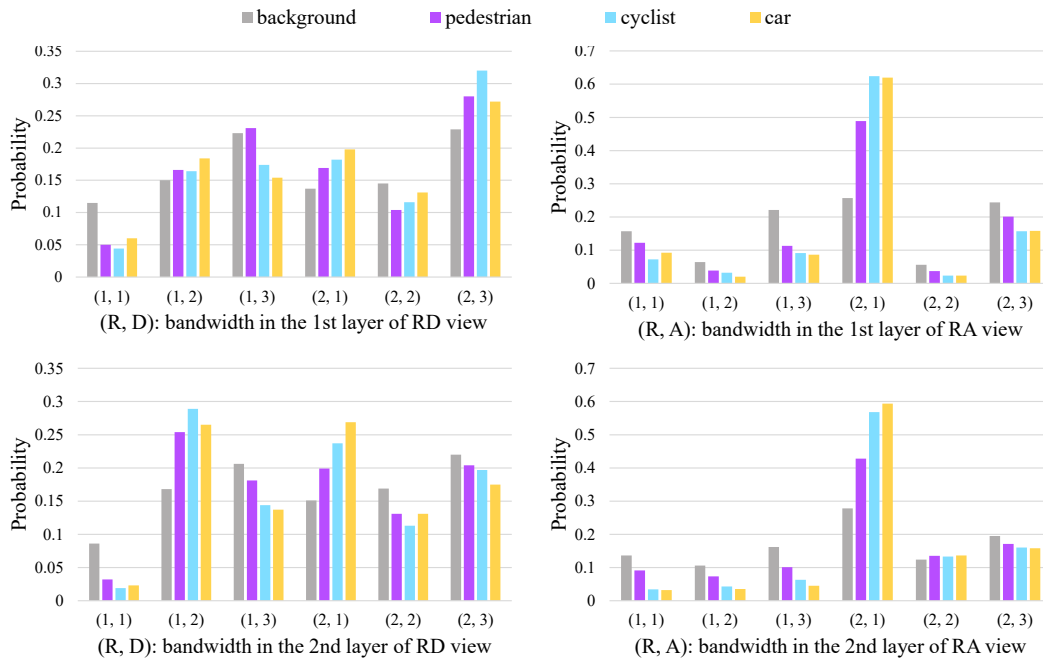


Figure S5: The category-wise probability distribution histogram of guard bandwidths in AdaPKC<sup>ξ</sup>-Net.

Table S1: Comparison of the variance values for different categories.  $\text{Var}^{1\text{st}}$  and  $\text{Var}^{2\text{nd}}$  represents the variance value of the first and second AdaPKC layer, respectively.

Category	RD View		RA View	
	$\text{Var}^{1\text{st}}(\%)$	$\text{Var}^{2\text{nd}}(\%)$	$\text{Var}^{1\text{st}}(\%)$	$\text{Var}^{2\text{nd}}(\%)$
<b>Bkg.</b>	0.19	0.19	0.67	0.33
<b>Ped.</b>	0.58	0.49	2.39	1.47
<b>Cyc.</b>	0.69	0.77	4.38	3.43
<b>Car</b>	0.42	0.72	4.33	3.90

## E.2 Comparison between AdaPKC and Dynamic CFAR Detectors

We compare the performance of AdaPKC and dynamic CFAR detectors, and the results are shown in Tab. S2. The comparison reveals several significant limitations of CFAR methods: i) They can only detect foreground targets without the ability to categorize them; ii) They show poor target identification performance, struggling with complex target and interference scenarios; iii) They rely on manual parameter tuning and lack adaptive learning capabilities. Therefore, it is evident that enhancing radar target perception paradigms through deep learning is both necessary and practical, serving as one of the key motivations for PKC and AdaPKC.

Table S2: Performance comparison between AdaPKC and dynamic CFAR detectors on the CAR-RADA dataset. *Bkg.*, *Ped.*, *Cyc.*, *Car* and *Frg.* represents *Background*, *Pedestrian*, *Cyclist*, *Car* and *Foreground* class, respectively, with *Foreground* class including *Pedestrian*, *Cyclist* and *Car* classes. FPS is calculated by the total inference time of the detection algorithm on both RA and RD views, using a workstation with an Intel(R) Xeon(R) Platinum 8255C CPU and a Tesla V100-SXM2 GPU.

View	Methods	FPS	IoU(%)					Dice(%)				
			Bkg.	Ped.	Cyc.	Car	Frg.	Bkg.	Ped.	Cyc.	Car	Frg.
RD	CA-CFAR	<b>47.5</b>	82.9	-	-	-	0.9	90.6	-	-	-	1.7
	SO-CFAR	36.0	77.8	-	-	-	1.2	87.5	-	-	-	2.4
	GO-CFAR	36.6	97.7	-	-	-	2.9	98.8	-	-	-	5.6
	OS-CFAR	42.8	93.8	-	-	-	3.1	96.8	-	-	-	6.1
	AdaPKC <sup>ξ</sup> -Net	20.7	<b>99.7</b>	55.1	31.9	58.2	-	<b>99.8</b>	71.0	48.2	73.4	-
	AdaPKC <sup>θ</sup> -Net	18.8	<b>99.7</b>	54.6	33.7	58.1	-	99.8	70.6	50.4	73.5	-
RA	CA-CFAR	<b>47.5</b>	89.2	-	-	-	0.4	94.3	-	-	-	0.8
	SO-CFAR	36.0	75.8	-	-	-	0.4	86.2	-	-	-	0.8
	GO-CFAR	36.6	97.2	-	-	-	0.6	98.6	-	-	-	1.2
	OS-CFAR	42.8	91.6	-	-	-	0.5	95.6	-	-	-	1.0
	AdaPKC <sup>ξ</sup> -Net	20.7	<b>99.8</b>	24.4	17.1	35.0	-	<b>99.9</b>	39.3	29.2	51.9	-
	AdaPKC <sup>θ</sup> -Net	18.8	<b>99.8</b>	26.5	15.8	32.3	-	<b>99.9</b>	41.9	27.3	48.9	-

## E.3 Exploration of Adaptive Sampling Strategies for Reference Band

We demonstrate the exploration of adaptive sampling strategies for reference band. Sampling for reference band plays an indispensable role in multiple applications of peak convolutions: i) for PKC with different guard bandwidth settings, sampling fixed  $N$  reference points ensures consistent parameter counts; ii) for AdaPKCs, different  $\mathbf{x}_c$  within the feature map might claim diverse PRFs, necessitating reference points sampling to meet the requirement of convolution operations; iii) for future research on reference bandwidth adjustment, this sampling mechanism operates in a manner akin to that in AdaPKCs. For simplicity of exploration, we take the mentioned application in PKC as the subject of our experiment. In addition to the common uniform sampling, we explore two adaptive sampling strategies based on the similar principle of AdaPKC<sup>ξ</sup>: For the candidates of reference points  $\{\mathbf{x}_r^{(i)}\}_{i=1}^{N_r}$  in a fixed reference band, we first compute their inner product (similarity) with  $\mathbf{x}_c$ , i.e.,  $\mathbf{x}_c \mathbf{x}_r^T$ . Then, the first version (v1) of adaptive sampling selects the  $N$  least similar reference points to locate target-interfering noise, while the second version (v2) of adaptive sampling chooses the  $N$  reference points with intermediate similarity instead. To evaluate the effectiveness

of these sampling strategies, we conduct comparative experiments on the CARRADA dataset, with the guard bandwidth of PKC fixed at 2 for all dimensions. The evaluation results are shown in Tab. S3. Compared with uniform sampling, adaptive sampling v1 consistently shows improvement across two views, further affirming the effectiveness of our introduced metric. In the case of adaptive sampling v2, we observe a similar trend as with AdaPKC<sup>ξ</sup>, which exhibits a more pronounced performance enhancement for RA view. Nevertheless, in comparison to the adaptive guard bandwidth adjustment in AdaPKCs, these adaptive samplings of reference band encompass a more limited range of adjustable receptive field, thus demonstrating less improvements on RSS performance.

Table S3: Comparison of different sampling strategies for reference band on the CARRADA dataset. The best and secondary results are marked with **bold** and underline, correspondingly.

Sampling Strategy	#Params @Frames	RD View		RA View	
		mIoU	mDice	mIoU	mDice
Uniform sampling	6.3M@5	60.9%	<u>72.9%</u>	42.8%	53.6%
Adaptive sampling v1	6.3M@5	<b>61.3%</b>	<b>73.1%</b>	<u>43.2%</u>	<u>53.7%</u>
Adaptive sampling v2	6.3M@5	<u>61.0%</u>	72.7%	<b>43.5%</b>	<b>54.1%</b>

#### E.4 Exploration of Enlarging PRF Search Space in AdaPKC<sup>ξ</sup>

To achieve AdaPRF estimation in a wider range of PRF search space under fixed resource constraints, we propose a novel training strategy named *Voting-driven Multi-round Training (Vot-MRT)*. Due to limitations in GPU memory resources, AdaPKC<sup>ξ</sup> restricts the selection of AdaPRF within a local range. To determine the AdaPRF in a wider range of PRF search space, *Vot-MRT* employs a step-wise local optimization strategy, inspired by greedy algorithms [4]. Concretely, *Vot-MRT* involves multiple rounds of training, and each successive round provides a better initialization for the guard bandwidth search space,  $\Omega^G$ , of the latter round. Given that different pixels may pick distinct guard bandwidths, *Vot-MRT* implements a voting mechanism to gather the most concentrated guard bandwidth for the subsequent round. The detailed algorithm of *Vot-MRT* strategy is presented in Algo. 1.

To evaluate the effectiveness of *Vot-MRT*, we conduct a quantitative experiment on CARRADA dataset. We keep  $\Omega_0^G$  consistent with  $\Omega^G$  denoted in § 3.1.1, and set the value of  $N$  to 3 by default, indicating that the original training process of AdaPKC<sup>ξ</sup>-Net will be repeated for three rounds in *Vot-MRT*. The results are shown in Tab. S4 and it can be observed that *Vot-MRT* brings visible segmentation performance improvements to AdaPKC<sup>ξ</sup>-Net by gradually enlarging the search space of alternative PRFs.

## F Supplementary Experiment Results

### F.1 Semantic Segmentation Results by Category on CARRADA

The category-wise semantic segmentation results on the CARRADA-Test dataset are shown in Tab. S5. Compared with TMVA-Net, PKCIn-Net employs PeakConv to replace the regular convolution, resulting in impressive performance improvements. Our proposed methods further enhance the capabilities of PeakConv, leading to significant enhancements across nearly all classes and views, with a minor decrease observed in the *car* class of RA view. Consequently, the proposed methods achieve superior RSS performance and a better trade-off among different classes, which still holds true when compared to other SoTA RSS models.

### F.2 Performance vs. Complexity

The trade-off between performance and complexity of different RSS models is presented in Fig. S6.

### F.3 Qualitative Comparisons on CARRADA

We present in Fig. S7 the qualitative comparisons of different methods on three frames from the CARRADA test split, with each frame exhibiting varying levels of interference with targets. In the

---

**Algorithm 1:** Voting-driven Multi-round Training

---

**Input:** Initial guard bandwidth search space:  $\Omega_0^G = \{\{b_x^G, b_y^G\} \mid b_x^G \in \Omega_{x|0}^G, b_y^G \in \Omega_{y|0}^G\}$ ;

Number of training rounds:  $N$ .

**Initialization:**

The maximum selection frequency of  $b_x^G$  and  $b_y^G$ :  $\{C_x^* = 0, C_y^* = 0\}$ .

**for**  $i \leftarrow 0$  **to**  $N - 1$  **do**

    AdaPKC $^\xi$ -Net\*  $\leftarrow$  Getting AdaPKC $^\xi$ -Net trained from scratch with  $\Omega_i^G$ ;

$F_i \leftarrow$  Getting the feature map calculated by AdaPKC $^\xi$ -Net\*;

**for**  $d$  **in**  $\{x, y\}$  **do**

**for**  $b_d^G$  **in**  $\Omega_{d|i}^G$  **do**

$C_d \leftarrow$  Getting the cumulative frequency of  $b_d^G$  being selected across all units in  $F_i$ ;

**if**  $C_d \geq C_d^*$  **then**

$C_d^* = C_d$ ;

$b_d^{G*} = b_d^G$ ;

**end**

**end**

$C_d^* = 0$ ;

**end**

$\mathbf{b}^{G*} = \{b_x^{G*}, b_y^{G*}\}$ ;

$\Omega_{i+1}^G \leftarrow$  Getting a new local search space extending from  $\mathbf{b}^{G*}$ ;

**end**

**Output:** Return AdaPKC $^\xi$ -Net\*.

---

Table S4: Results of AdaPKC $^\xi$ -Net with *Vot-MRT* strategy on CARRADA dataset. The number of training rounds is set to 3 by default.

Round	RD View		RA View	
	mIoU	mDice	mIoU	mDice
1	61.2%	73.1%	44.1%	55.1%
2	61.6%	73.6%	43.7%	54.5%
3	<b>61.8%</b>	<b>73.8%</b>	<b>44.6%</b>	<b>55.7%</b>

first frame, where the target is interfered with minor noise/clutter, results indicate that all methods can accurately locate and classify targets in relatively clean RD view. However, in the RA view, methods without peak convolution struggle to identify target regions completely in the presence of stronger interference. The interference suppression capability equips PKCIn-Net with superior identification performance, while AdaPKC $^\xi$ -Net<sup>*FiT*</sup> further strengthens this capability and recognizes complete target regions. In the second frame, as clutter interference on targets intensifies, PKCIn-Net can still accurately locate targets in both RD and RA views but faces challenges in differentiating target categories in the highly cluttered RA view. Nevertheless, AdaPKC $^\xi$ -Net<sup>*FiT*</sup> correctly classifies targets in both views. In the third frame, where the signal of the distant car target is weak and the clutter interference is strong, existing methods miss the car target whereas AdaPKC $^\xi$ -Net<sup>*FiT*</sup> can still identify it in the RD view. These results suggest that AdaPKC $^\xi$ -Net<sup>*FiT*</sup> exhibits stronger interference suppression and target recognition capabilities compared to PKCIn-Net and other RSS methods.

#### F.4 Semantic Segmentation Results by Category on KuRALS

Considering the limited samples of both *ship* and land *vehicle* classes, we merge the two classes into a single *vehicle* class to mitigate the negative impact of class-imbalance on model training. This class merging is primarily based on two observations: i) both ships and land vehicles belong to rigid transportation vehicles, whose reflected radar signals share considerable similarities; ii) ships and land vehicles are primarily used for marine monitoring tasks and land detection tasks, respectively, and there exist nearly no overlapping application scenarios between them, which minimizes the risk

Table S5: RSS performance comparison by category on the CARRADA benchmark.

View	Frameworks	#Params @Frames	IoU(%)					Dice(%)				
			Bkg.	Ped.	Cyc.	Car	mIoU	Bkg.	Ped.	Cyc.	Car	mDice
RD	FCN	134.3M@3	99.7	47.7	18.7	52.9	54.7	99.8	24.8	16.5	26.9	66.3
	U-Net	17.3M@3	99.7	51.0	33.4	37.7	55.4	99.8	67.5	50.0	54.7	68.0
	DeepLabv3+	59.3M@3	99.7	43.2	11.2	49.2	50.8	99.9	60.3	20.2	66.0	61.6
	RSS-Net	10.1M@3	99.3	0.1	4.1	25.0	32.1	99.7	0.2	7.9	40.0	36.9
	RAMP-CNN	106.4M@9	99.7	48.8	23.2	54.7	56.6	99.9	65.6	37.7	70.8	68.5
	TMVA-Net	5.6M@5	99.7	49.5	22.8	52.5	56.1	99.8	66.2	37.1	68.9	68.0
	TransRadar	4.9M@5	99.6	49.5	24.7	54.8	57.2	99.8	66.2	39.7	70.8	69.1
	PKCIn-Net	6.3M@5	99.7	54.0	30.4	58.5	60.7	99.8	70.1	46.7	73.9	72.6
	AdaPKC <sup>θ</sup> -Net(ours)	6.3M@5	99.7	54.6	33.7	58.1	61.5	99.8	70.6	50.4	73.5	73.6
	AdaPKC <sup>ξ</sup> -Net <sup>FIT</sup> (ours)	6.3M@5	99.7	56.3	32.8	59.6	62.1	99.9	72.0	49.4	74.7	74.0
RA	FCN	134.3M@3	99.8	14.8	0.0	23.3	34.5	99.9	25.8	0.0	37.8	40.9
	U-Net	17.3M@3	99.8	22.4	8.8	0.0	32.8	99.9	36.6	16.1	0.0	38.2
	DeepLabv3+	59.3M@3	99.9	3.4	5.9	21.8	32.7	99.9	6.5	11.1	35.7	38.3
	RSS-Net	10.1M@3	99.5	7.3	5.6	15.8	32.1	99.8	13.7	10.5	27.4	37.8
	RAMP-CNN	106.4M@9	99.8	1.7	2.6	7.2	27.9	99.9	3.4	5.1	13.5	30.5
	TMVA-Net	5.6M@5	99.8	25.6	7.5	17.8	37.7	99.9	40.7	13.9	30.2	46.2
	TransRadar	4.9M@5	99.7	20.7	11.1	28.2	39.9	99.8	34.3	20.0	44.0	49.5
	T-RODNet	162.0M@4	99.9	25.4	9.5	39.4	43.5	99.9	40.5	17.4	56.6	53.6
	PKCIn-Net	6.3M@5	99.8	24.0	14.8	33.7	43.1	99.9	38.7	25.8	50.4	53.7
	AdaPKC <sup>θ</sup> -Net(ours)	6.3M@5	99.8	26.5	15.8	32.3	43.6	99.9	41.9	27.3	48.9	54.5
	AdaPKC <sup>ξ</sup> -Net <sup>FIT</sup> (ours)	6.3M@5	99.8	26.2	19.2	32.0	44.3	99.9	41.5	32.2	48.4	55.5

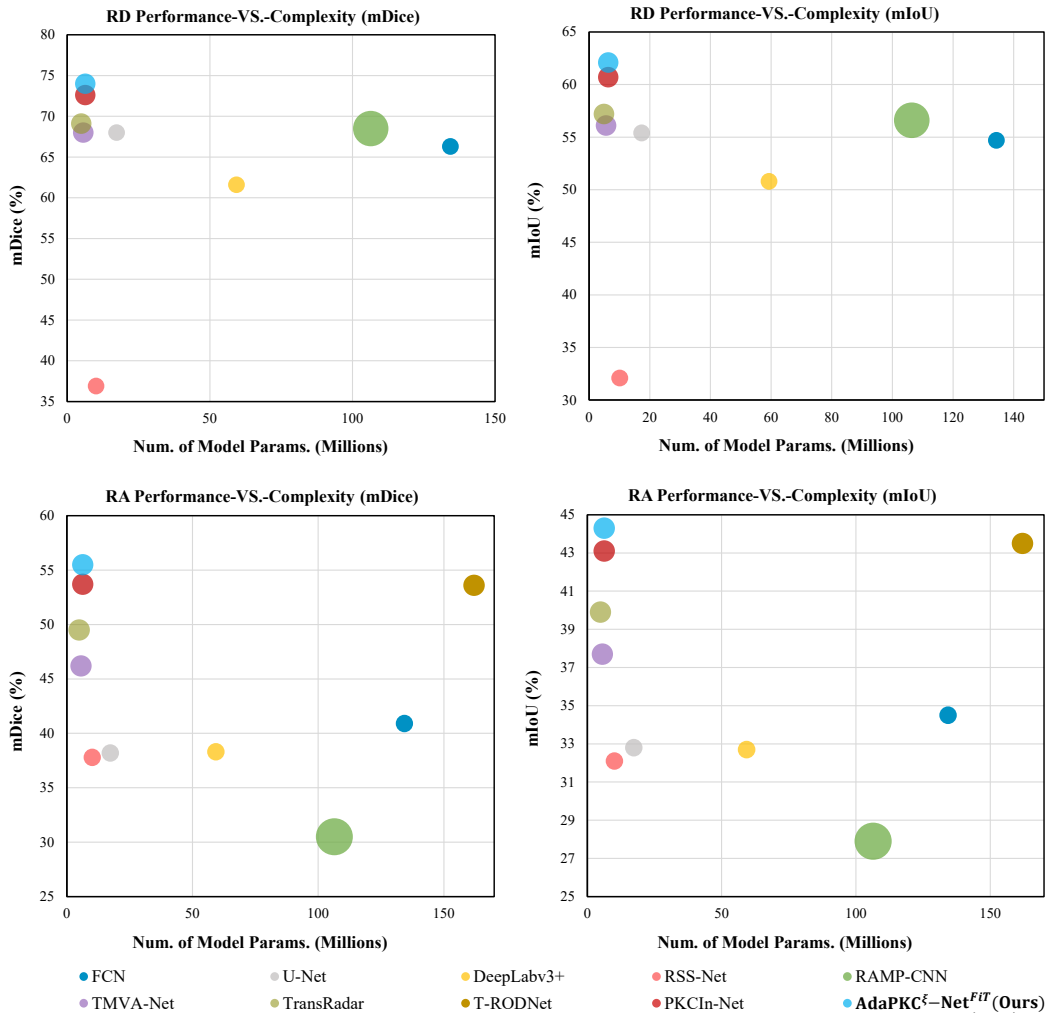


Figure S6: Performance vs. Complexity.

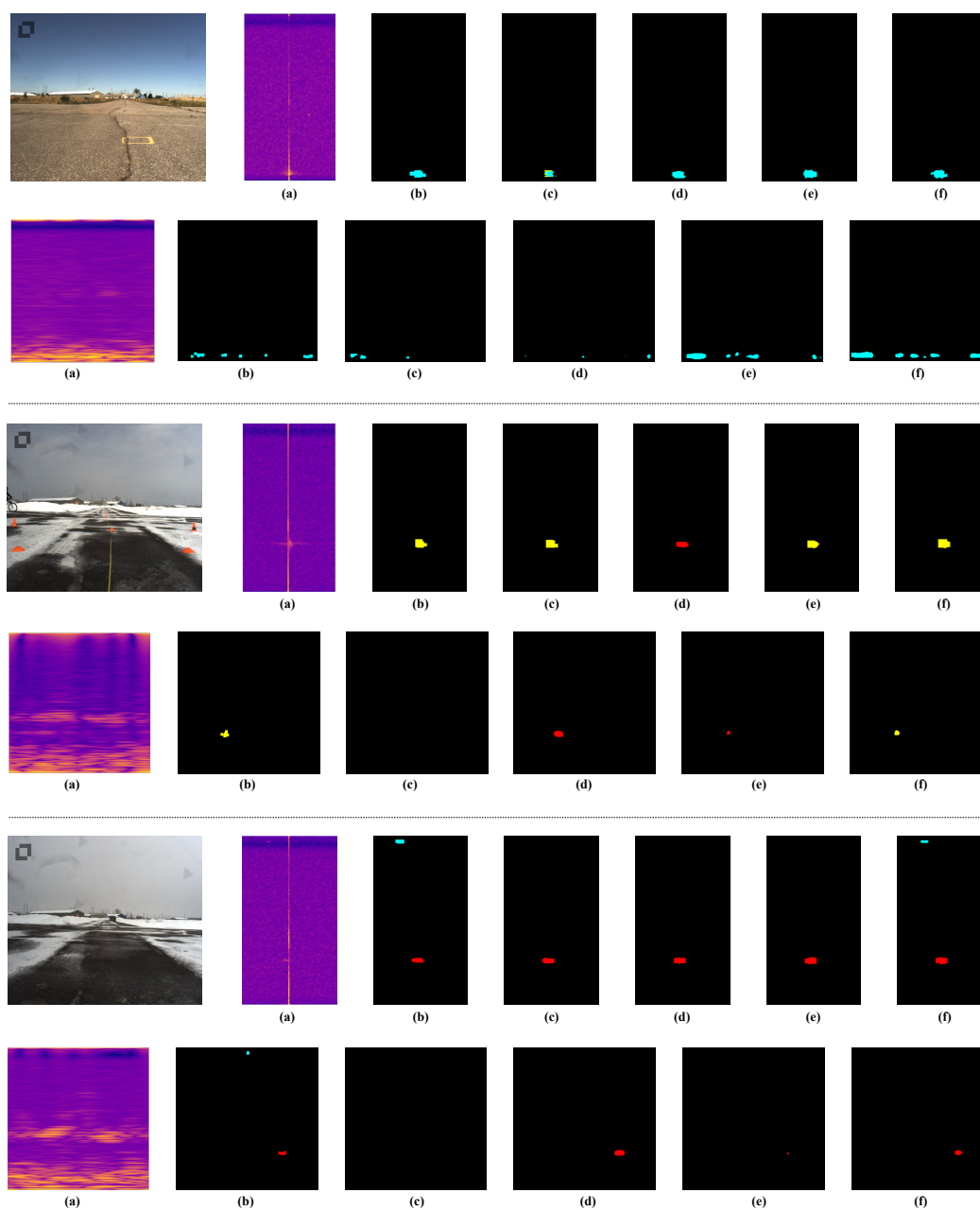


Figure S7: Qualitative comparison of different methods on CARRADA. The three frames from CARRADA test split are with varying levels of interference on targets. For each frame, the *top* row shows the camera image and the results in RD view, while the *bottom* row shows the results in RA view. (a) RD/RA tensor, (b) the label of mask, (c) TMVA-Net, (d) TransRadar, (e) PKCIn-Net and (f) AdaPKC $^{\xi}$ -Net $^{FiT}$  (ours). Different colors represent different object categories. Black: background, Red: pedestrian, Yellow: cyclist, Cyan: car.



of practical application issues due to recognition confusion. As a result, we perform the segmentation task with four categories: *background*, *UAV*, *pedestrian* and *vehicle*. The semantic segmentation results by category on the KuRALS-Test dataset are presented in Tab. S6.

Table S6: RSS performance comparison by category on the KuRALS-Test dataset.

Frameworks	#Params @Frames	IoU(%)					Dice(%)				
		Bkg.	UAV	Ped.	Veh.	mIoU	Bkg.	UAV	Ped.	Veh.	mDice
FCN	134.3M@5	99.9	68.1	21.9	11.5	50.4	99.9	81.0	35.9	20.6	59.4
U-Net	17.3M@5	99.9	74.8	30.9	4.0	52.4	99.9	85.6	47.2	7.7	60.1
DeepLabv3+	59.3M@5	99.9	73.1	23.0	14.4	52.6	99.9	84.5	37.4	25.2	61.8
KuRALS-Net(ours-baseline)	1.2M@5	99.9	72.8	<b>39.0</b>	12.2	56.0	99.9	84.3	<b>56.1</b>	21.8	65.5
KuRALS-Net w/ PKC(ours-baseline)	1.2M@5	99.9	<u>79.7</u>	31.9	15.3	56.7	99.9	<u>88.7</u>	48.4	26.5	<u>65.9</u>
KuRALS-Net w/ AdaPKC <sup><math>\theta</math></sup> (ours)	1.2M@5	99.9	78.4	<u>33.2</u>	<u>19.4</u>	<u>57.8</u>	99.9	87.9	<u>49.9</u>	<u>32.6</u>	<b>67.6</b>
KuRALS-Net w/ AdaPKC <sup><math>\xi</math> FiT</sup> (ours)	1.2M@5	99.9	<b>81.5</b>	31.6	<b>19.6</b>	<b>58.2</b>	99.9	<b>89.8</b>	48.0	<b>32.7</b>	<b>67.6</b>

## F.5 Qualitative Comparison on KuRALS

The qualitative results on KuRALS dataset are illustrated in Fig. S8. Compared to our baseline methods and other segmentation models, our AdaPKCs demonstrate more accurate localization and classification of targets.

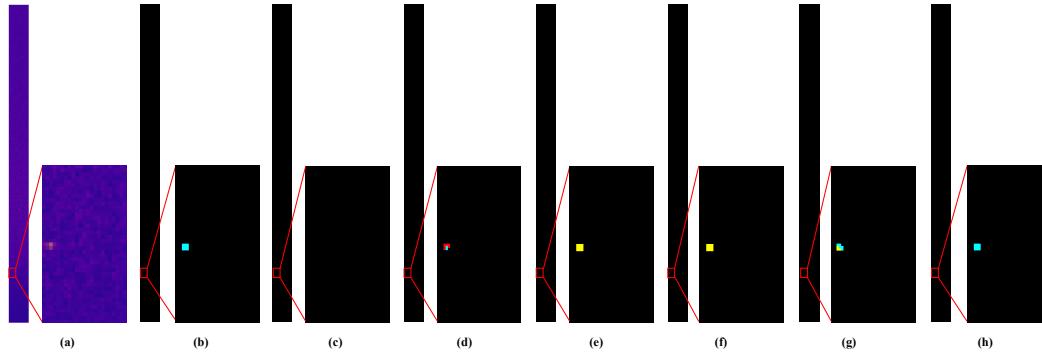


Figure S8: Qualitative comparison on KuRALS. (a) RD tensor, (b) the label of mask, (c) FCN, (d) U-Net, (e) DeepLabv3+, (f) KuRALS-Net (ours-baseline), (g) KuRALS-Net w/ PKC (ours-baseline), (h) KuRALS-Net w/ AdaPKC <sup>$\xi$  FiT</sup>(ours). Different colors represent different object categories. Black: background, Red: UAV, Yellow: Pedestrian, Cyan: Vehicle.

## F.6 More Results in Ablation Study for FiTOS

Fig. S9 illustrates the segmentation results of mIoU for AdaPKC <sup>$\xi$</sup> -Net<sup>FiT</sup> with different values of  $\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ .

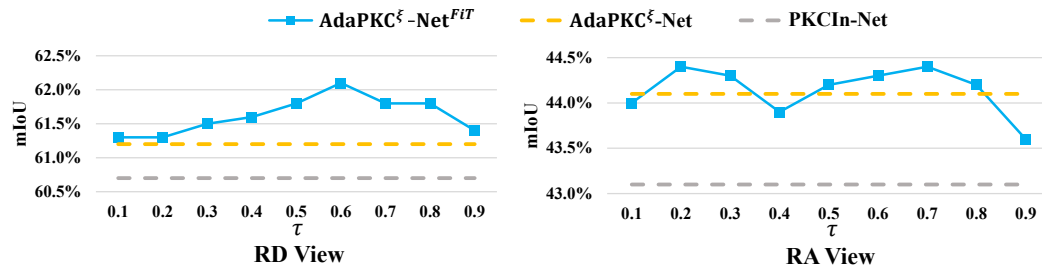


Figure S9: RSS performance of mIoU for AdaPKC <sup>$\xi$</sup> -Net<sup>FiT</sup> with different values of  $\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ .

## G Broader Discussions

**Limitations.** We have collected a Ku-band continuous wave radar dataset to validate the effectiveness of our proposed method in surveillance radar detection scenarios. Nevertheless, pulse-Doppler radar is also commonly used in these scenarios and investigation on such radar datasets would further uncover the potentials and issues of our methods in practical applications. However, there is currently a lack of publicly available datasets for pulse-Doppler monitoring radar. To alleviate this issue, we will try to collect and release a pulse-Doppler radar dataset to enable more comprehensive validation and analysis of our methods and other works.

**Societal Impacts.** Our approach is applicable to various practical applications, such as perceptions for autonomous driving, UAV surveillance and marine monitoring. However, inappropriate usage may lead to decreased reliability, potentially resulting in safety and other issues.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: This paper discusses the limitations of our work in Section G.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: For each theoretical result, this paper provides the full set of assumptions and a complete (and correct) proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: This paper fully discloses all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

#### 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: In abstract, this paper provides an link to the data and code, with sufficient instructions to faithfully reproduce the main experimental results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: This paper specifies all the training and test details in Section 4.1 and Section D.2.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: This paper reports error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: This paper provides sufficient information on the computer resources needed to reproduce the experiments in Section D.2.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [\[Yes\]](#)

Justification: The research conducted in the paper conforms, in every respect, with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

## 10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#)

Justification: The paper discuss the potential societal impacts of our work in Section G.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

#### 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The creators or original owners of assets, used in the paper, are properly credited and the license and terms of use are explicitly mentioned and properly respected.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.



- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [\[Yes\]](#)

Justification: New assets introduced in the paper are well documented and the documentation is provided alongside the assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [\[NA\]](#)

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [\[NA\]](#)

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.