Appendix

A. Ablations on the transferability of the proposed AKE module.

We take a step forward to examine the transferability of the AKE module between different architectures. Specifically, we directly apply the AKE module trained with one teacher on the intermediate features of other teachers to distill RetinaNet with ResNet 18 student. For example, the second column in the first row of Table 1 indicates that the AKE modules are first finetuned on ATSS and then used on RetinaNet directly to distill the student. In Table 1, we choose RetinaNet, ATSS, Faster R-CNN, and FCOS with ResNet 50 as the teachers. It can be observed that using shared decoder can still improve the student significantly, especially on Faster R-CNN teacher where comparable improvement is achieved compared with using decoder that is specifically trained. This demonstrates that there are indeed some general knowledge across different architectures and our UniKD can effectively extract it to help the student model.

AKE applied on (Teacher)	AKE trained on			
	RetinaNet	ATSS	Faster R-CNN	FCOS
RetinaNet	**35.3**	34.7	34.1	33.9
ATSS	32.5	**34.4**	32.2	32.8
Faster R-CNN	34.6	34.1	**34.8**	34.4
FCOS	33.6	34.0	33.1	**34.7**

Table 1: Ablation on the transferability of the AKE module.

B. More visualizations of the attention of the queries. We provide more visualizations to show the location and weight of the attention for two types of queries in Figure 1. For traditional detectors, we can see that the queries learn to gather information from the salient and marginal parts of objects to fulfill the detection task, and the two types of queries are complementary to each other. It is worth noting that the spatial relations between feature maps and the input image are not preserved in DETR-based detectors. The visualization of feature density also proves that. So the learned offsets show a different pattern from traditional detectors, causing inefficiency in pixel-to-pixel feature imitation. In UniKD, we introduce deformable transformer encoders as the adaptor to learn this spatial transformation, which assures that the knowledge absorbed by AKE modules can be effectively propagated to the student.

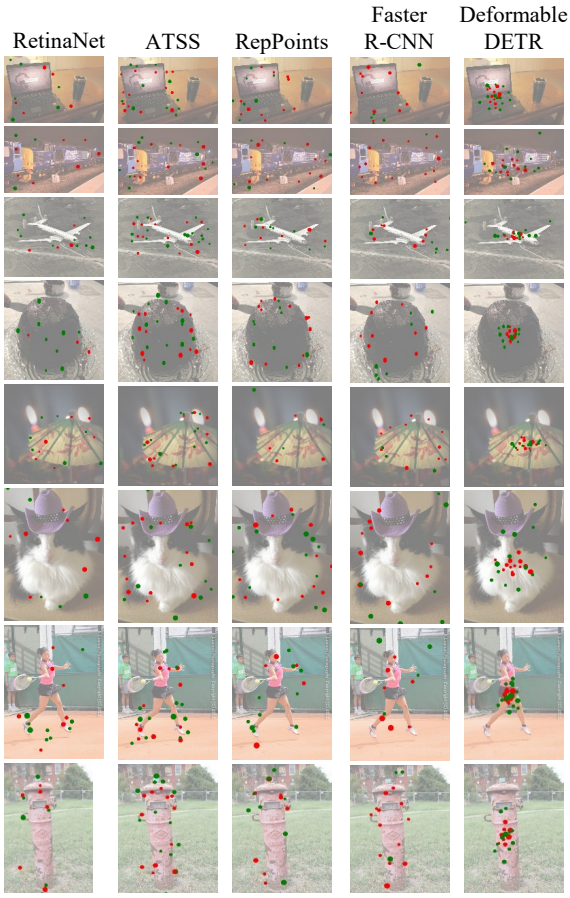

Figure 1: More visualizations of the attention of the queries. The content knowledge queries are denoted in red dots and the positional knowledge queries are in green dots. The size of the dots indicates the attention weights.