Supplementary Material:
Exploring Inter-Channel Correlation for Diversity-preserved Knowledge Distillation

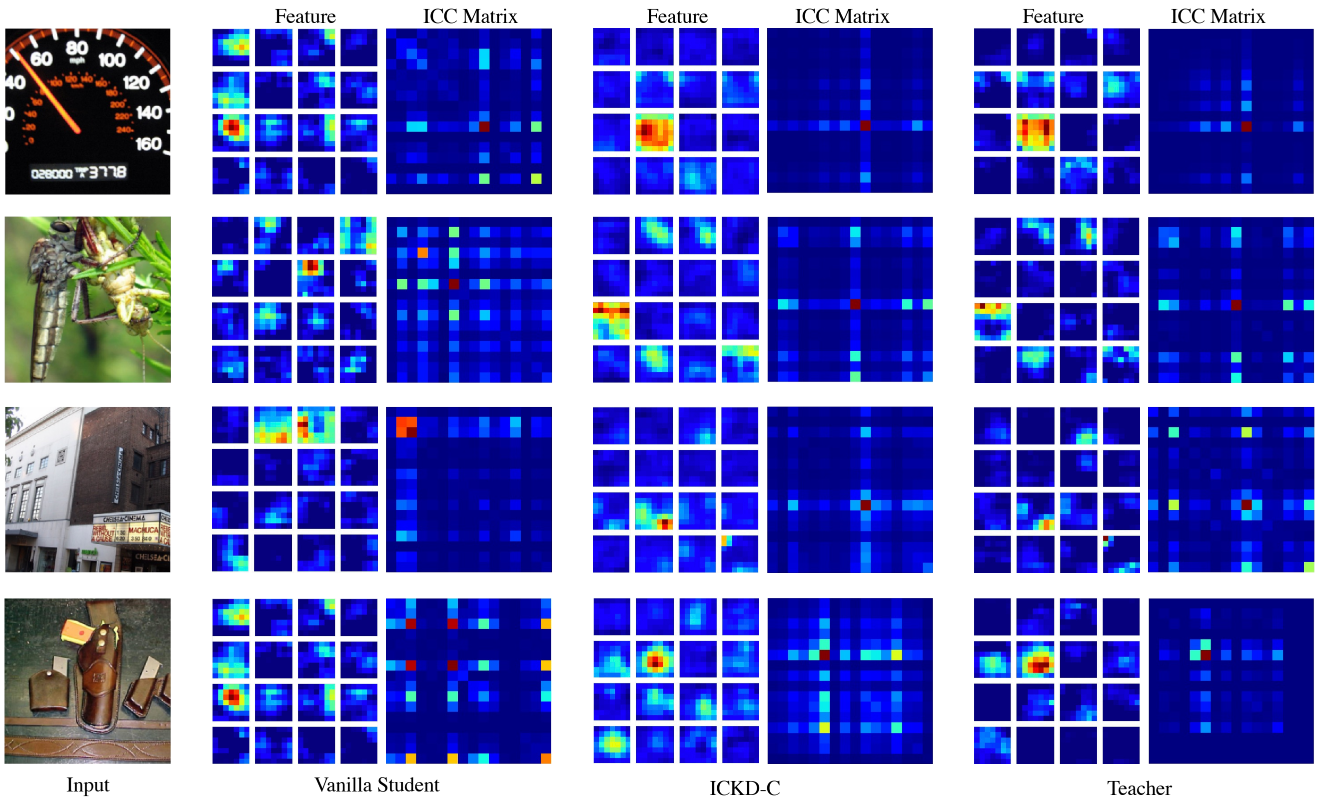
Li Liu[1] Qingle Huang[1] Sihao Lin[2] Hongwei Xie[1] Bing Wang[1] Xiaojun Chang[3] Xiaodan Liang[4]
[1]Alibaba Group [2]Monash University [3]RMIT University [4]Sun Yat-sen University
{liuli.ll9412, qingle.hql, linsihao6, hongwei.xie.90, xdliang328}@gmail.com
fengquan.wb@alibaba-inc.com, xiaojun.chang@rmit.edu.au

1. Visualization of Features and ICC Matrices

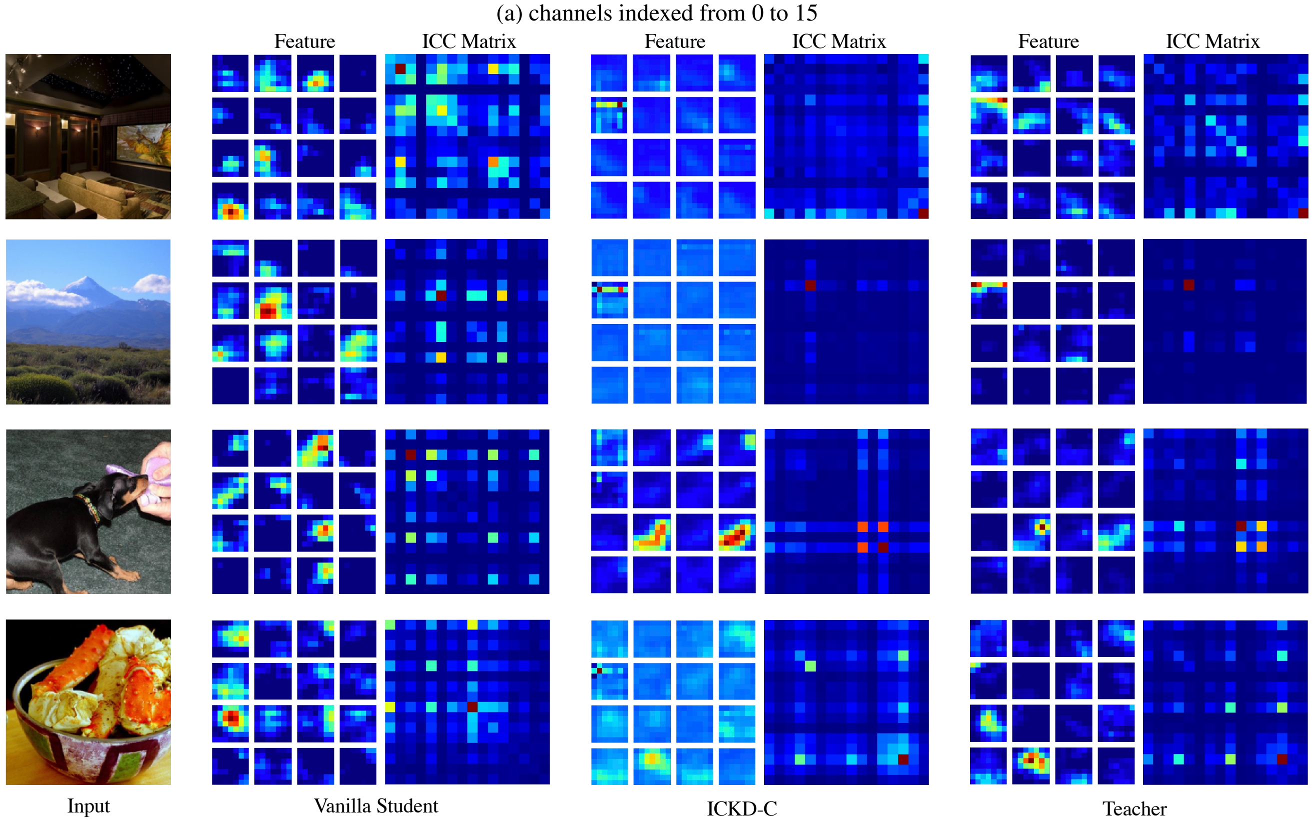
In the main text we have quantitatively evaluated our ICKD-C on Cifar-100 and ImageNet. We also have qualitatively compared our model with the vanilla student by the visualization of features and ICC matrices. The result is impressive since both the feature pattern and ICC matrix are fairly similar with those of the teacher model, demonstrating that transferring the inter-channel correlation can help the student align with the feature space of teacher. Here we show more results on ImageNet testing set (See Fig. 1). In our experimental setting, the teacher network is ResNet34 and student network is ResNet18. The 4-th block of the teacher and student are adopted as distillation layer, which has 64 channels in total. For detailed visualization, we *orderly* select 16 channels of each sample. In Fig. 1 (a), the channels indexed from 0 to 15 are selected for visualization, while the channels indexed from 16 to 31 are selected for visualization in Fig. 1 (b). As shown by the results, our ICKD-C, compared to the vanilla student, has more similar ICC matrix and feature pattern with the teacher.

2. Visualization of Segmentation

In the main text our ICKD-S is quantitatively compared to the vanilla student and is able to improve the small student model by a large margin. Here we show some segmentation result for a perceptual insight. The teacher backbone is ResNet101 and student backbone is ResNet18. As shown in Fig. 2, our ICKD-S is superior to vanilla student especially when there exists complex background or overlap between two objects. According to the segmentation results, our ICKD-S can reduce the errors that misclassify the background into objects (2nd and 3rd row in Fig. 2). When it comes to overlap between two objects, ICKD-S can effectively perceive the spatial context to give a more complete prediction(4th row in Fig. 2).



(a) channels indexed from 0 to 15



(b) channels indexed from 16 to 31

Figure 1: **Visualization of feature and ICC matrix.** Each row corresponds to a sample. We *orderly* select 16 channels of each sample for better visualization. (a) Channels indexed from 0 to 15 are selected. (b) Channels indexed from 16 to 31 are selected.

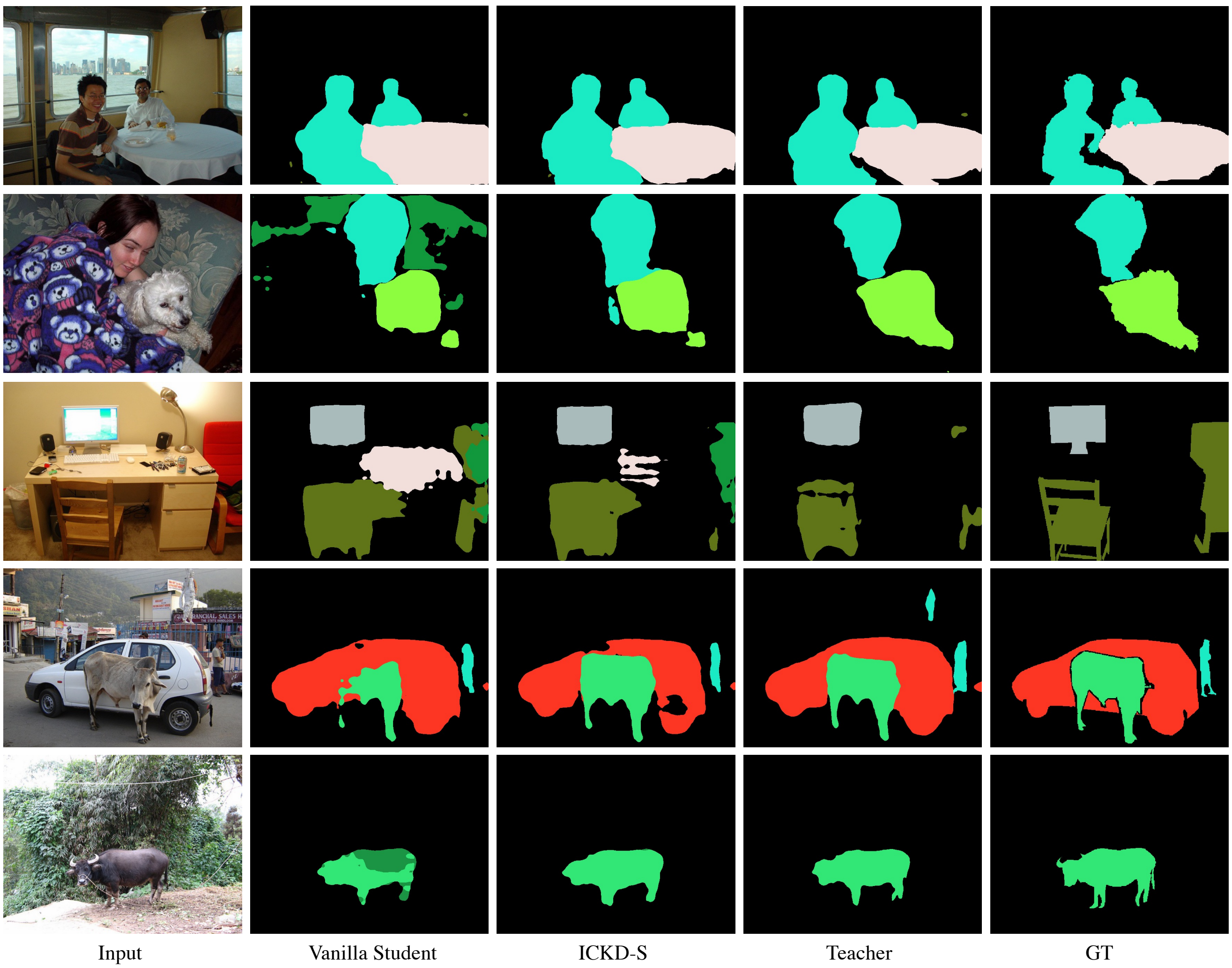


Figure 2: **Semantic segmentation results on Pascal VOC validation set.** Compared to the vanilla student, our ICKD-S can prevent from misclassifying the background (2nd, 3rd row) and in terms of overlap between objects, our model can give a more complete prediction (4th row).