DENSO IT LAB

Effectiveness of Function Matching in Driving Scene Recognition

2022/10/23　The 3rd AVVision Workshop

Shingo Yashima (Denso IT Laboratory, inc.)

Learning Compact DNNs for Driving Scene Recognition

Purpose Obtaining a lightweight and well-preforming recognition model for autonomous driving (e.g., segmentation, detection)

Issue The trained model sometimes does not reach the desired performance

Cause 1. Not enough model capacity → Need to change the architecture

Learning Compact DNNs for Driving Scene Recognition

Purpose Obtaining a lightweight and well-preforming recognition model for autonomous driving (e.g., segmentation, detection)

Issue The trained model sometimes does not reach the desired performance

Cause
1. Not enough model capacity → Need to change the architecture
2. Enough model capacity, but not optimized sufficiently

Compact models are often harder to generalize than large models,
even though they may have the capacity to represent solutions of the large models

Learning Compact DNNs for Driving Scene Recognition

Purpose Obtaining a lightweight and well-preforming recognition model for autonomous driving (e.g., segmentation, detection)

Issue The trained model sometimes does not reach the desired performance

Cause
1. Not enough model capacity → Need to change the architecture
2. Enough model capacity, but not optimized sufficiently

Method to exploit full generalization power of compact models ?

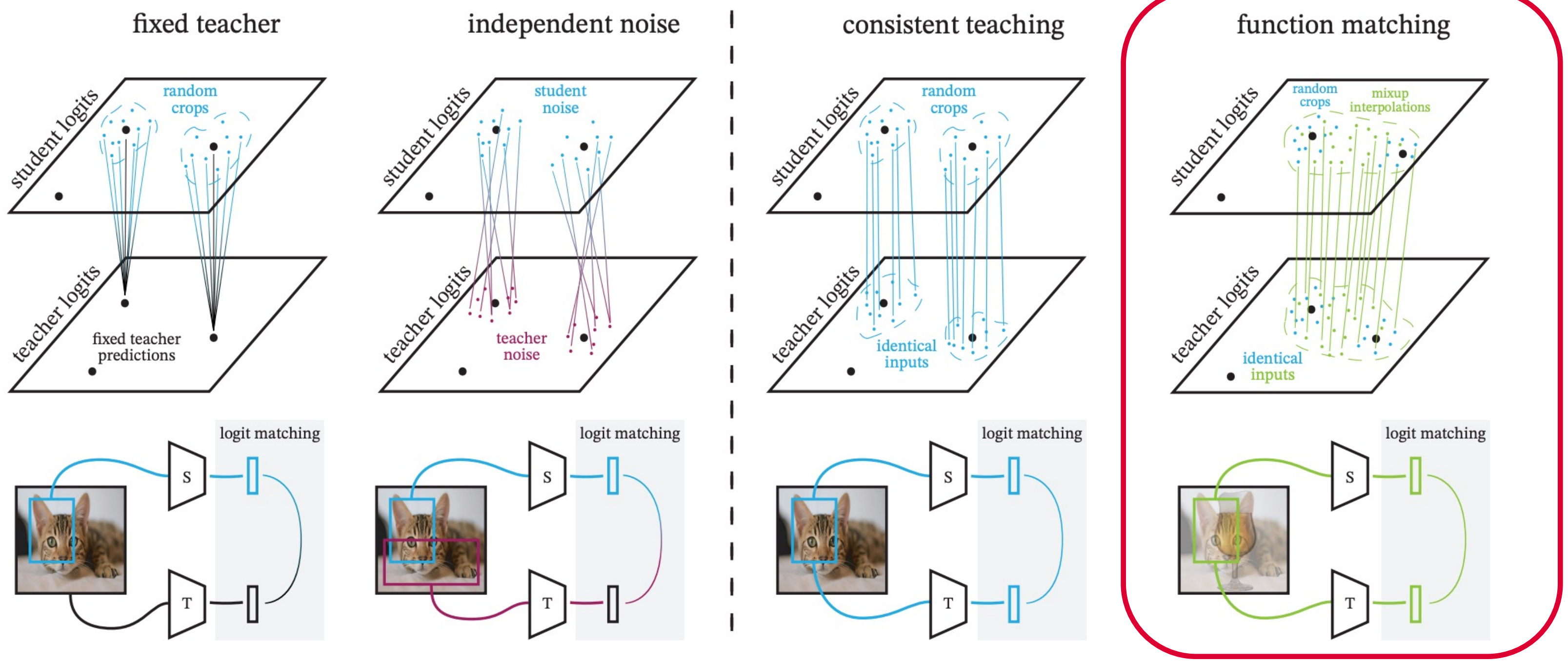
→ **Knowledge Distillaion**

Knowledge Distillation as Function Matching

Recent study showed critical components of distillation to improve performance [1] :

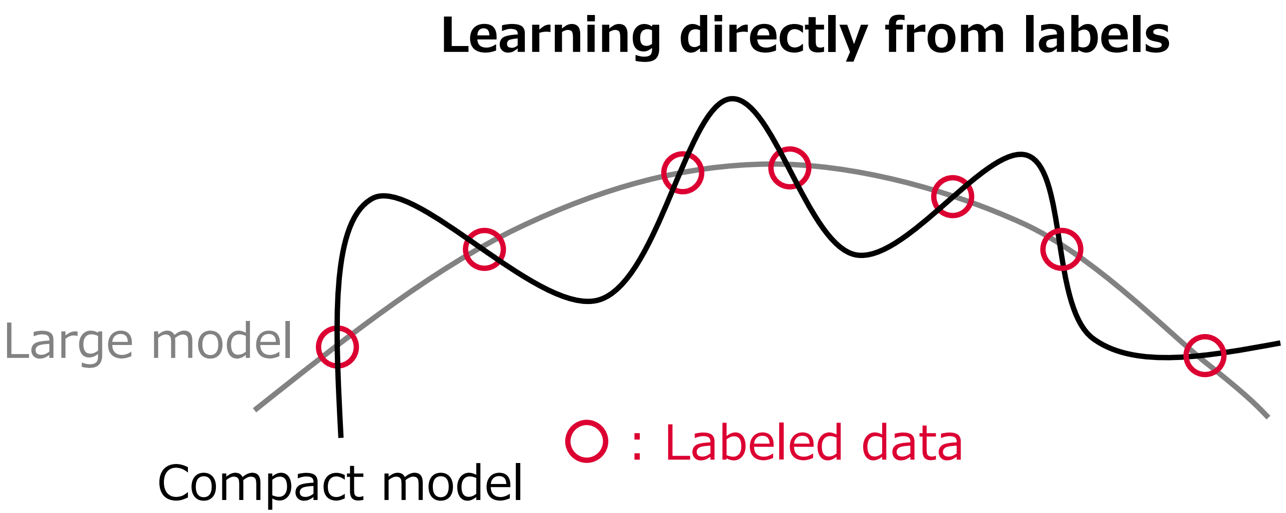
1. Consistency of teacher's and student's inputs
2. Distilling on a **wide range of data points** for large number of epochs

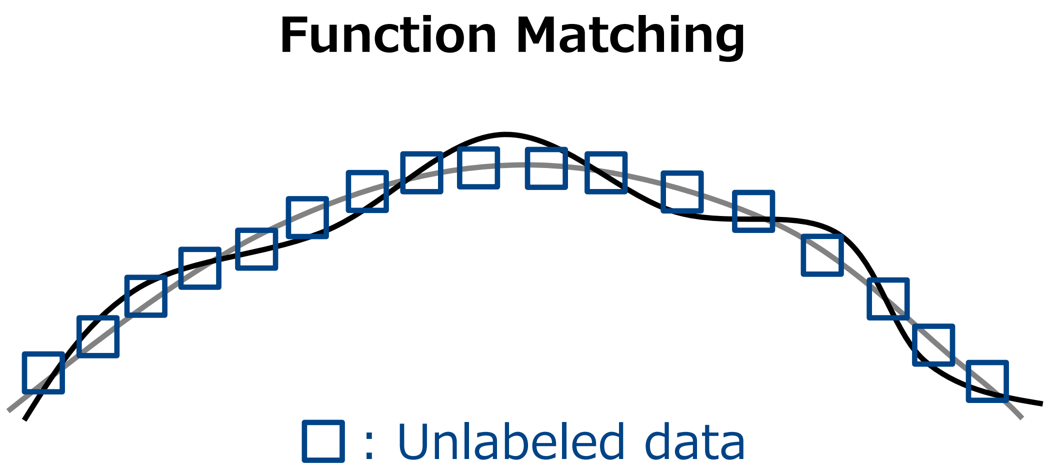
→ **Imitating teacher as a function is critical (function matching)**



[1] L Beyer et al., "Knowledge Distillation: A Good Teacher Is Patient and Consistent." *CVPR2022*.

Function Matching on Driving Scene Recognition





Compact models are hard to generalize when directly learned from labels

Imitating the large model on a large number of points yields generalization

Generally, we can access an almost **infinite amount of** (unlabeled) driving scene data

How effective is using large amounts of unlabeled scene data for distillation?

Experiments

Evaluation on semantic segmentation (SS) and object detection (OD) in BDD100K

Datasets

SS Labeled: 10,000 images
OD Labeled: 100,000 images
Unlabeled: 16,000,000 images (clipped from original videos)

Learning Methods

① Train the compact model directly from the **labeled** dataset (**Supervised**)

② Train the large model from the **labeled** dataset
→ Distill it to the compact model with the **labeled** dataset (**Distill**)

③ Train the large model from the **labeled** dataset
→ Distill it to the compact model with the **unlabeled** dataset (**FunMatch**)

Results（Semantic Segmentation）

Model Large: PSPNet with ResNet-101 backbone

Compact: PSPNet with ResNet-18 backbone

Results

	Method	Training Data	mIoU
Large (teacher)	Supervised	Labeled (70K)	64.83
Compact ①	Supervised	Labeled (70K)	61.48
Compact ②	Distill	Labeled (70K枚)	61.70
Compact ③	FunMatch	Labeled (70K) + Unlabeled (16M)	64.74

(picked the best performing model among all training schedules)

The peformance of the compact model is improved dramatically
by function matching with a large number of unlabeled data

Results（Object Detection）

Model　　Large:　　FCOS with ResNet-101 + FPN backbone

　　　　Compact:　FCOS with ResNet-18 + FPN backbone

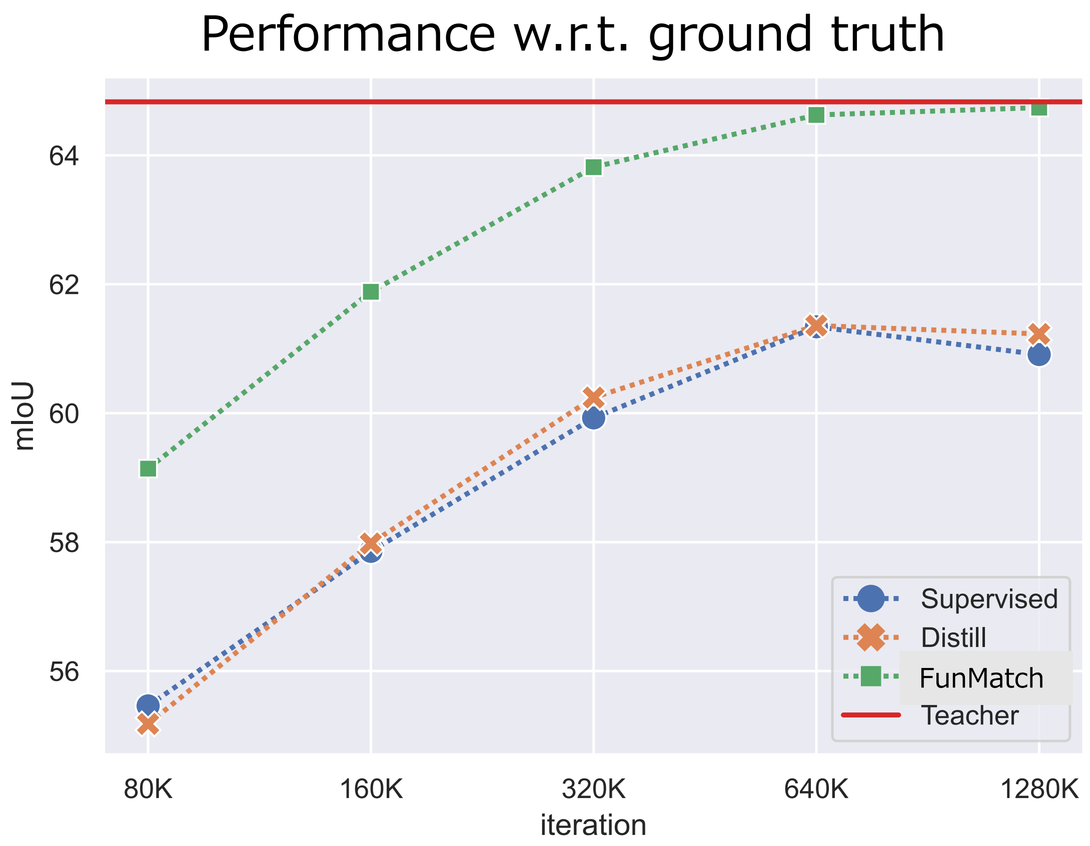
Results

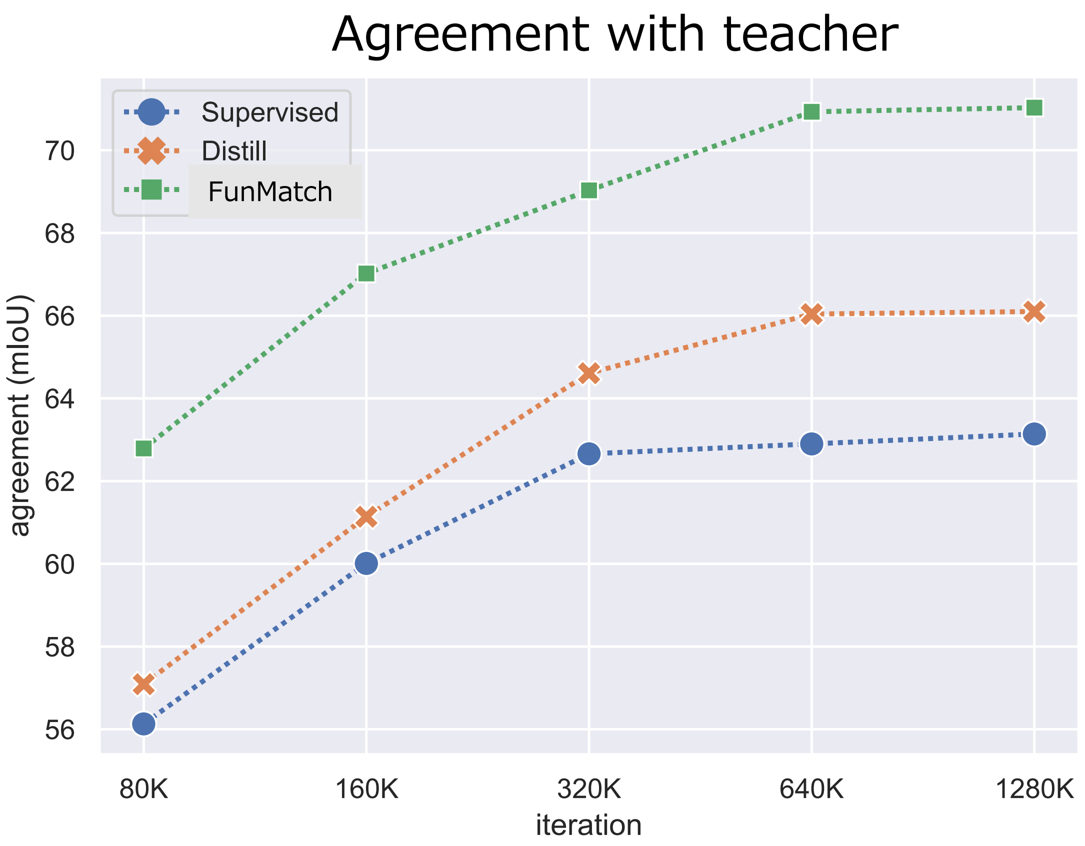
	Method	Training Data	mAP
Large (teacher)	Supervised	Labeled (70K)	31.4
Compact ①	Supervised	Labeled (70K)	29.5
Compact ②	Distill	Labeled (70K枚)	30.4
Compact ③	FunMatch	Labeled (70K) + Unlabeled (16M)	31.2

(picked the best performing model among all training schedules)

The peformance of the compact model is improved dramatically by function matching with a large number of unlabeled data

Results（Agreement with teacher）





Use of unlabeled data and longer training yields better agreement with teacher, resulting in better performance

DENSO IT LAB