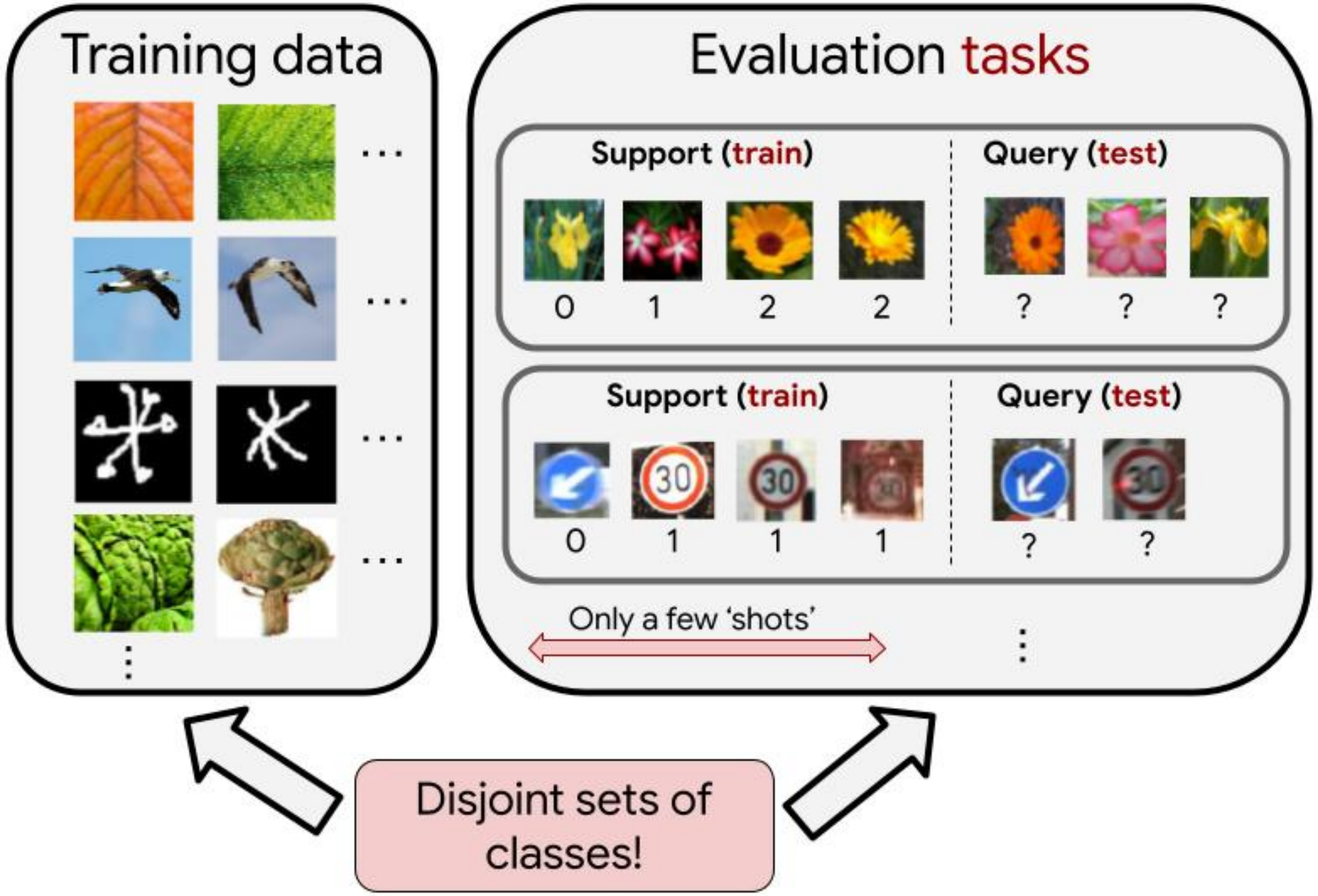
Learning a Universal Template for Few-shot Dataset Generalization

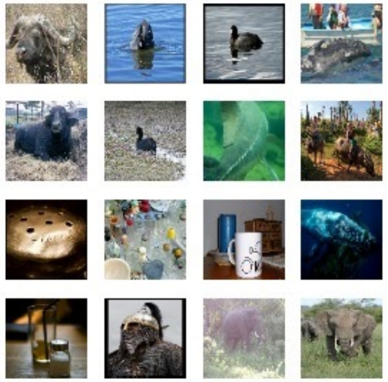
Eleni Triantafillou, Hugo Larochelle, Richard Zemel, Vincent Dumoulin

Few-shot classification

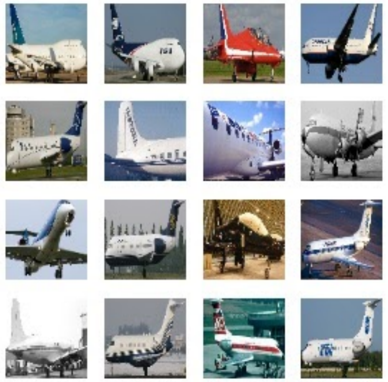
Goal: leverage a training dataset to create a model that can then *learn new classes* from few examples

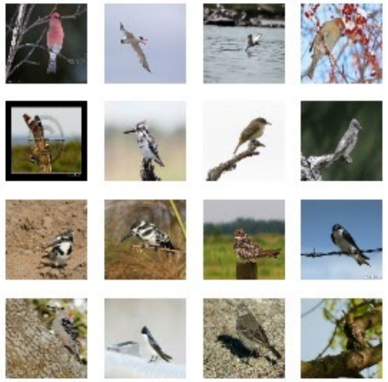


The Meta-Dataset benchmark

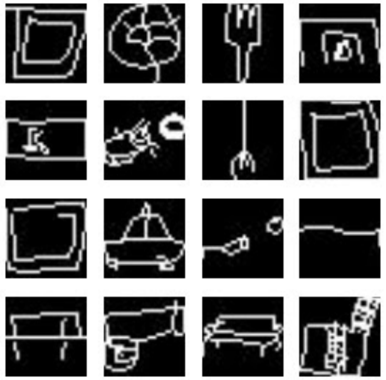

ImageNet


Omniglot


Aircraft


Birds


Textures


Quickdraw


Fungi


VGG Flower

Following Requeima et al, we add 3 additional held-out datasets for evaluation:

CIFAR-10, CIFAR-100, MNIST

Triantafillou, Zhu, Dumoulin, Lamblin, Evci, Xu, Goroshin, Gelada, Swersky, Manzagol, and Larochelle, ICLR 2020.
Requeima, Gordon, Bronskill, Nowozin, Turner. Fast and Flexible Multi-Task Classification Using Conditional Neural Adaptive Processes. NeurIPS 2019.

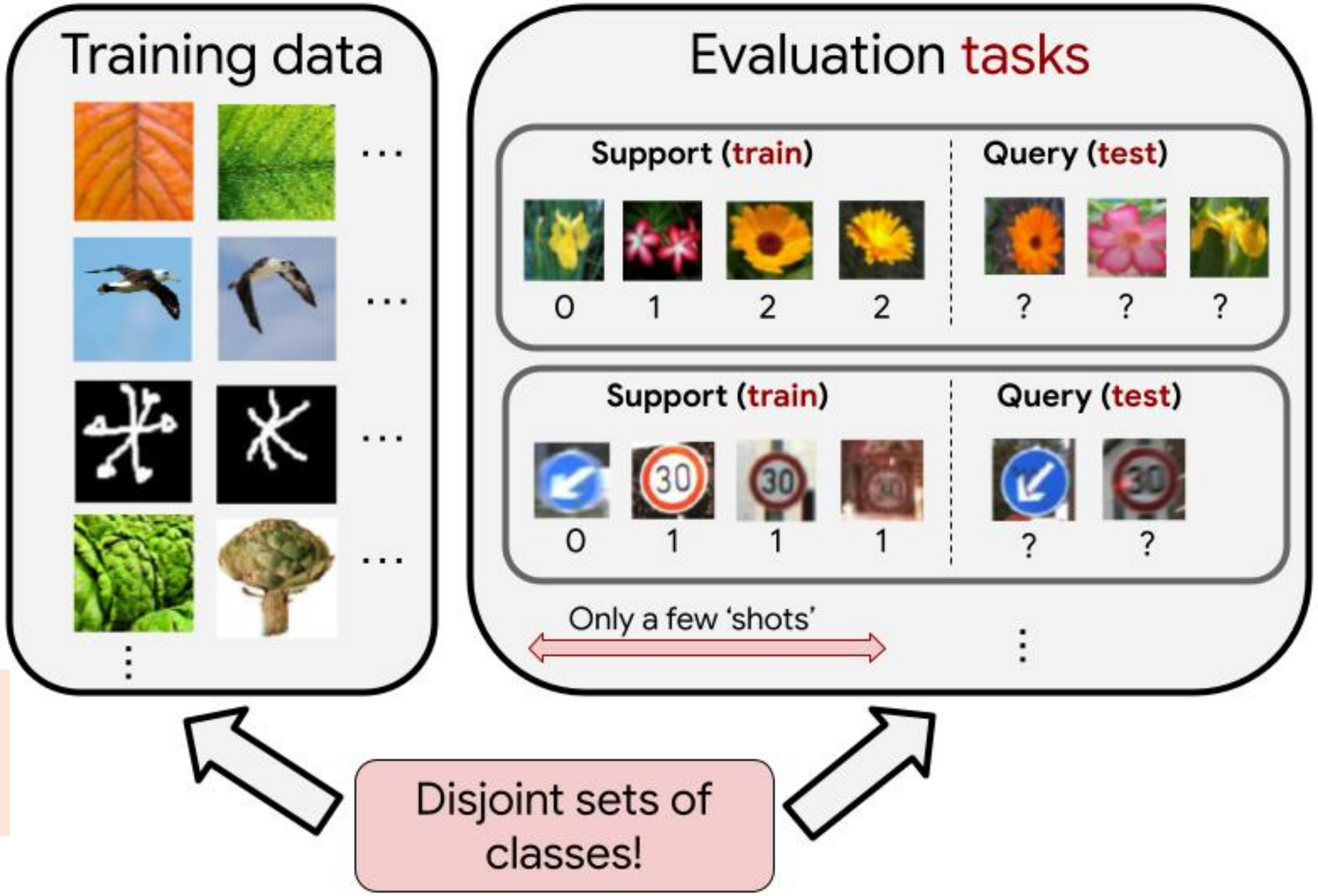
Few-shot classification on Meta-Dataset

Two types of *test tasks*:

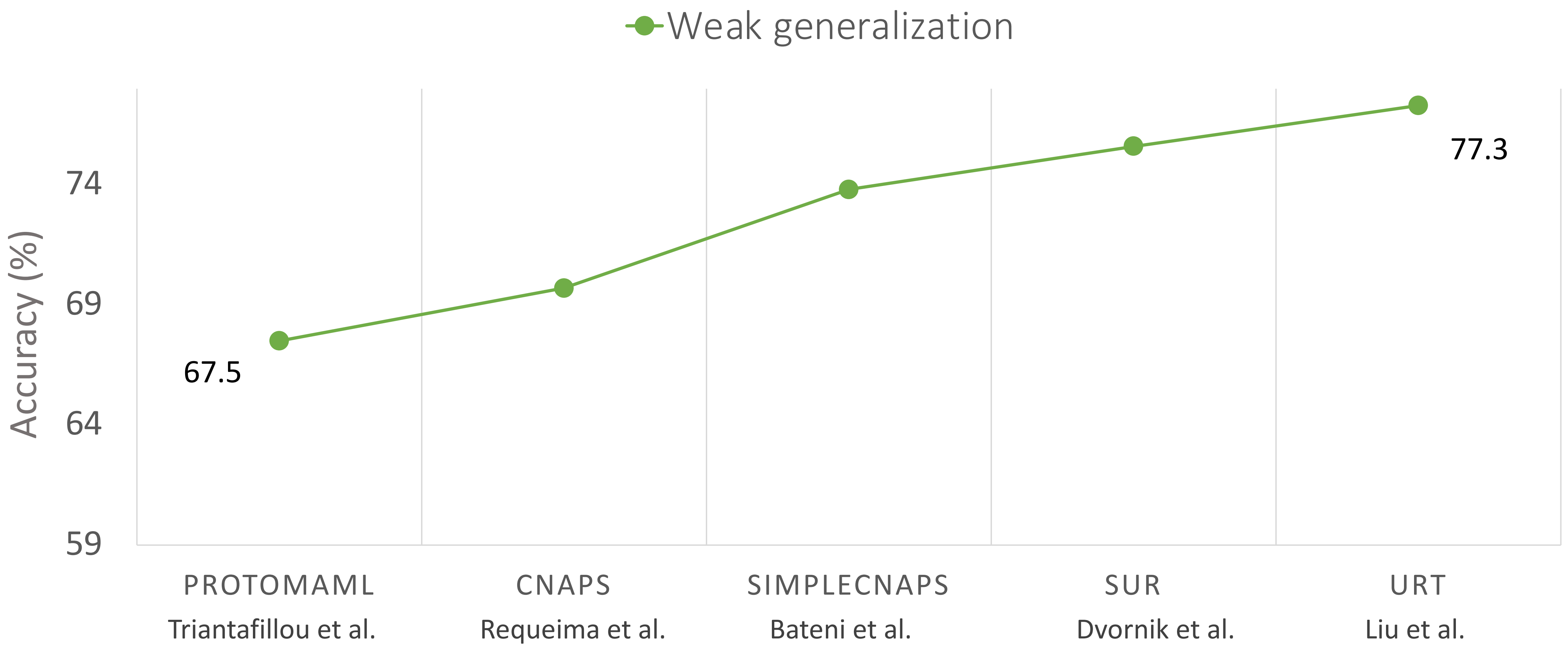
1. Weak generalization (WG):
   Sampled from held-out classes of a *training dataset*

2. Strong generalization (SG):
   Sampled from a held-out *dataset*

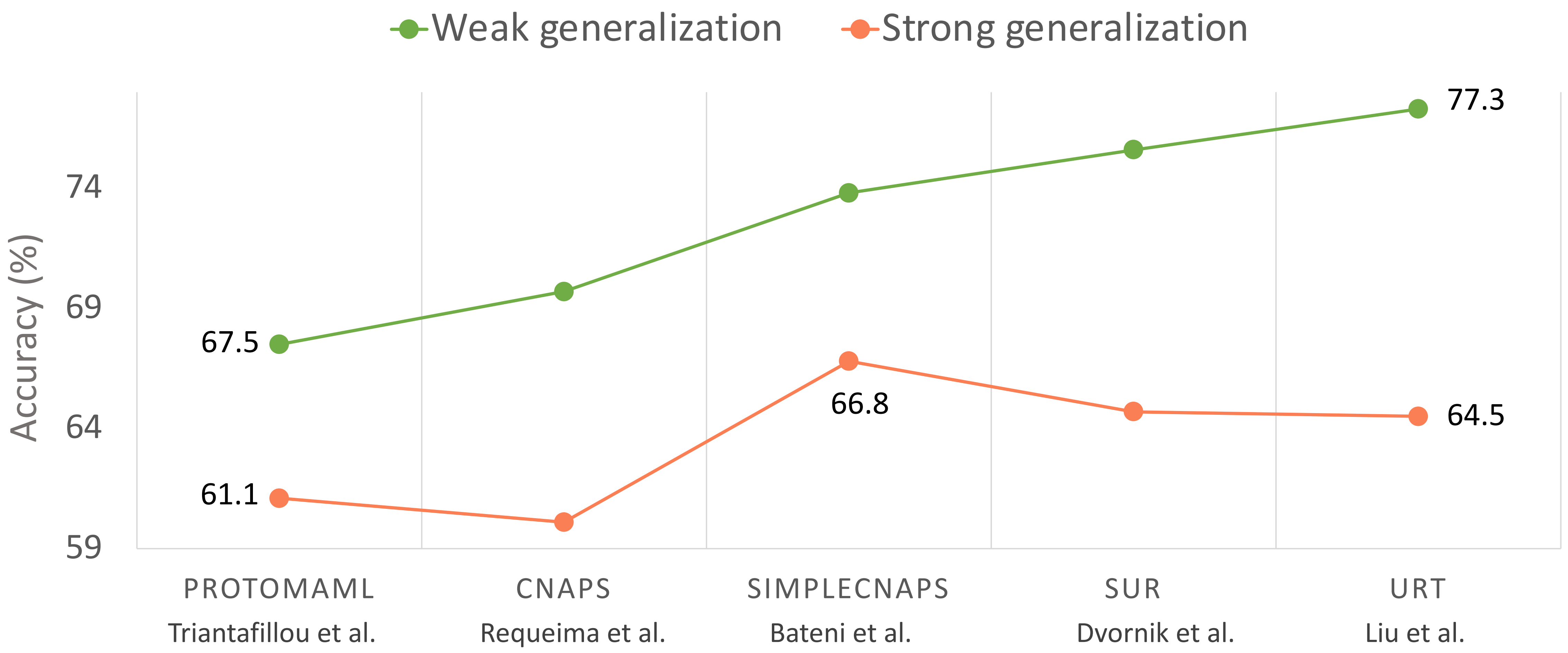
‘Few-shot Dataset Generalization’



Exciting progress on Meta-Dataset’s WG tasks



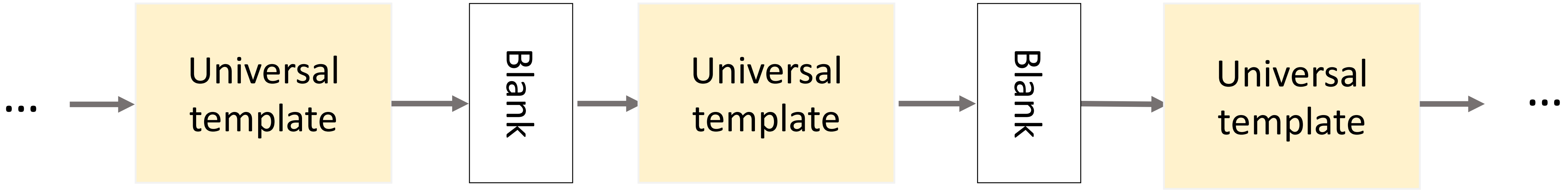
We aim to improve on the harder SG tasks



Learning a *universal template*

- A partially-parameterized model that can define a wide array of feature extractors by appropriately ‘filling in the blanks’

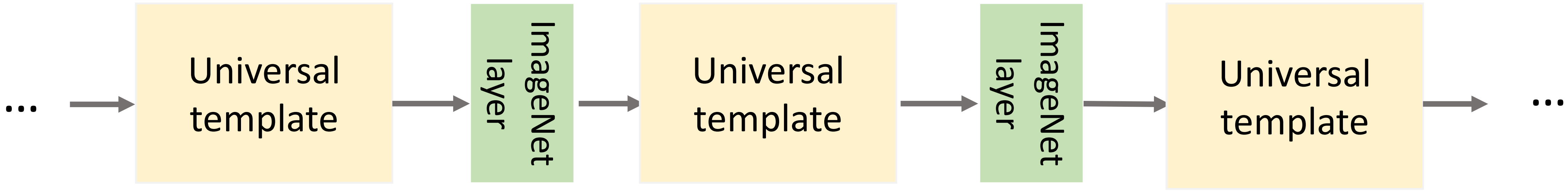
Partial feature extractor



Learning a *universal template*

- A partially-parameterized model that can define a wide array of feature extractors by appropriately ‘filling in the blanks’

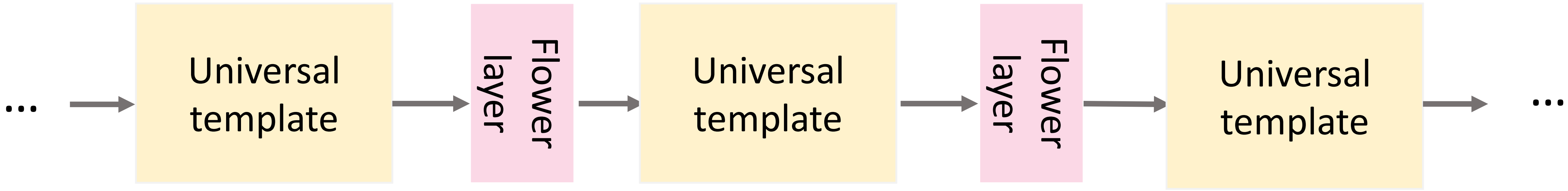
ImageNet’s feature extractor



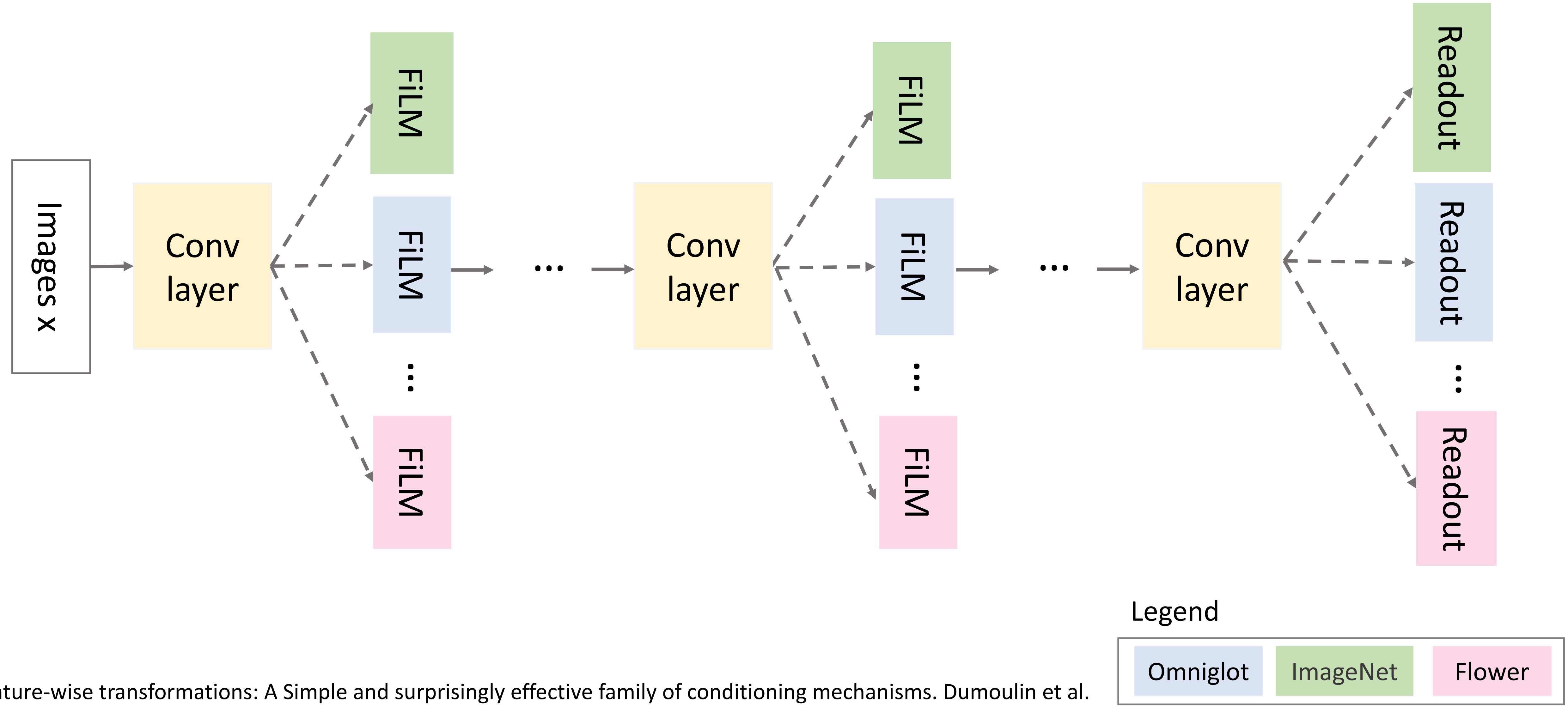
Learning a *universal template*

- A partially-parameterized model that can define a wide array of feature extractors by appropriately ‘filling in the blanks’

Flowers’ feature extractor

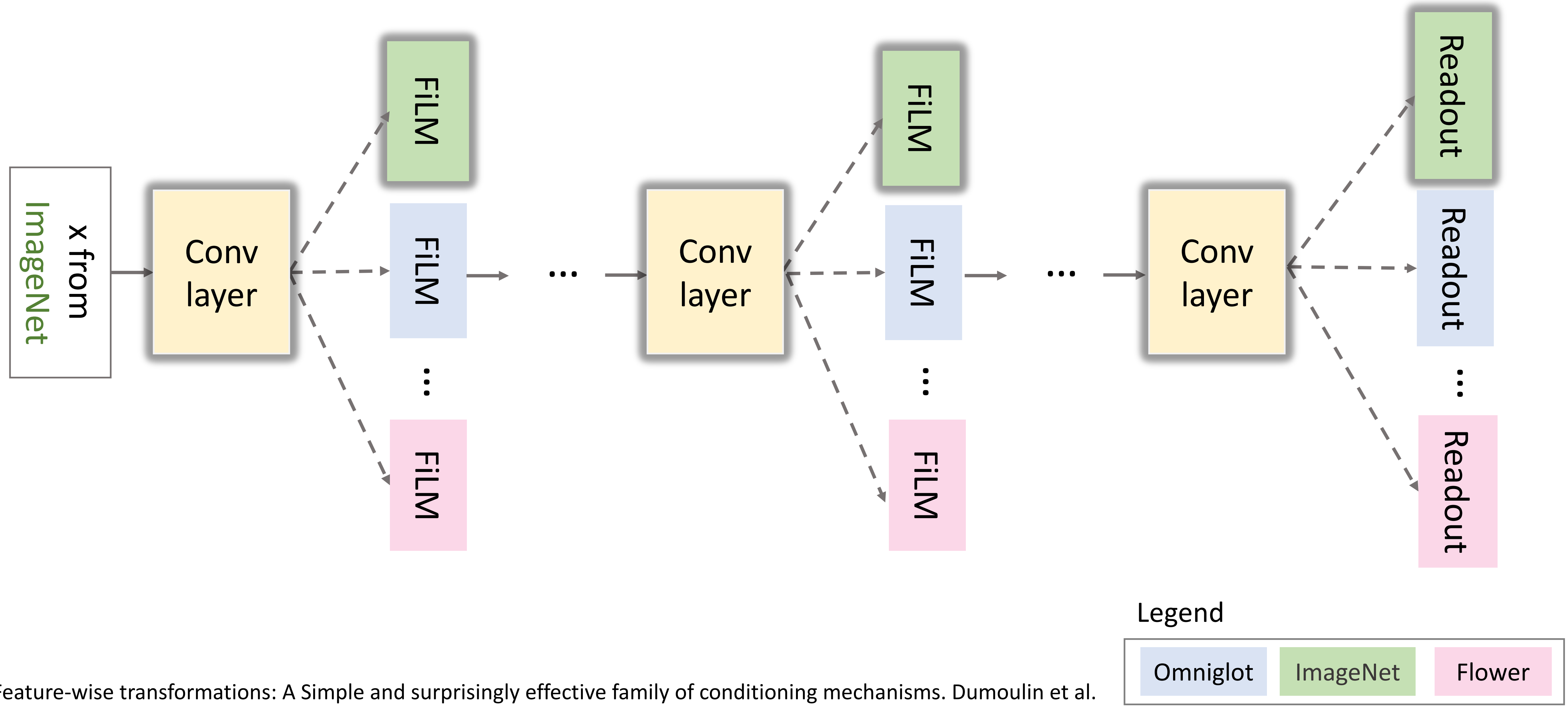


Joint training over all training datasets



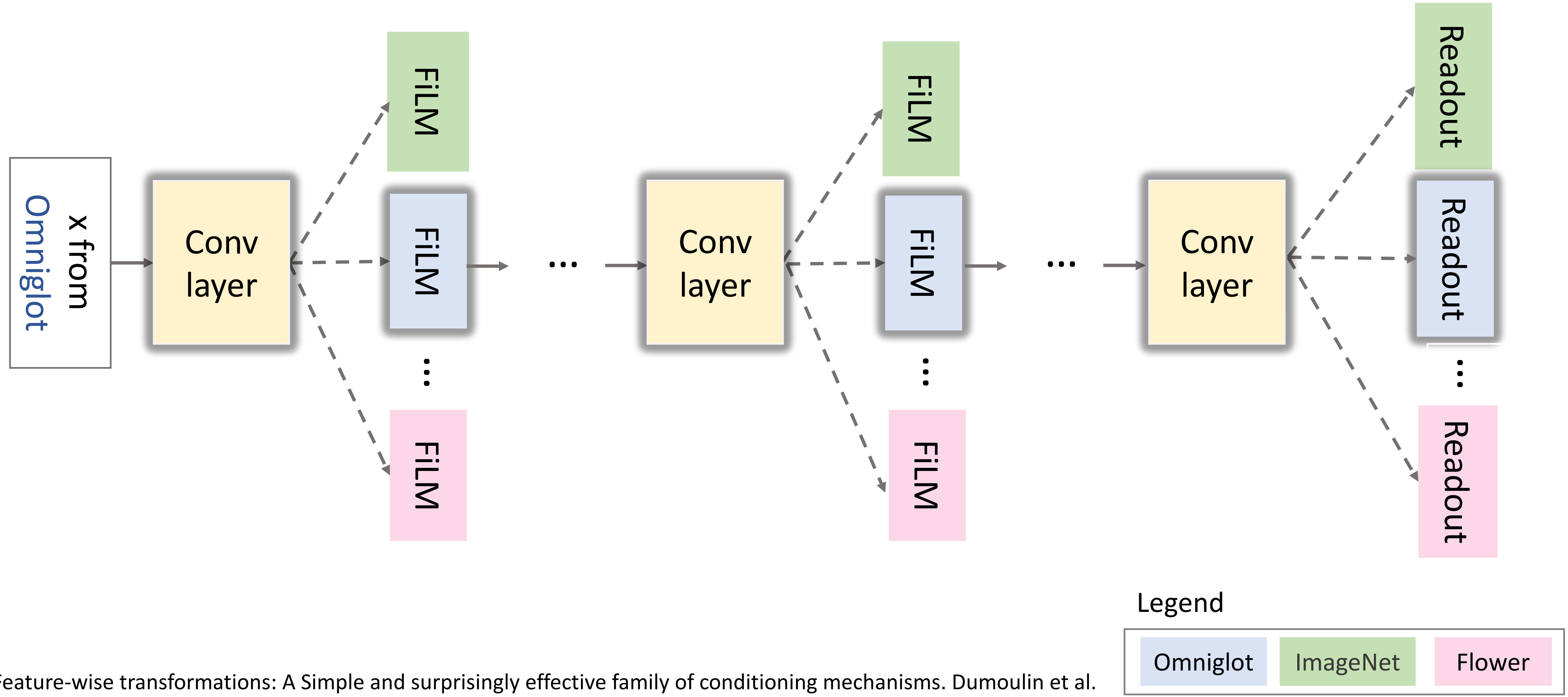
Feature-wise transformations: A Simple and surprisingly effective family of conditioning mechanisms. Dumoulin et al.

Joint training over all training datasets



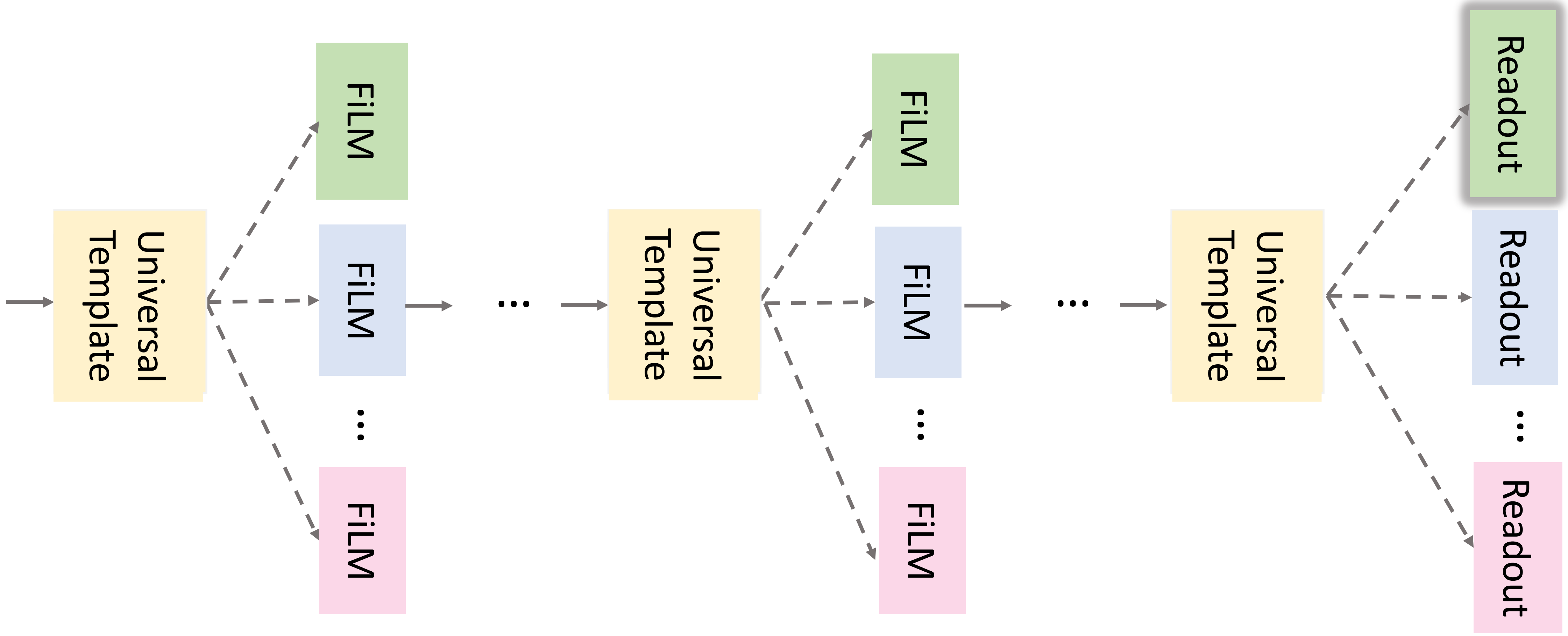
Feature-wise transformations: A Simple and surprisingly effective family of conditioning mechanisms. Dumoulin et al.

Joint training over all training datasets



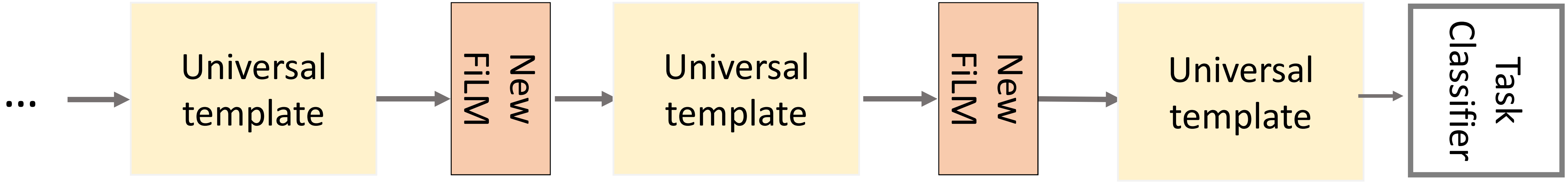
Feature-wise transformations: A Simple and surprisingly effective family of conditioning mechanisms. Dumoulin et al.

Joint training over all training datasets



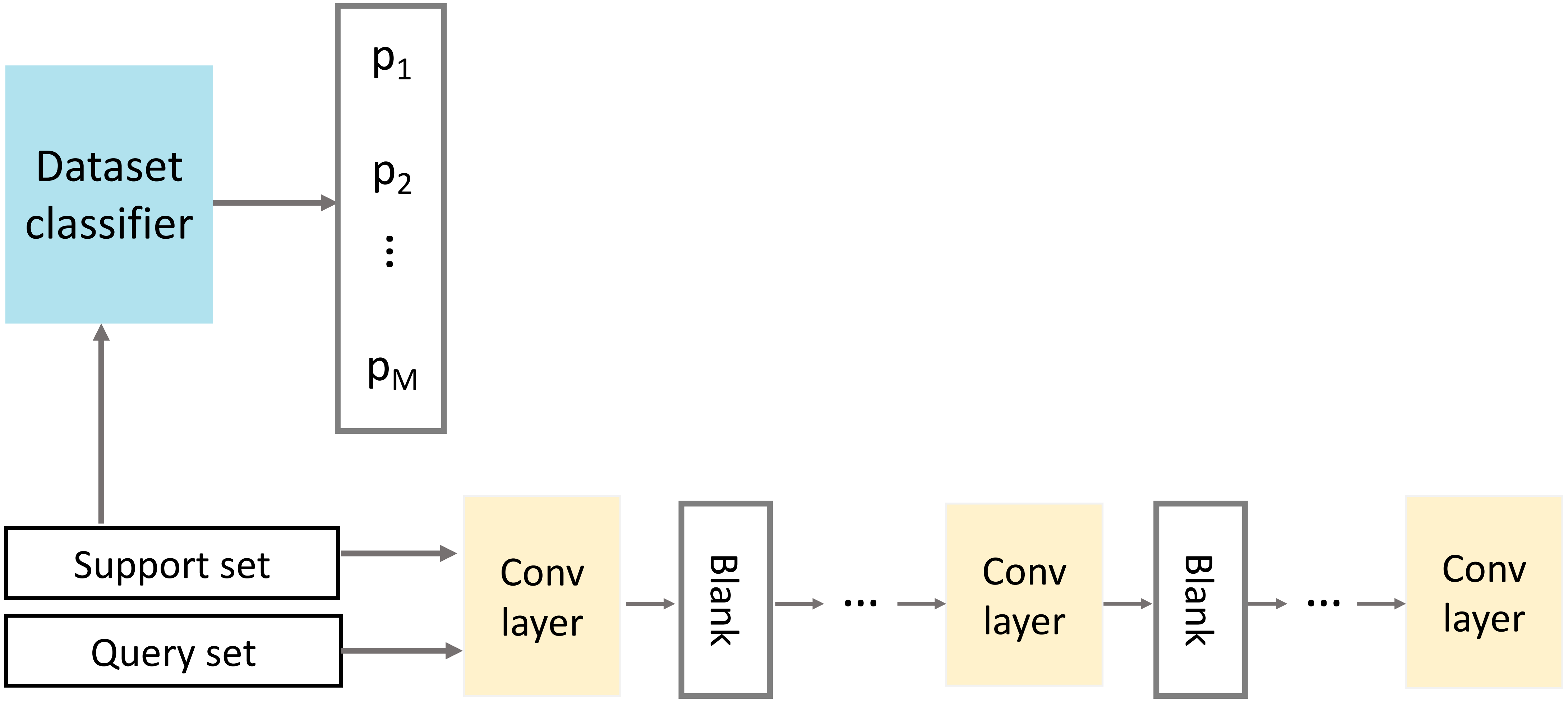
Sharing the convolutional layers across diverse datasets results in turning those into a Universal Template

Defining a feature extractor for a test task

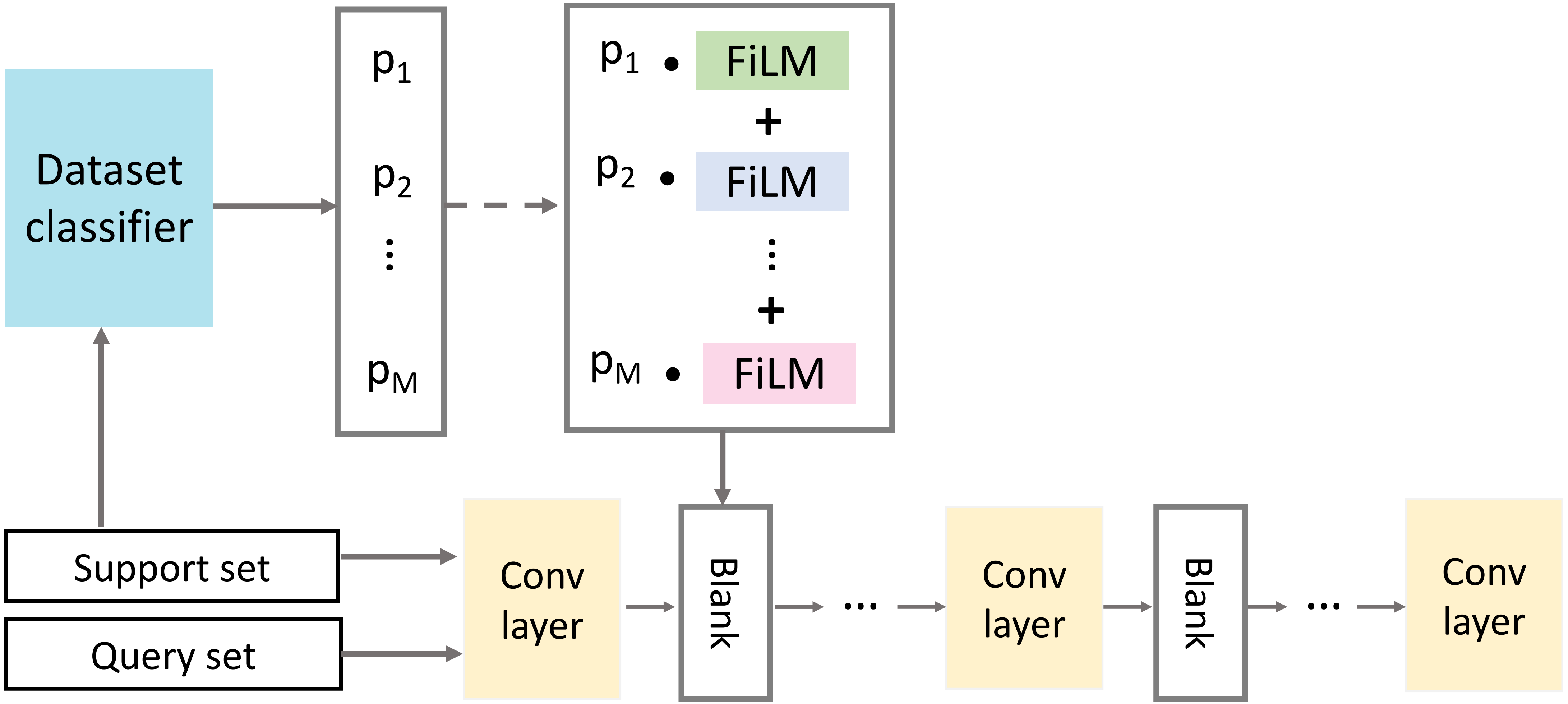


- We propose to learn how to fill in the blanks, via gradient descent on the support set
- We found that the init really matters, we propose to tailor it to each task

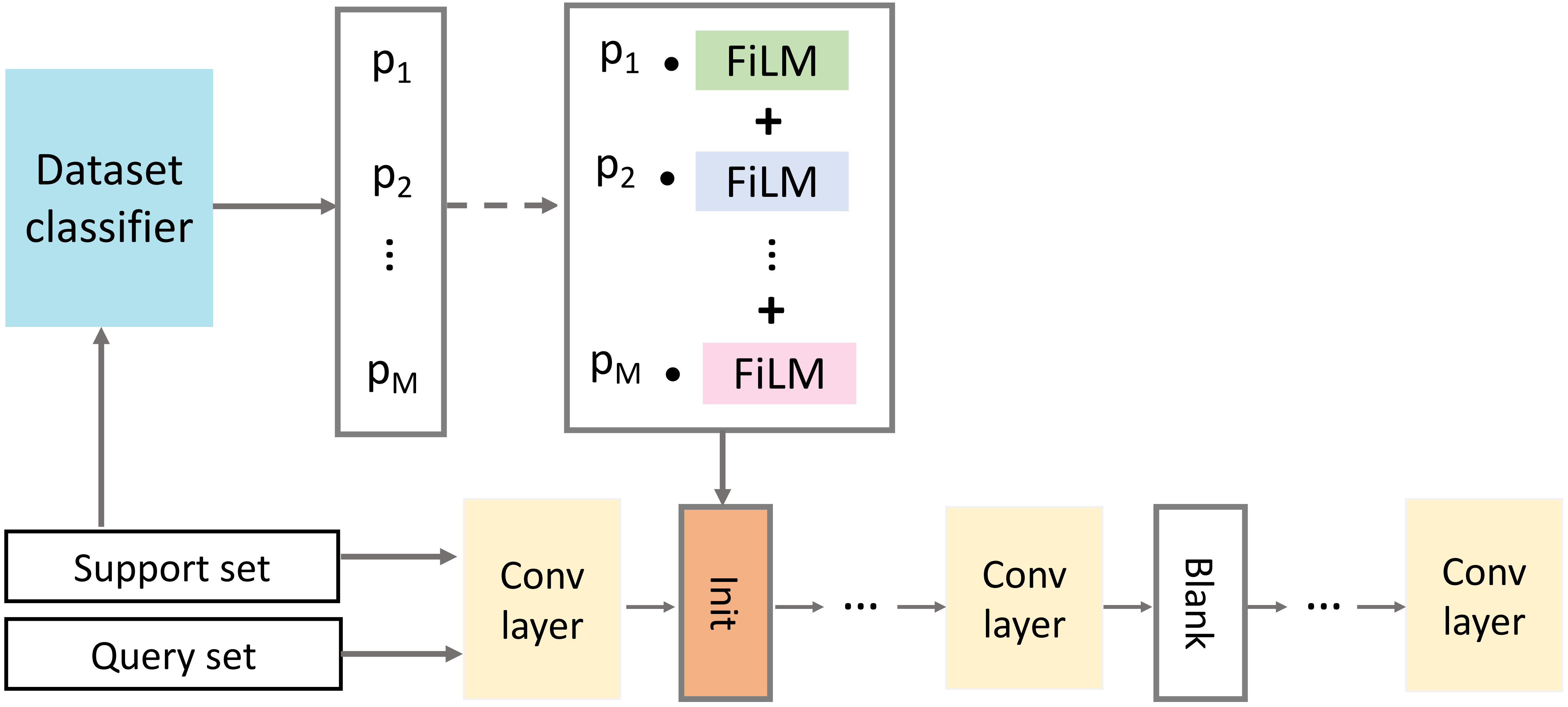
The Blender init scheme



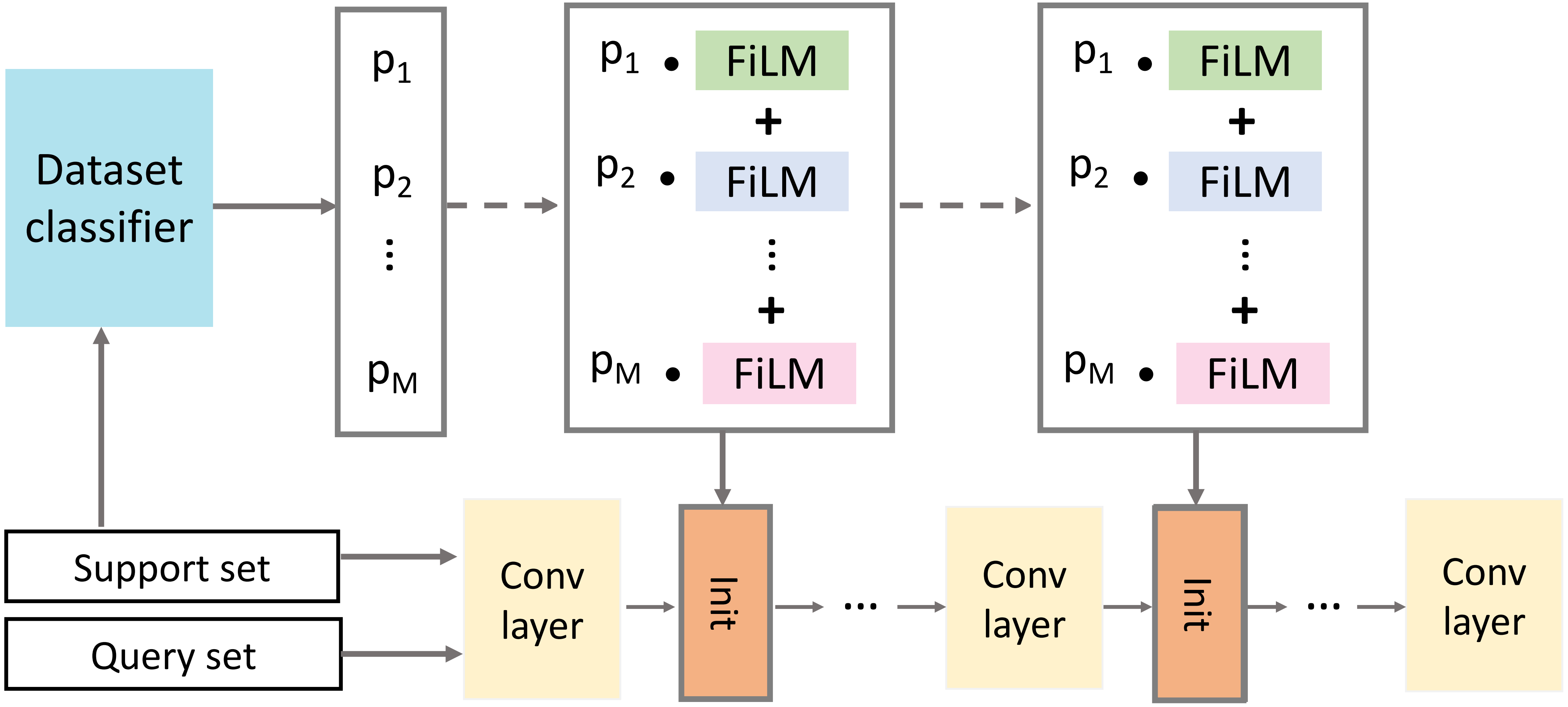
The Blender init scheme



The Blender init scheme



The Blender init scheme

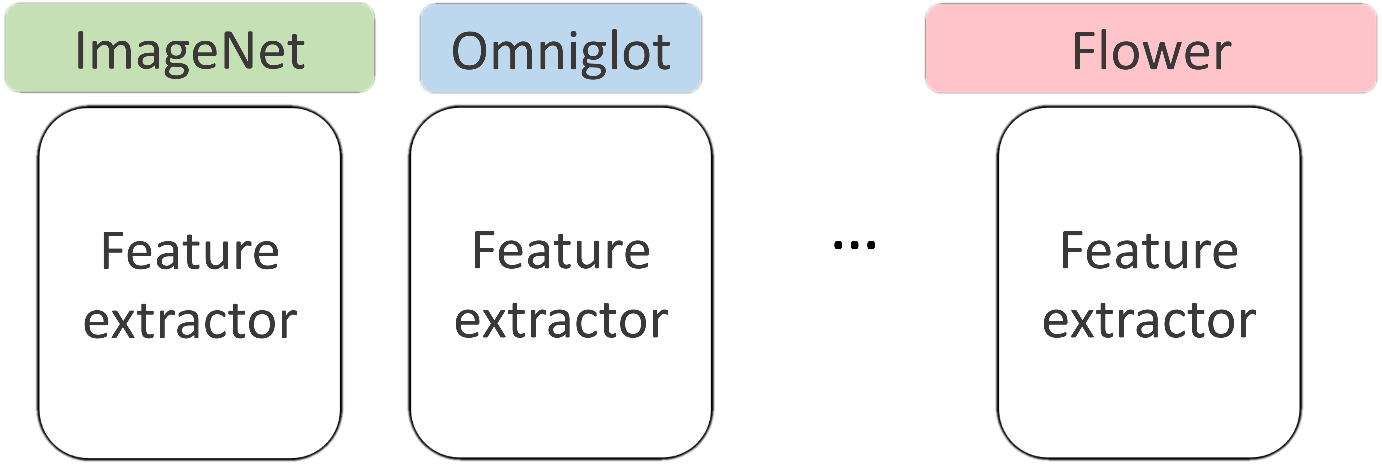


We refer to our approach as…

'Few-shot Learning with a Universal TEmplate'   (FLUTE)

Related work: universal representations

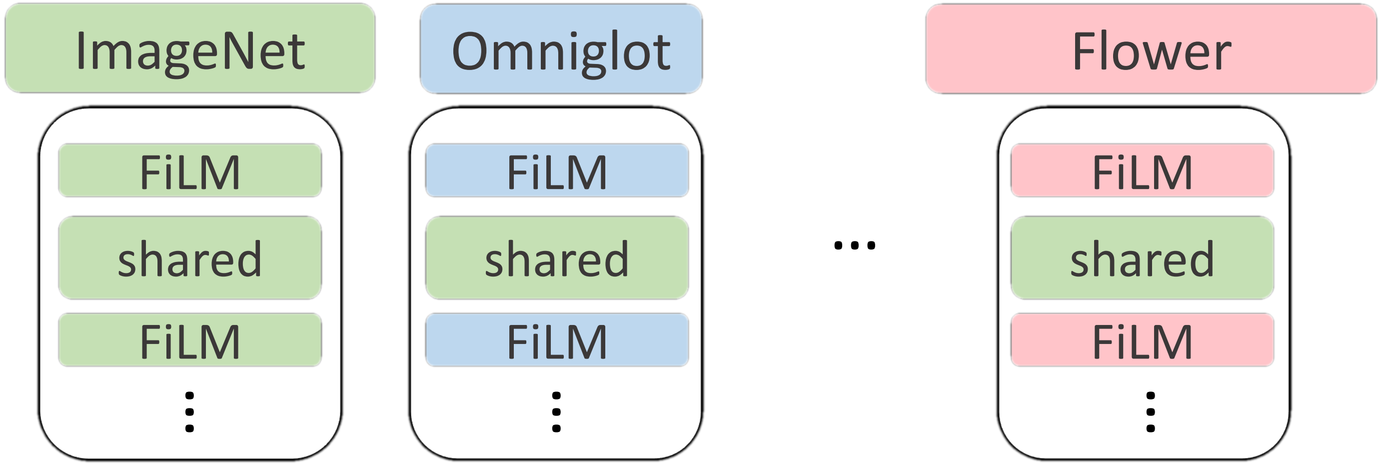
- Train a separate feature extractor on each training dataset



Dvornik, Schmid, Mairal. *Selecting Relevant Features from a Multi-domain Representation for Few-shot Classification.*
Liu, Hamilton, Guodong Long, Jing Jiang, Hugo Larochelle. *A Universal Representation Transformer Layer for Few-Shot Image Classification.*

Related work: universal representations

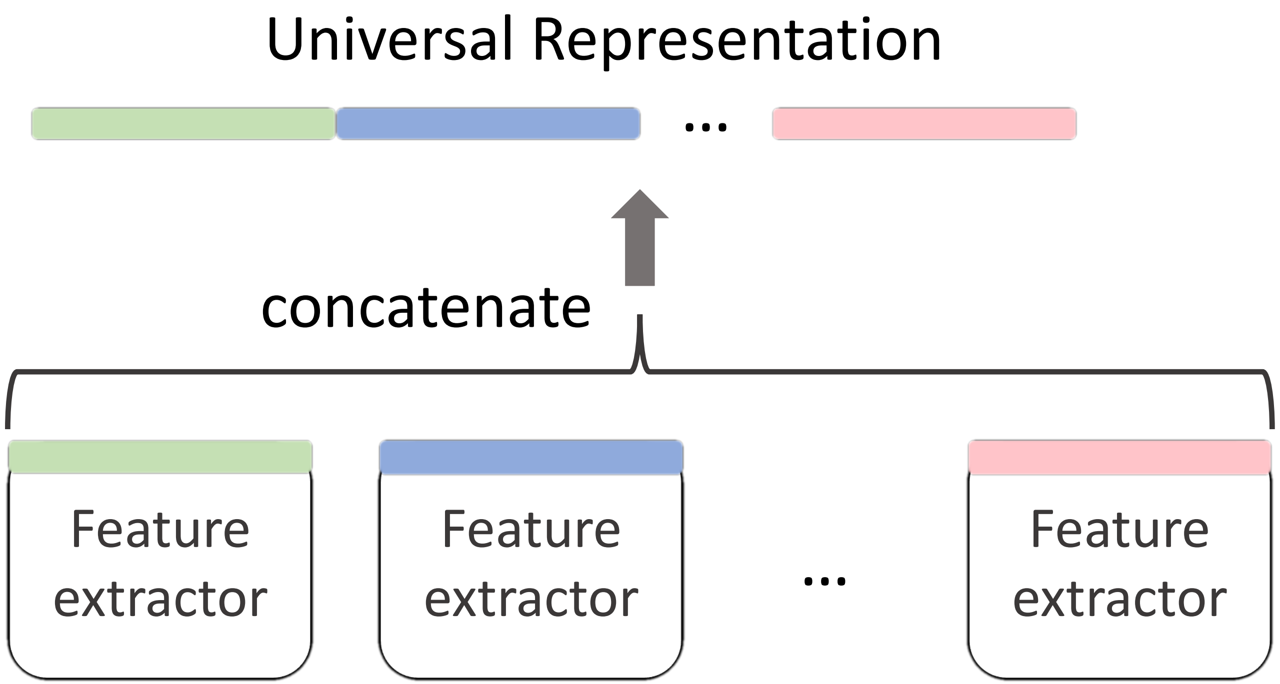
- The 'parametric family' variants share weights across feature extractors



Dvornik, Schmid, Mairal. *Selecting Relevant Features from a Multi-domain Representation for Few-shot Classification.*
Liu, Hamilton, Guodong Long, Jing Jiang, Hugo Larochelle. *A Universal Representation Transformer Layer for Few-Shot Image Classification.*

Related work: universal representations

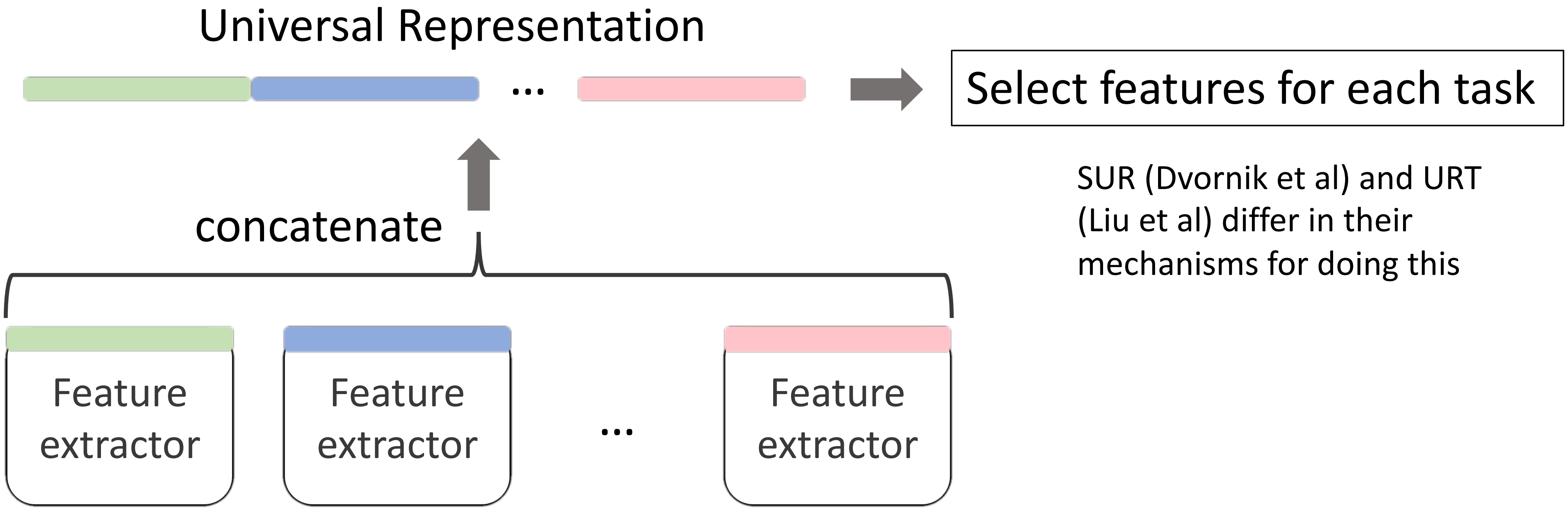
- The universal representation is the concatenation of the representations of the different feature extractors



Dvornik, Schmid, Mairal. *Selecting Relevant Features from a Multi-domain Representation for Few-shot Classification.*
Liu, Hamilton, Guodong Long, Jing Jiang, Hugo Larochelle. *A Universal Representation Transformer Layer for Few-Shot Image Classification.*

Related work: universal representations

- The universal representation is the concatenation of the representations of the different feature extractors



SUR (Dvornik et al) and URT (Liu et al) differ in their mechanisms for doing this

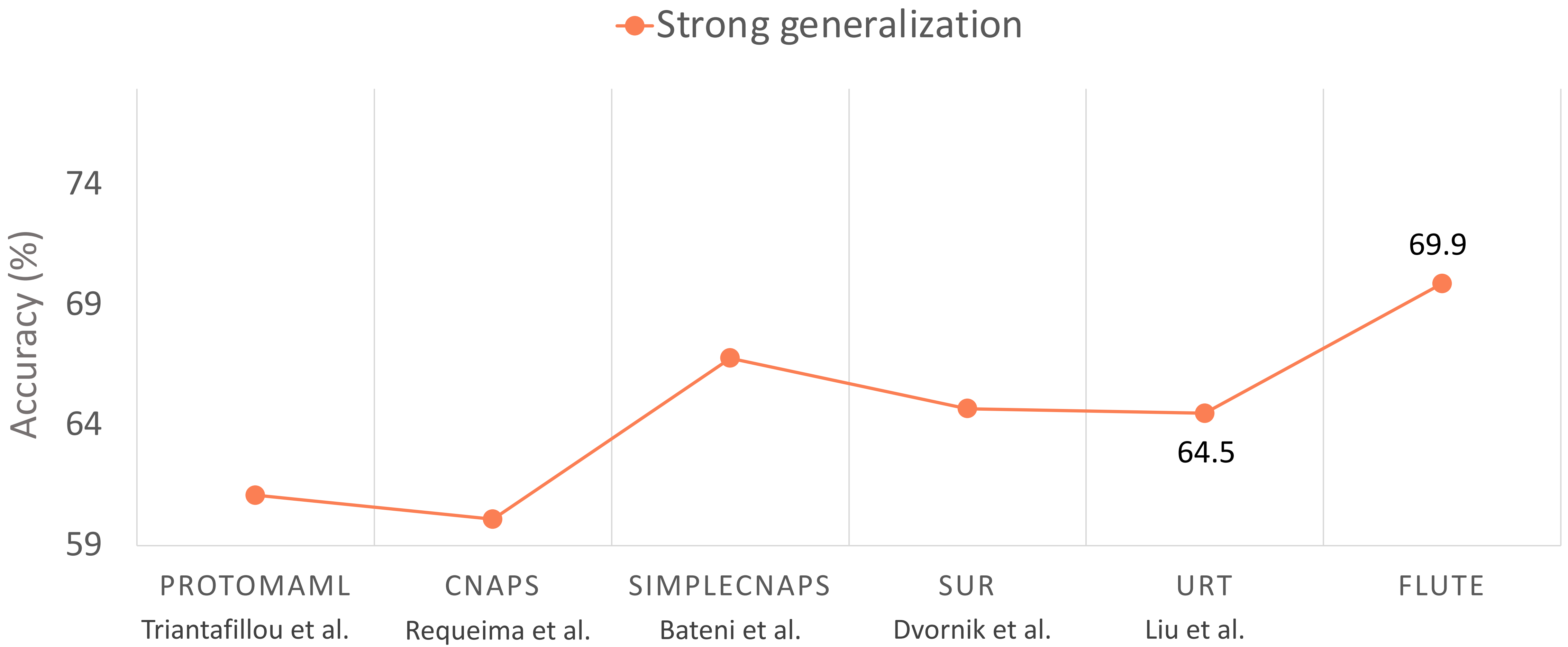
Dvornik, Schmid, Mairal. *Selecting Relevant Features from a Multi-domain Representation for Few-shot Classification.*
Liu, Hamilton, Guodong Long, Jing Jiang, Hugo Larochelle. *A Universal Representation Transformer Layer for Few-Shot Image Classification.*

Universal representations vs FLUTE

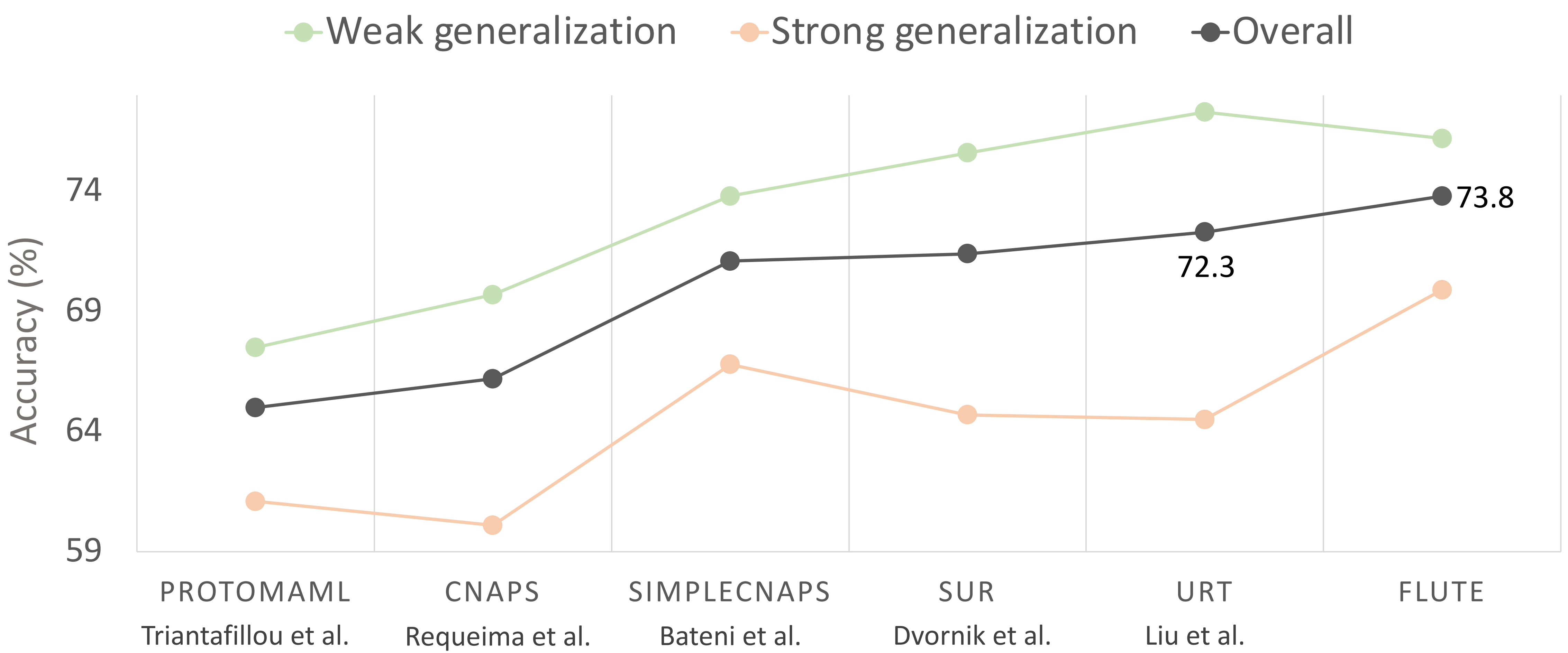
- Different inductive biases: “diverse feature vector” vs “template that can produce diverse feature extractors”

- Compared to SUR/URT, FLUTE is:
  - More expressive
  - More adaptable
  - More scalable

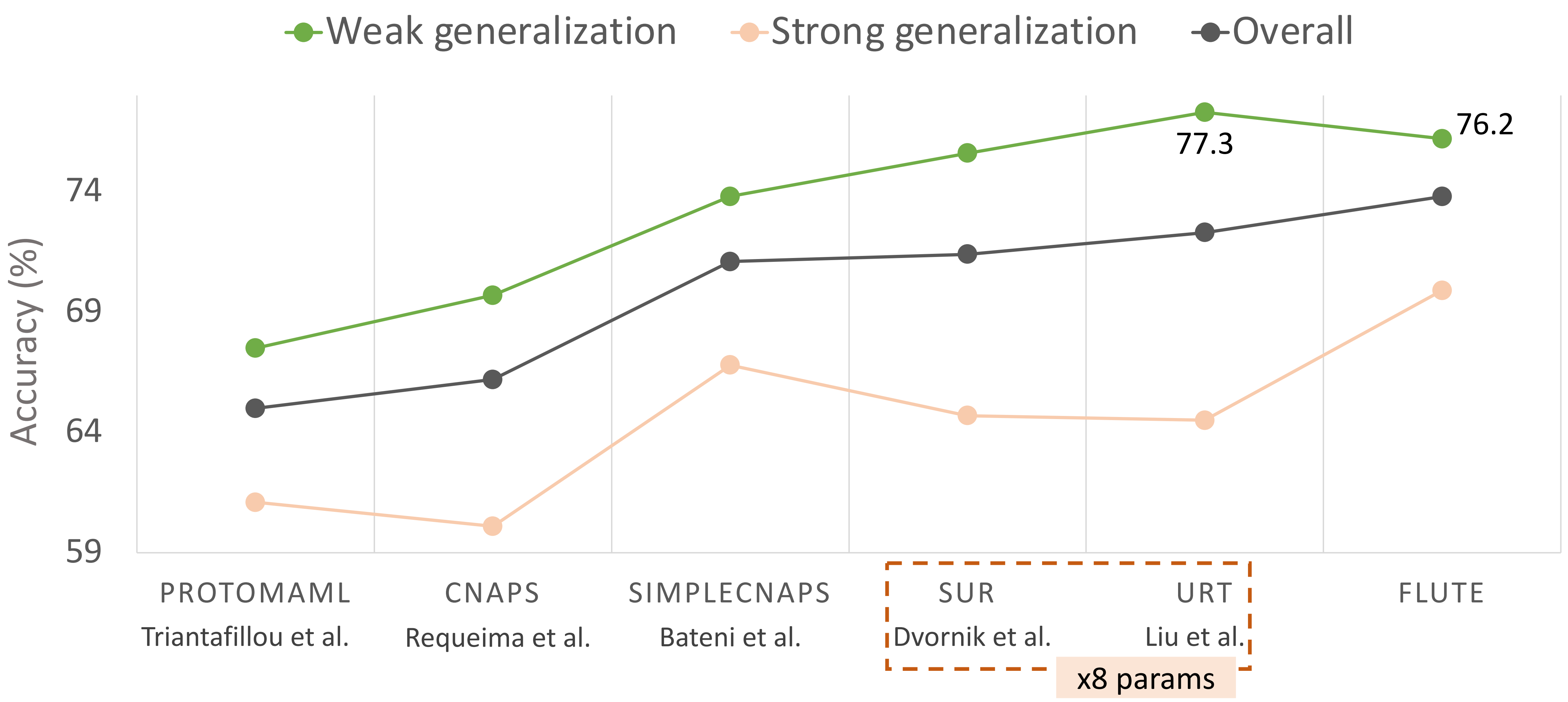
FLUTE improves on strong generalization



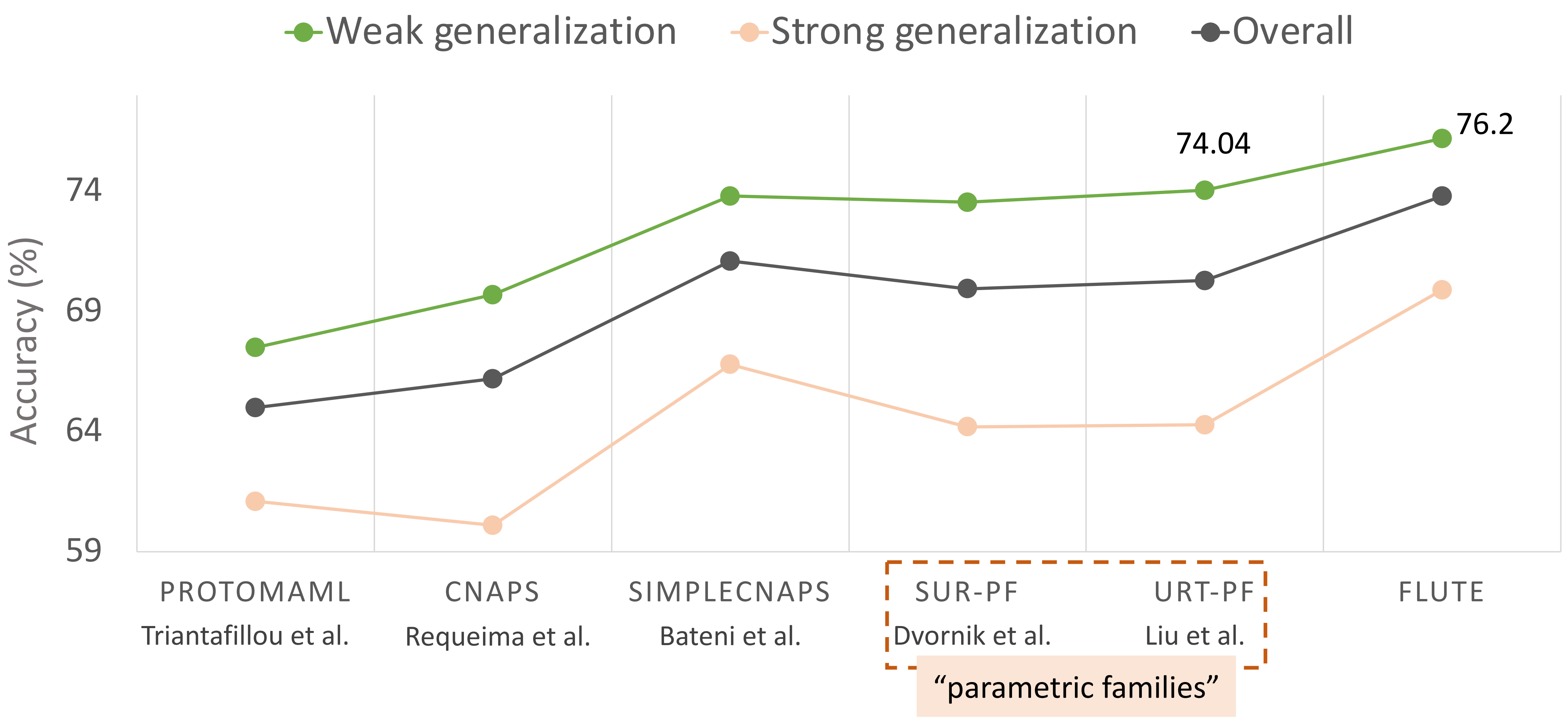
… and on overall performance



… and on overall performance

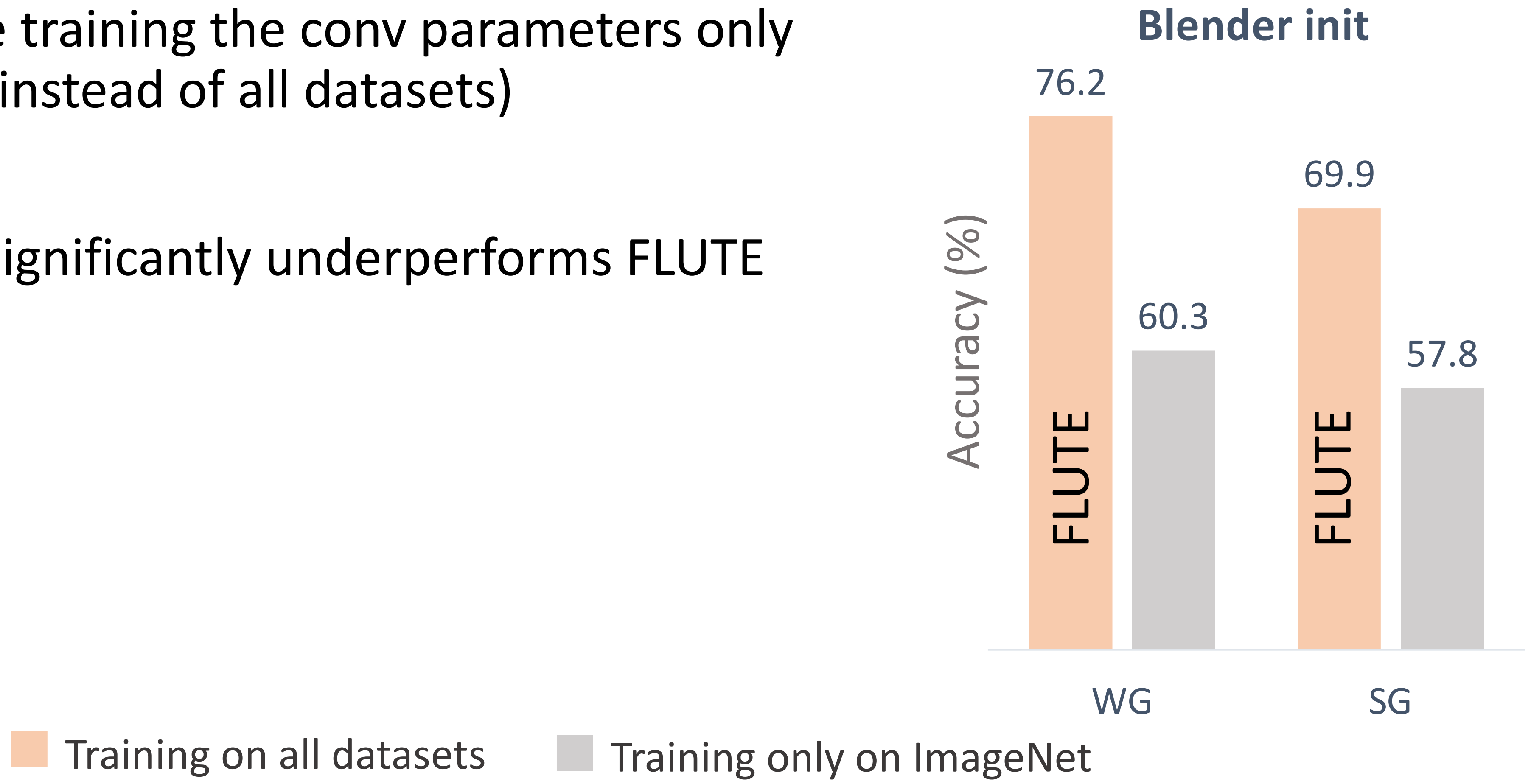


… and on overall performance



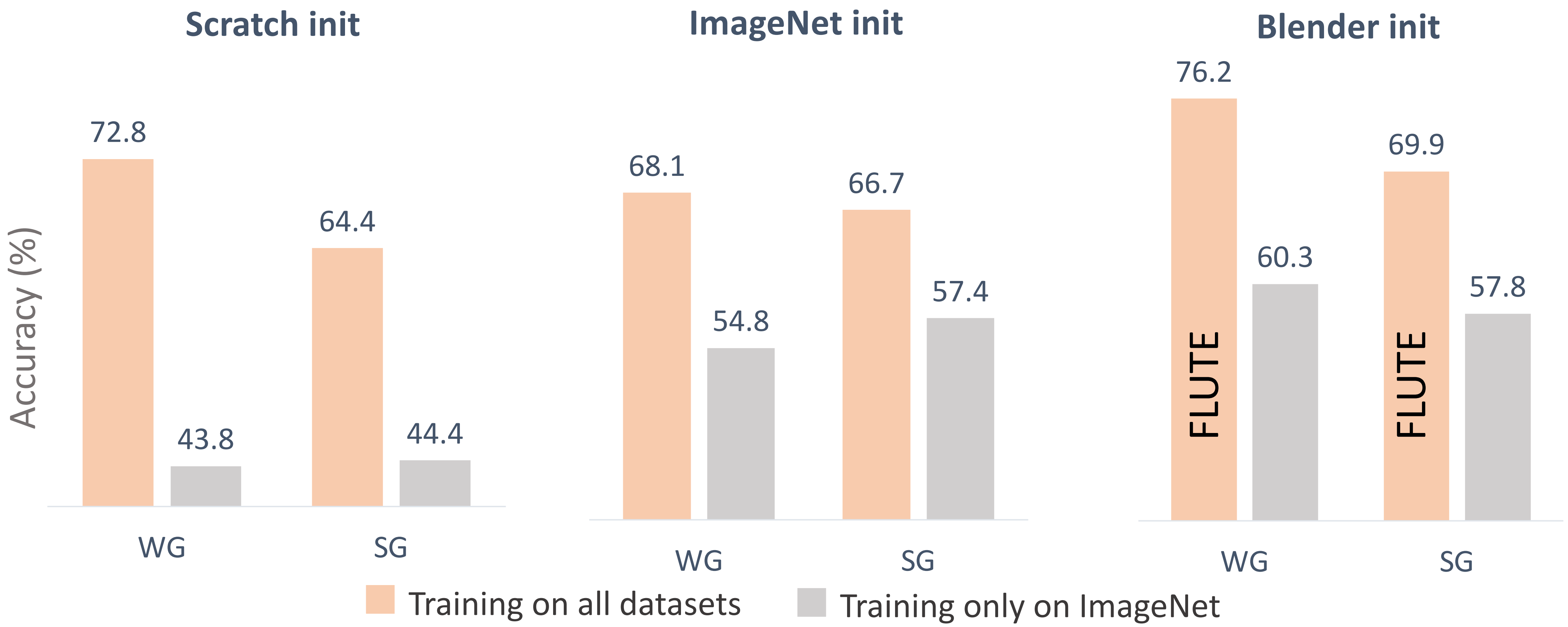
The importance of diverse training data

- We investigate training the conv parameters only on ImageNet (instead of all datasets)
- This baseline significantly underperforms FLUTE



The importance of diverse training data

- The lack of diversity is hurts across different init heuristics



Thank you for listening!

Thank you to my collaborators!