Parallel Network to Learn Novelty from the Known

Shuaiyuan Du, Chaoyi Hong, Zhiyu Pan, Chen Feng*, and Zhiguo Cao
e-mail: {sydu, cyhong, zhiyupan, chen_feng, zgcao}@hust.edu.cn

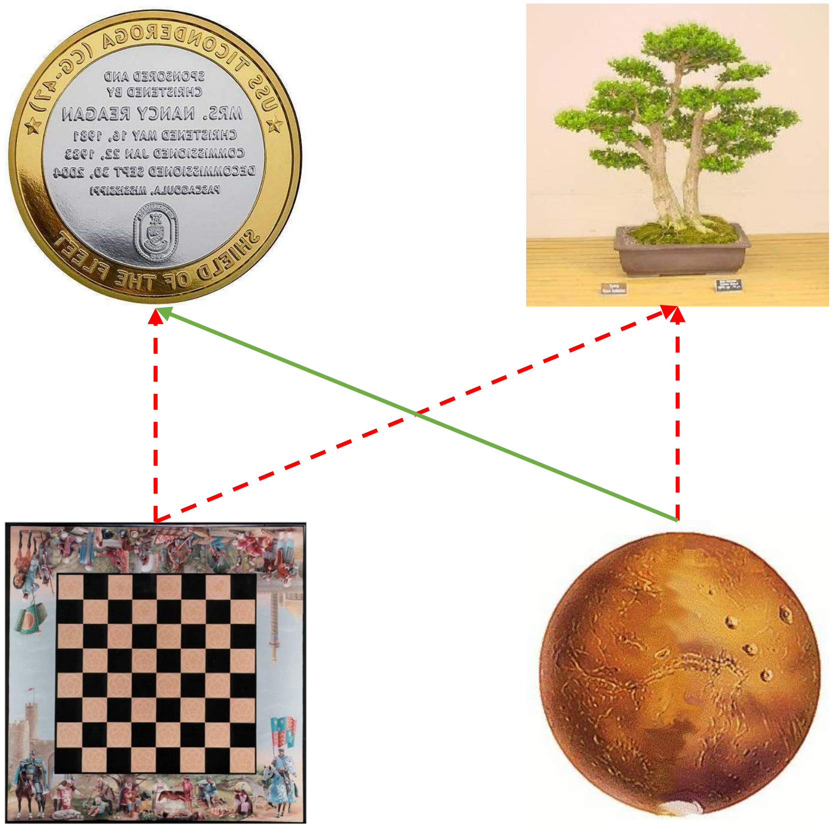
What is the Novelty Detection ?

Recognize the unseen classes, or technically, those classes never appearing during training.

The Challenge

Known
(train)

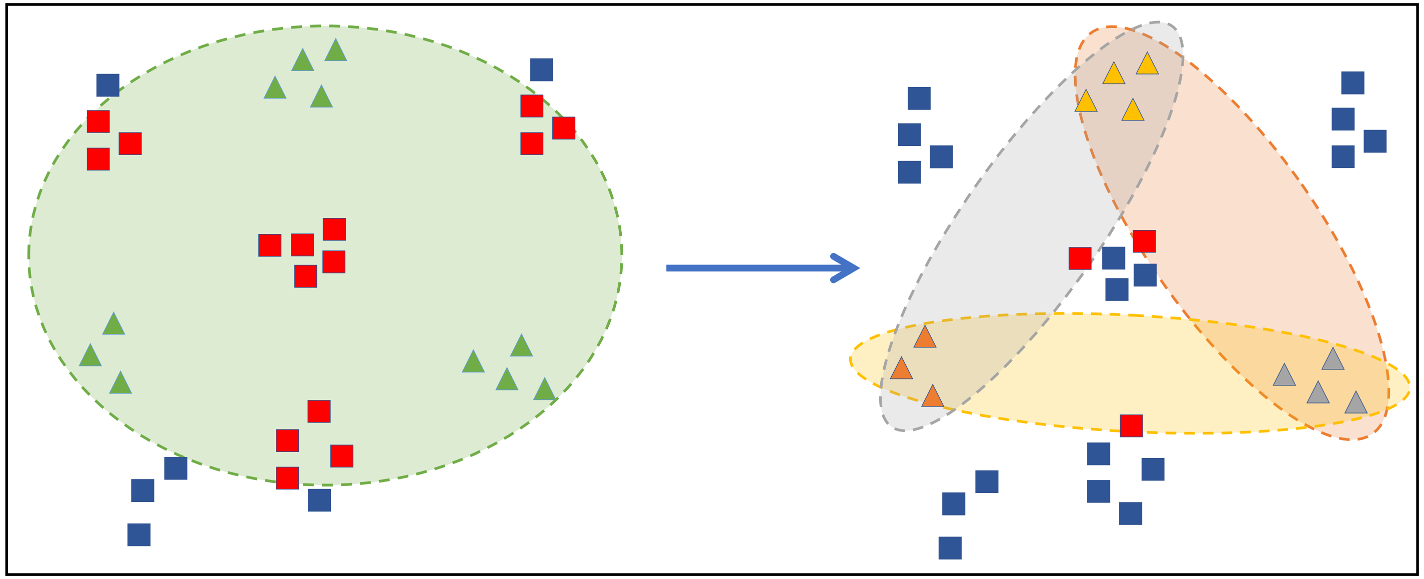
Novelty
(test)



→ misclassify - - → correct classify

The limit of classic classification network:
Cannot discriminate novel samples resembling training classes.

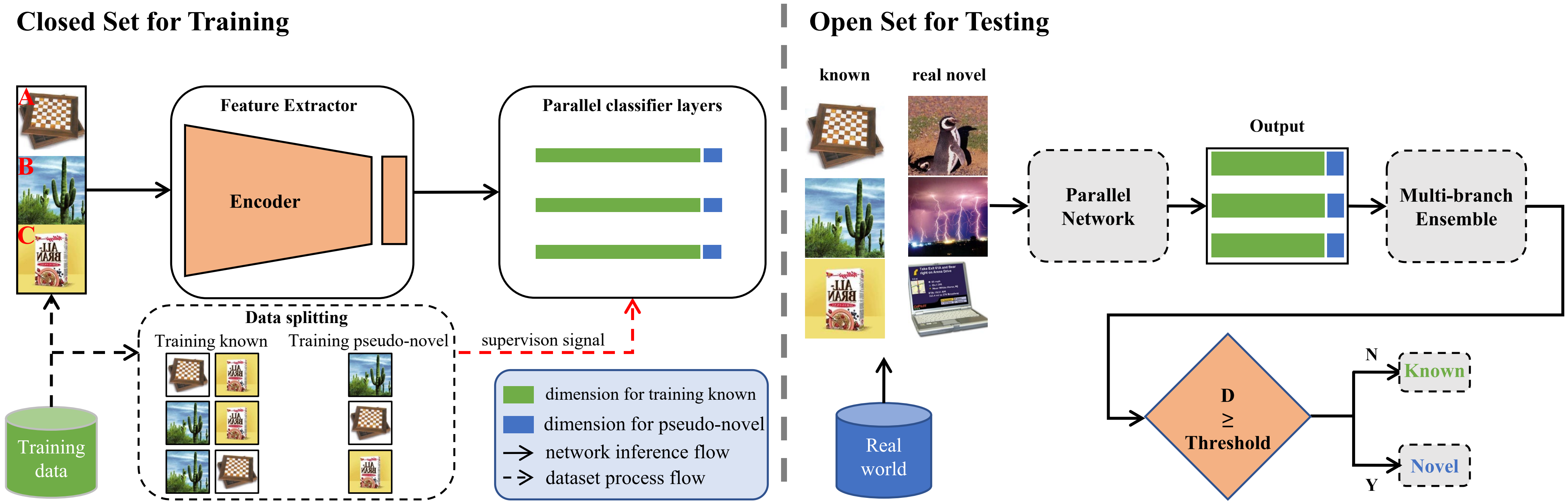
Key Idea



▲ known or pseudo-novel ■ real novel ◌ classification space

Construct subtasks of pseudo-novelty detection

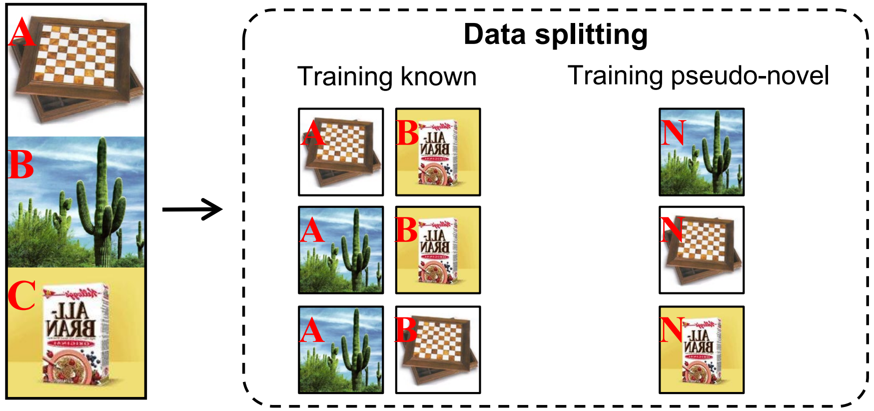
Pipeline



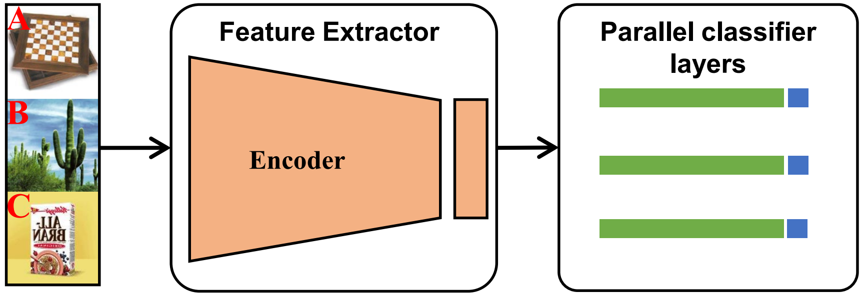
Closed Set: **lack** real novel samples in training
Open Set: **meet** real novel samples in testing

Closed Set for Training

Generate Pseudo-novel



Closed Set for Training



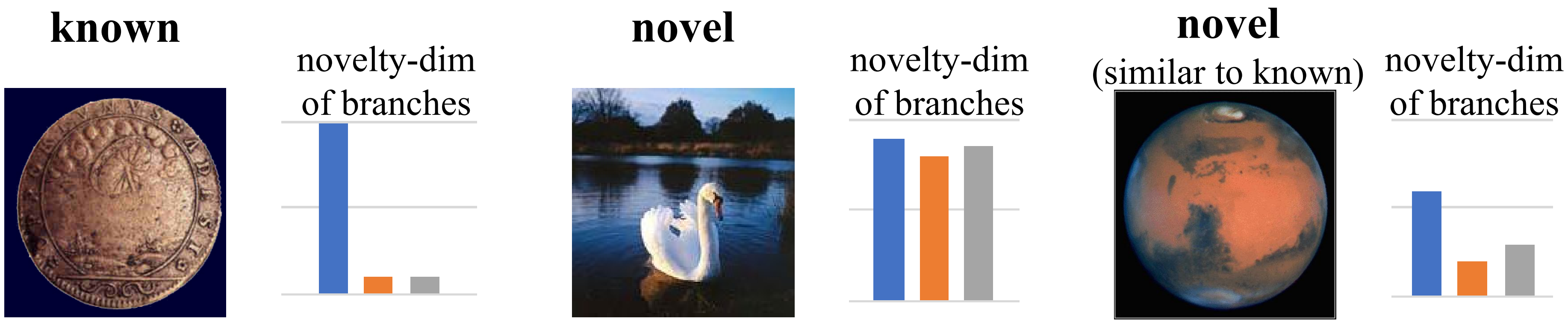
Generate Pseudo-novel Sub-tasks

- Use only the training set to create sub-tasks of pseudo-novelty detection.
- All training classes share equal possibility of being pseudo-novel during training.

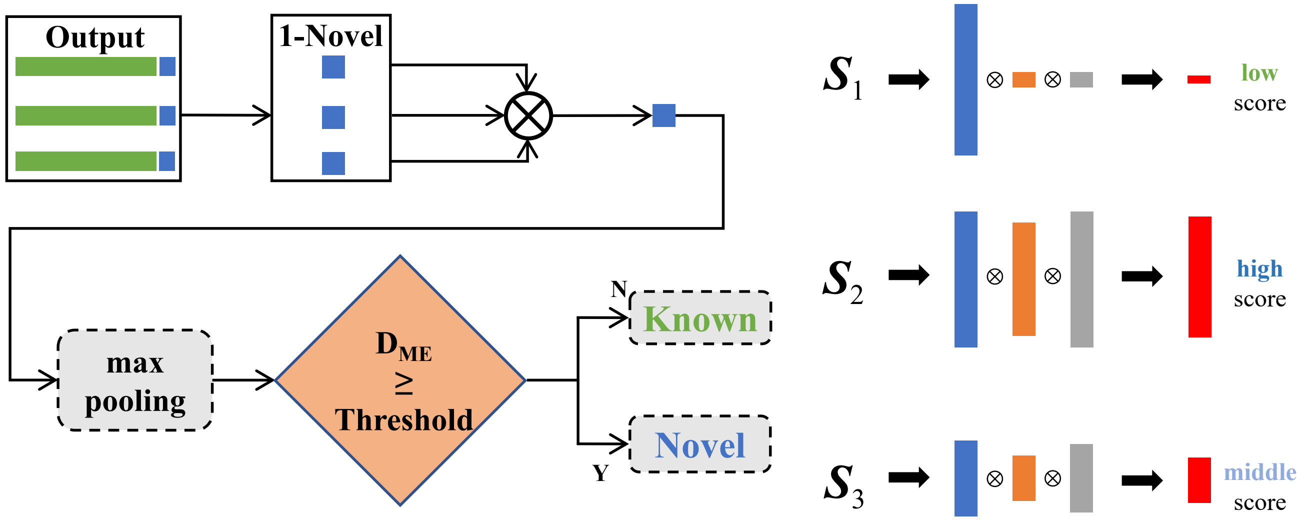
Train PN Within Closed Set

- “Parallel” denotes multiple branches of co-working FC classifiers.
- The classes acting as “pseudo-novel” to train each FC are unoverlapped.

Open Set for Testing



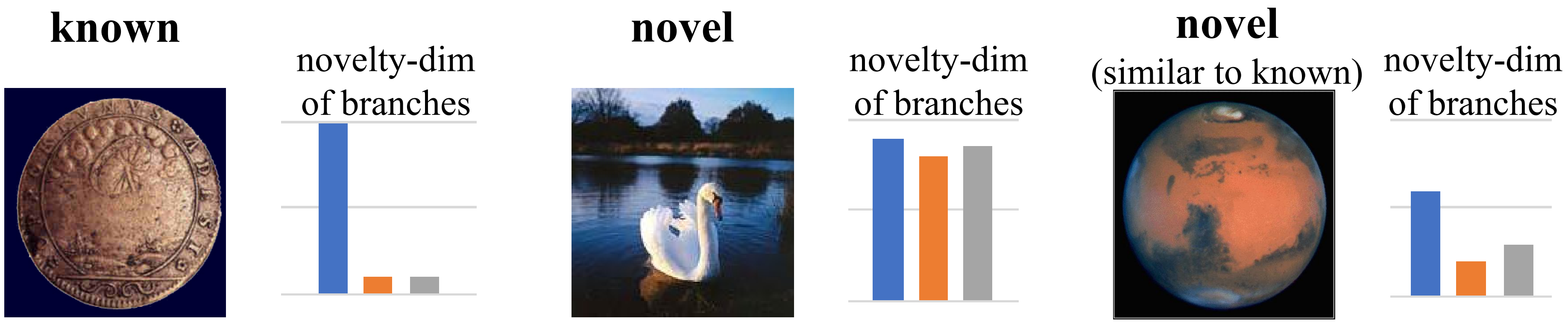
Samples within different classes shall have different output distributions.



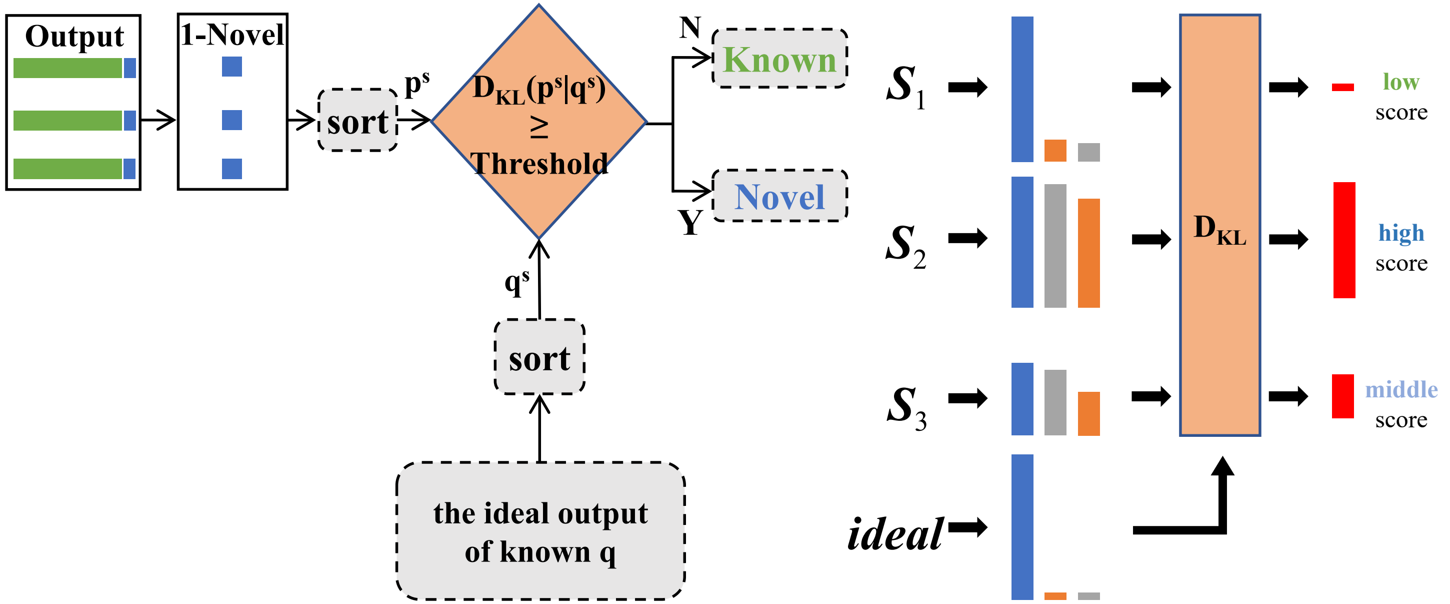
Multiplicative Ensemble

Multiply the novel dimensions in all FCs.

Open Set for Testing



Samples within different classes shall have different output distributions.



KL-Divergence based Ensemble

- Measure the Kullback-Leibler (KL) divergence between the ideal output of known classes and any test sample.

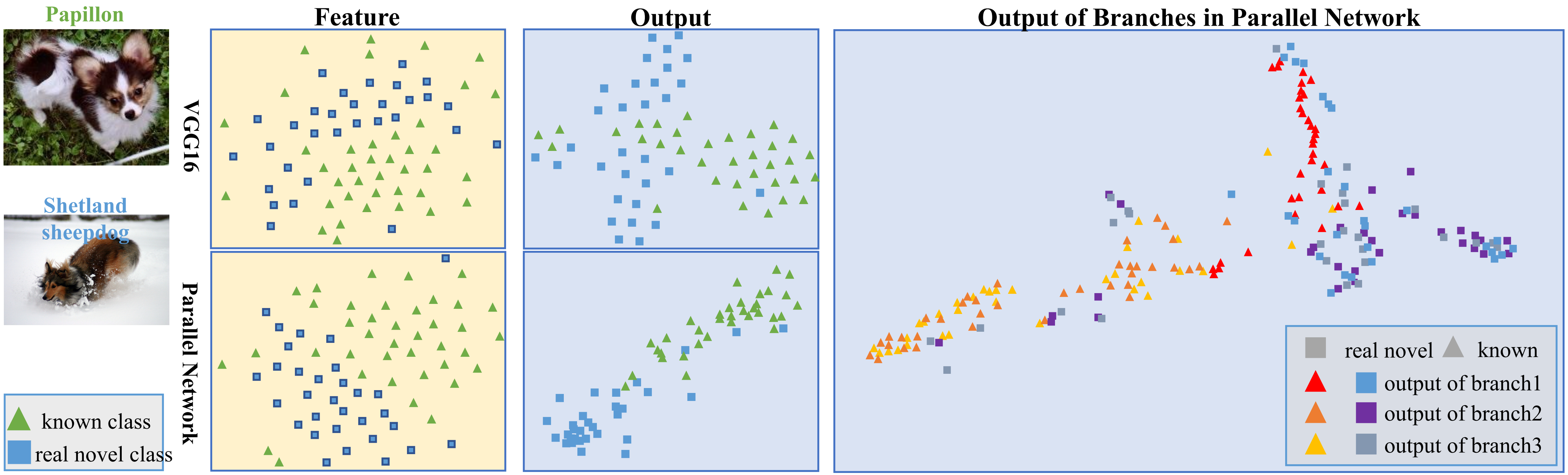
Comparison & Ablation Study

Method	Stanford Dogs	Caltech-256	Founder-Type200	Mean
FT(baseline)	0.766	0.827	0.841	0.811
One-class SVM	0.542	0.576	0.627	0.582
KNFST	0.649	0.743	0.870	0.754
Local KNFST	0.652	0.712	0.673	0.679
OpenMax	0.776	0.831	0.852	0.820
FT(c+C)	0.780	0.848	0.754	0.794
Deep Novelty	0.825	0.869	*0.893*	*0.862*
Ours (ME)	**0.833**	**0.882**	0.871	*0.862*
Ours (KLD)	*0.829*	*0.873*	**0.901**	**0.868**

Features	Classifier	AUC	imp.
baseline	baseline	0.689	+0.000
ours	baseline	0.709	+0.020
baseline	ours	0.725	+0.036
ours	ours	**0.829**	**+0.140**

State-of-the-art and robust on 3 kinds of public datasets.
Without additional data, PN still seemed to have mined the concept of novelty.
Feature and classifier benefit from each other.

Analysis Study



Compared to Baseline, out PN

- builds larger margin between visually similar classes;
- shows better discriminative power in both feature and classifier output.
- shows great difference between the integral distributions of known and real novel.

Summary

Propose end-to-end PN for novelty detection.

1. PN learns a more compact and discriminative feature or output space.
2. PN learns the concept of "novelty" with only training known classes, *i.e.*, no novel samples.
3. We designed careful experiments to validate our proposition.

THANKS FOR YOUR ATTENTION