Parallel Network to Learn Novelty from the Known

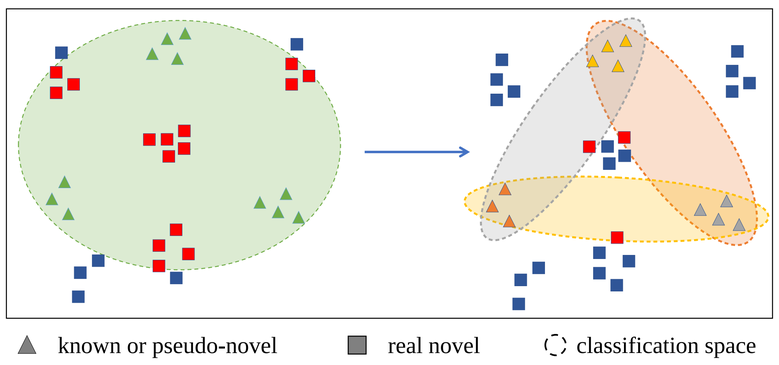
Shuaiyuan Du, Chaoyi Hong, Zhiyu Pan, Chen Feng*, and Zhiguo Cao

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Novelty Detection

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

Key Idea



- **Classic classification network** cannot discriminate novel samples resembling training classes.
- **We** divide the training set to construct subtasks of pseudo-novelty detection and learns the concept of "novelty".

Comparison & Ablation Study

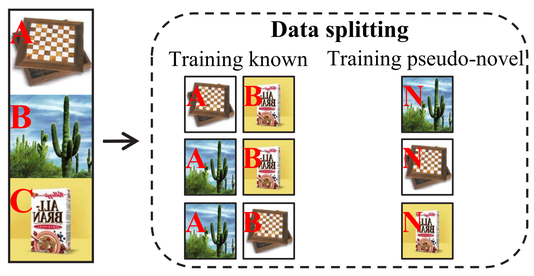
Method	Stanford Dogs	Caltech-256	Founder-Type200	Mean
FT(baseline)	0.766	0.827	0.841	0.811
One-class SVM	0.542	0.576	0.627	0.582
KNFST	0.649	0.743	0.870	0.754
Local KNFST	0.652	0.712	0.673	0.679
OpenMax	0.776	0.831	0.852	0.820
FT(c+C)	0.780	0.848	0.754	0.794
Deep Novelty	0.825	0.869	*0.893*	*0.862*
Ours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

State-of-the-art and robust.

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**

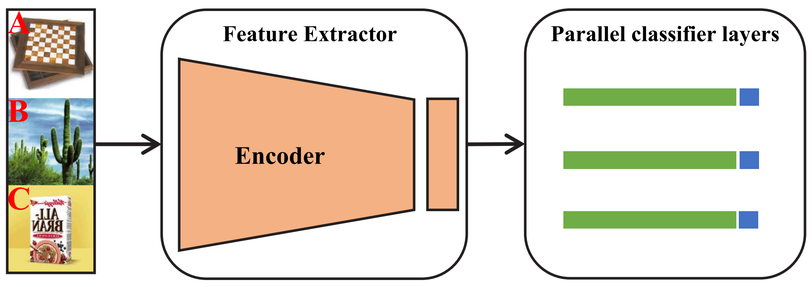
In our structure, the feature and classifier benefit from each other.

PN for closed-set Training



Generate Pseudo-novel Sub-tasks.

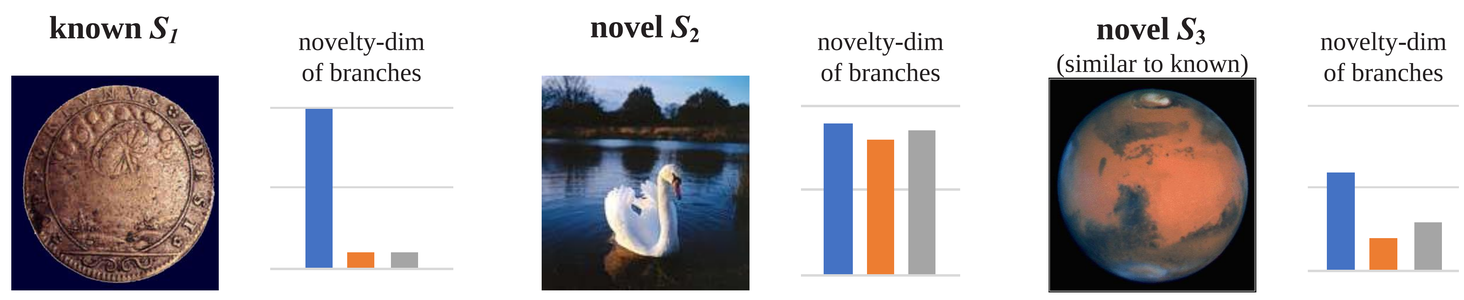
- Use only the training set to create sub-tasks of pseudo-novelty detection.
- All training classes share equal possibility of being pseudo-novel during training.



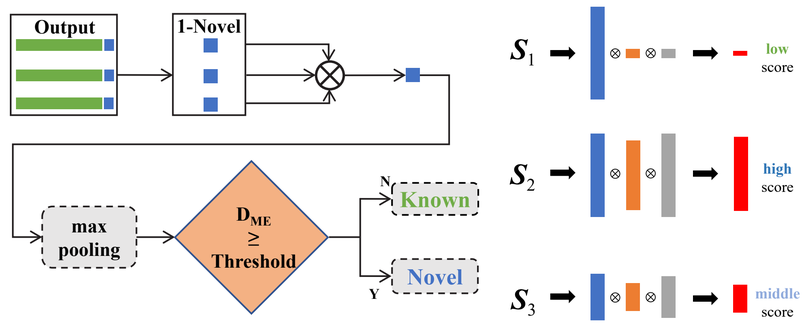
Train PN Within Closed Set

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

Multi-Branch Ensemble for Open-set Testing

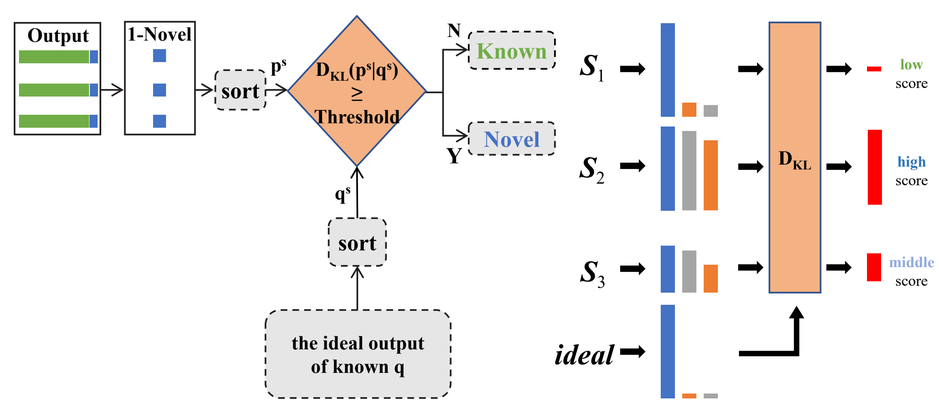


Samples within different classes shall have different output distributions.
(Totally three classes in this example.)

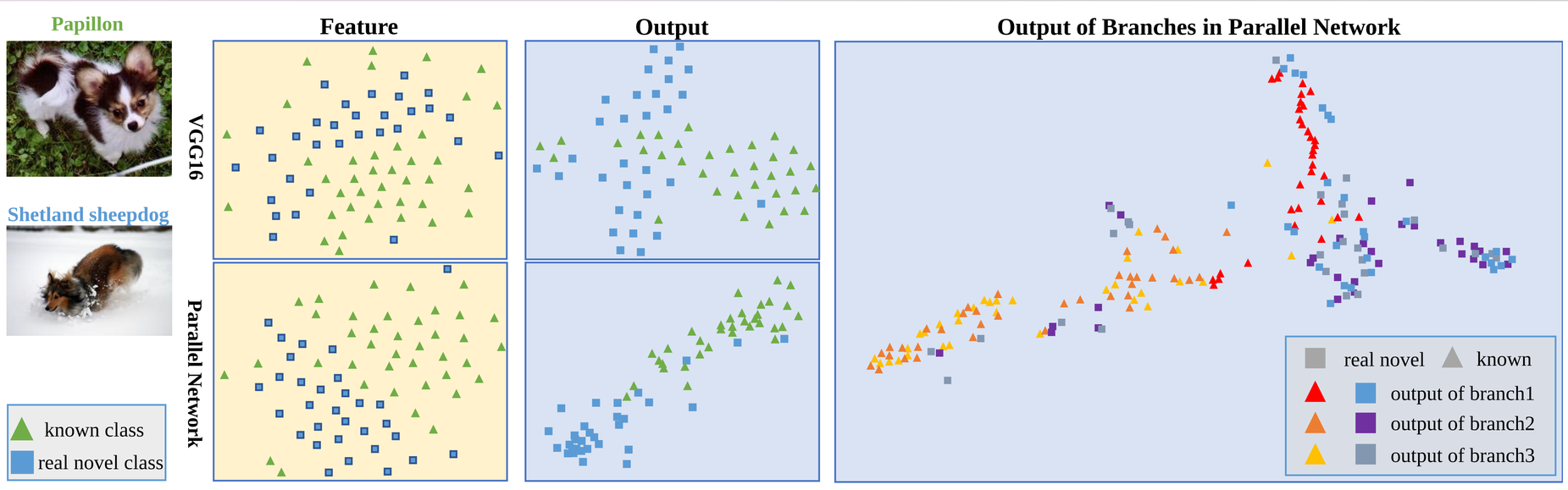


1: Multiplicative Ensemble
Multiply the novel dimensions in all classifiers.

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



Analysis Experiment



Compared with the baseline, our PN

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