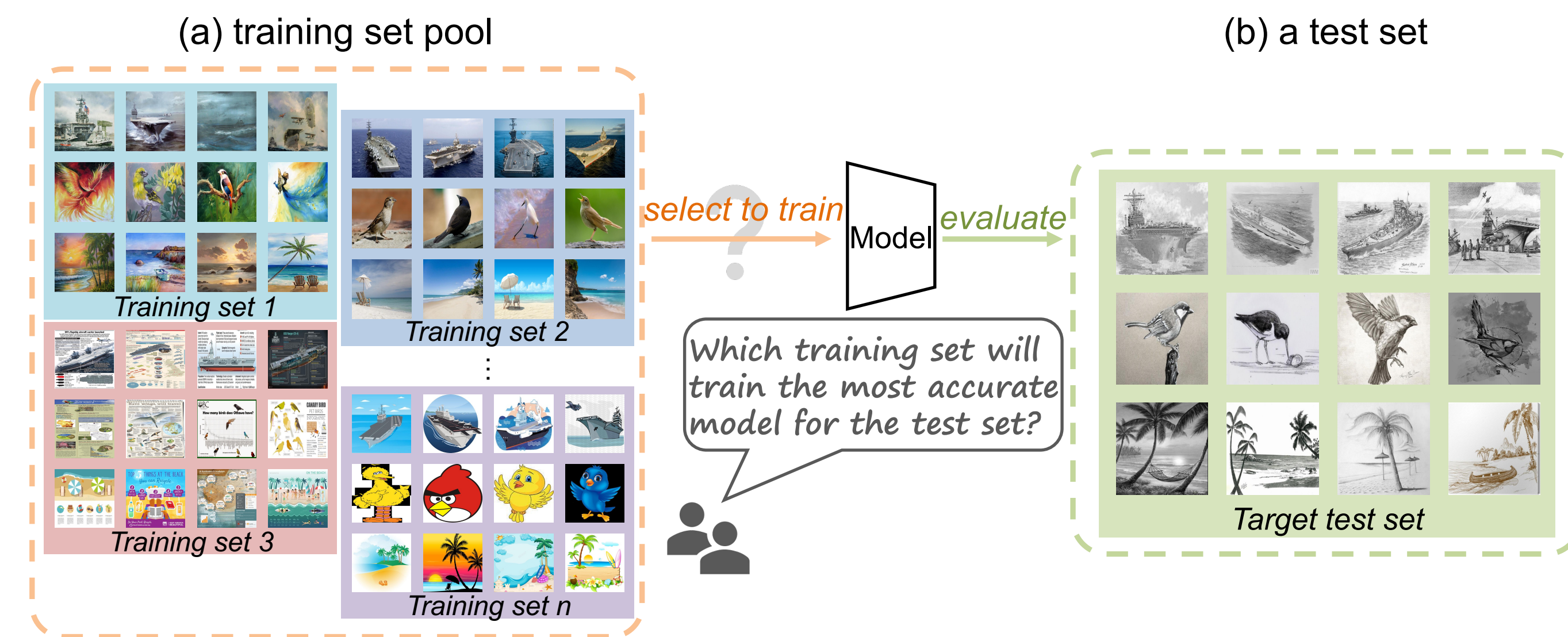
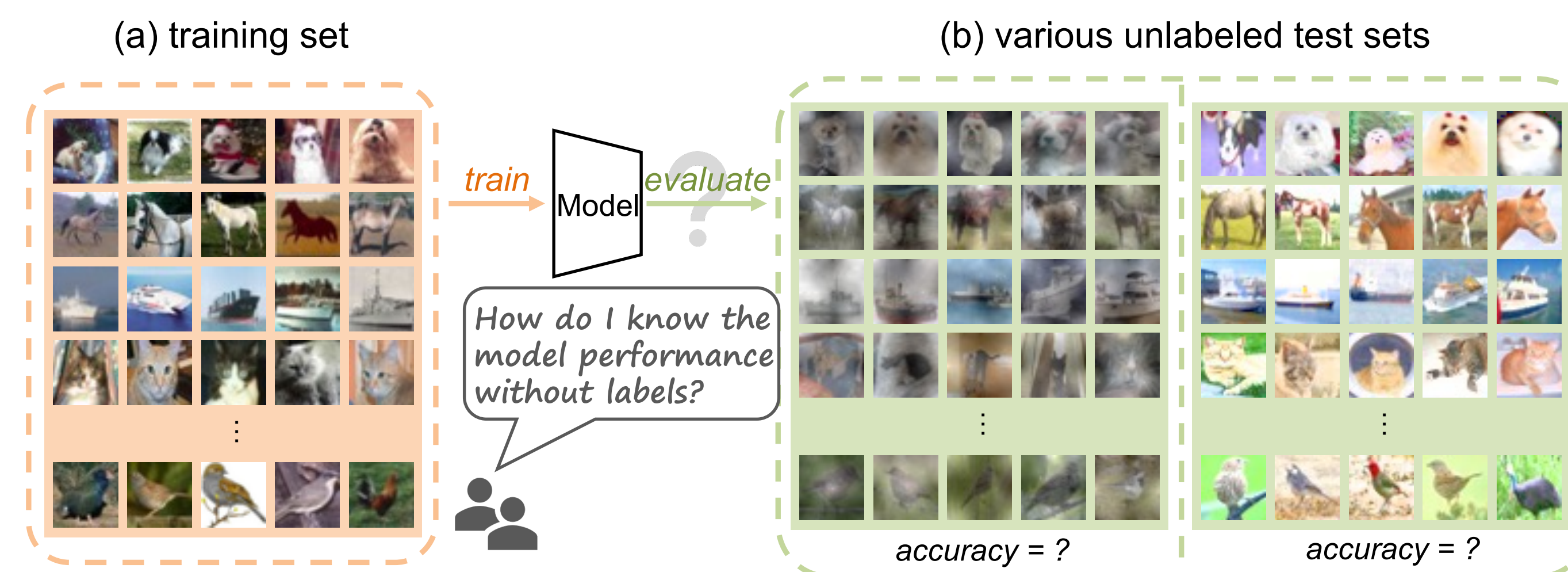


Two Dataset-Level Applications:

- **Training set suitability:** Given a target test set and multiple training sets from different distributions, the goal is to select (ideally **without training**) the most suitable training set.



- **Test set difficulty:** Given a trained model, the objective is to estimate its accuracy on various test datasets **without labels**.

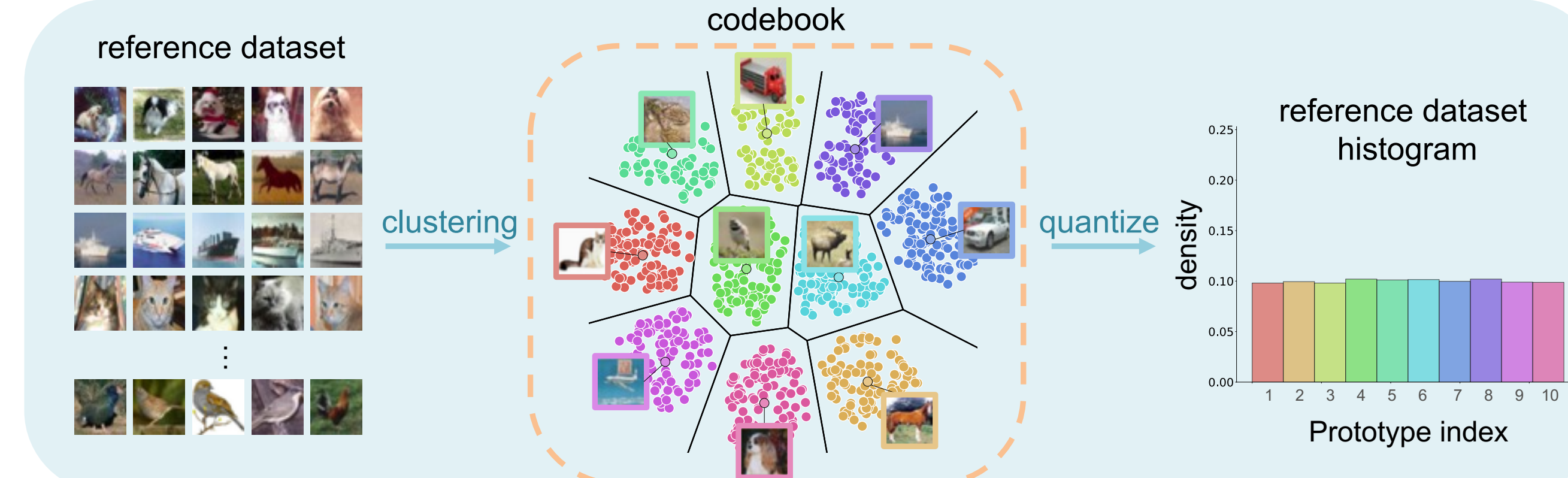


- **Image representation v.s. dataset representation:**

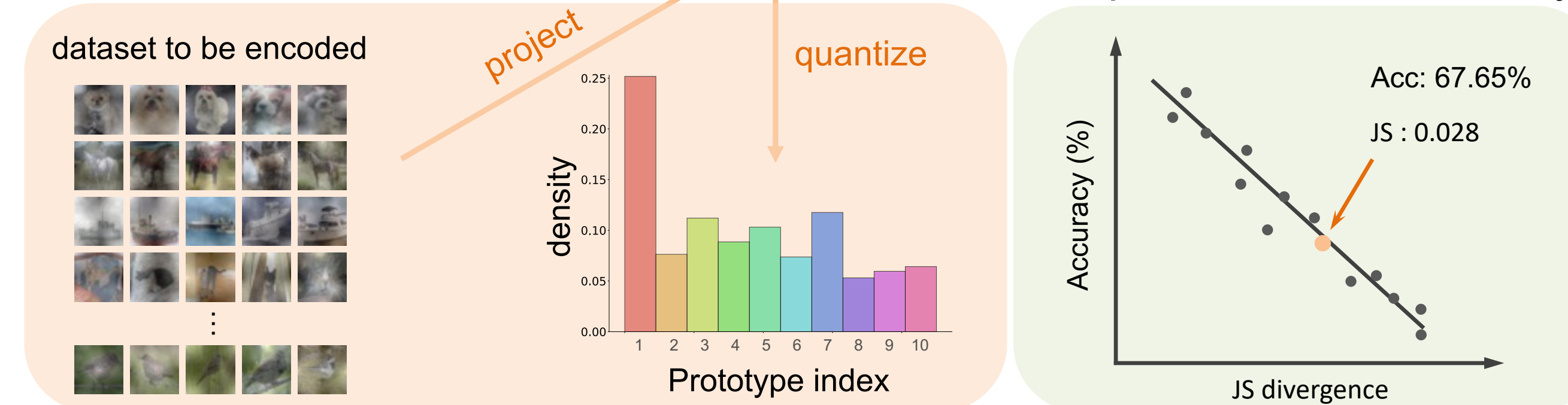


Method: Bag-of-Prototypes

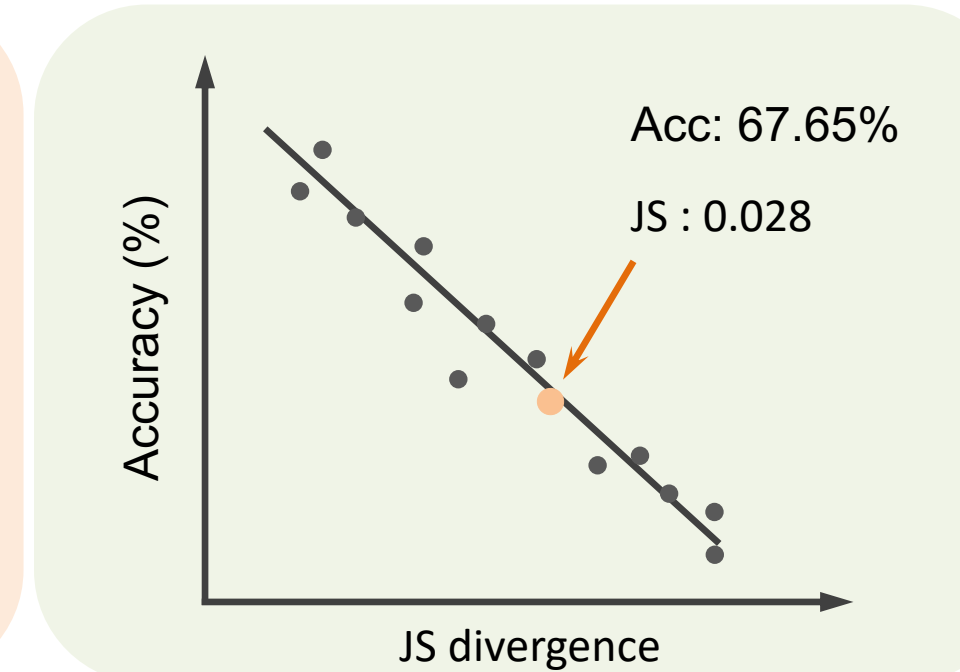
- Step I: codebook generation



- Step II: histogram computation

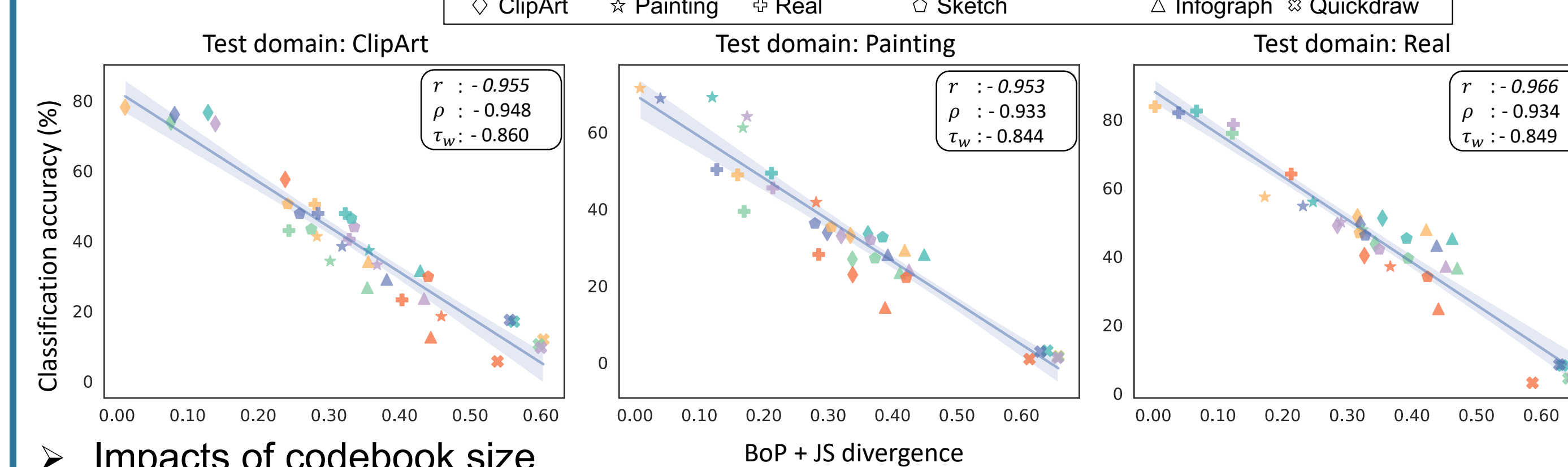


- Step III: measure similarity

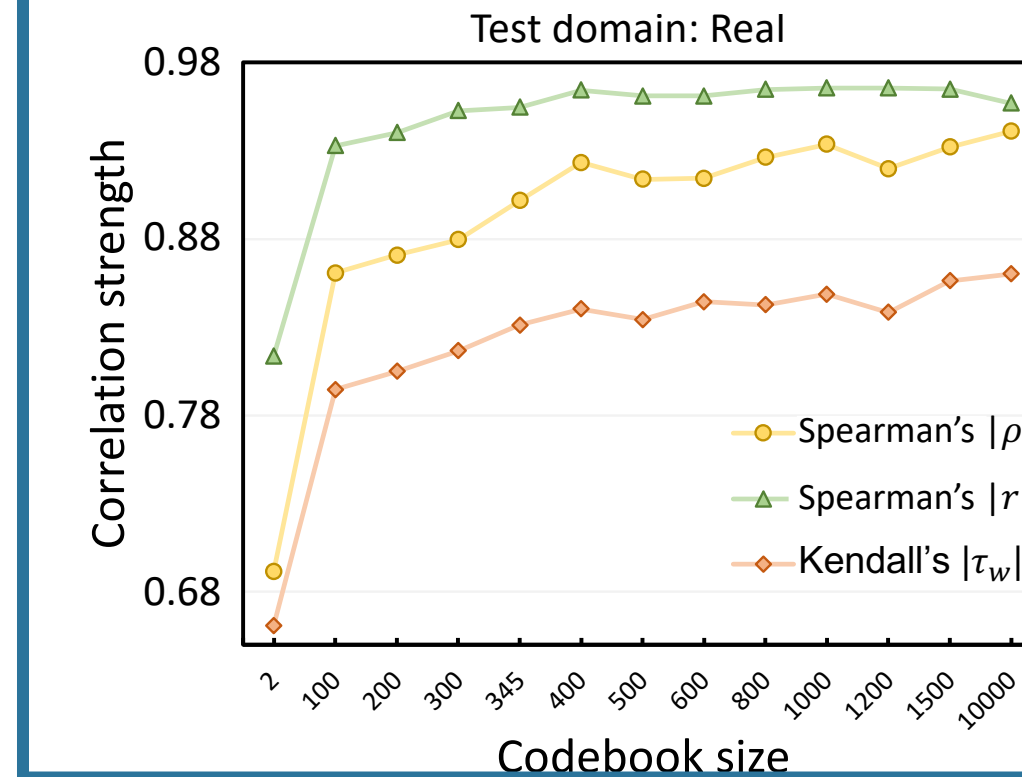


Experiment: Training set suitability

- Correlation study:



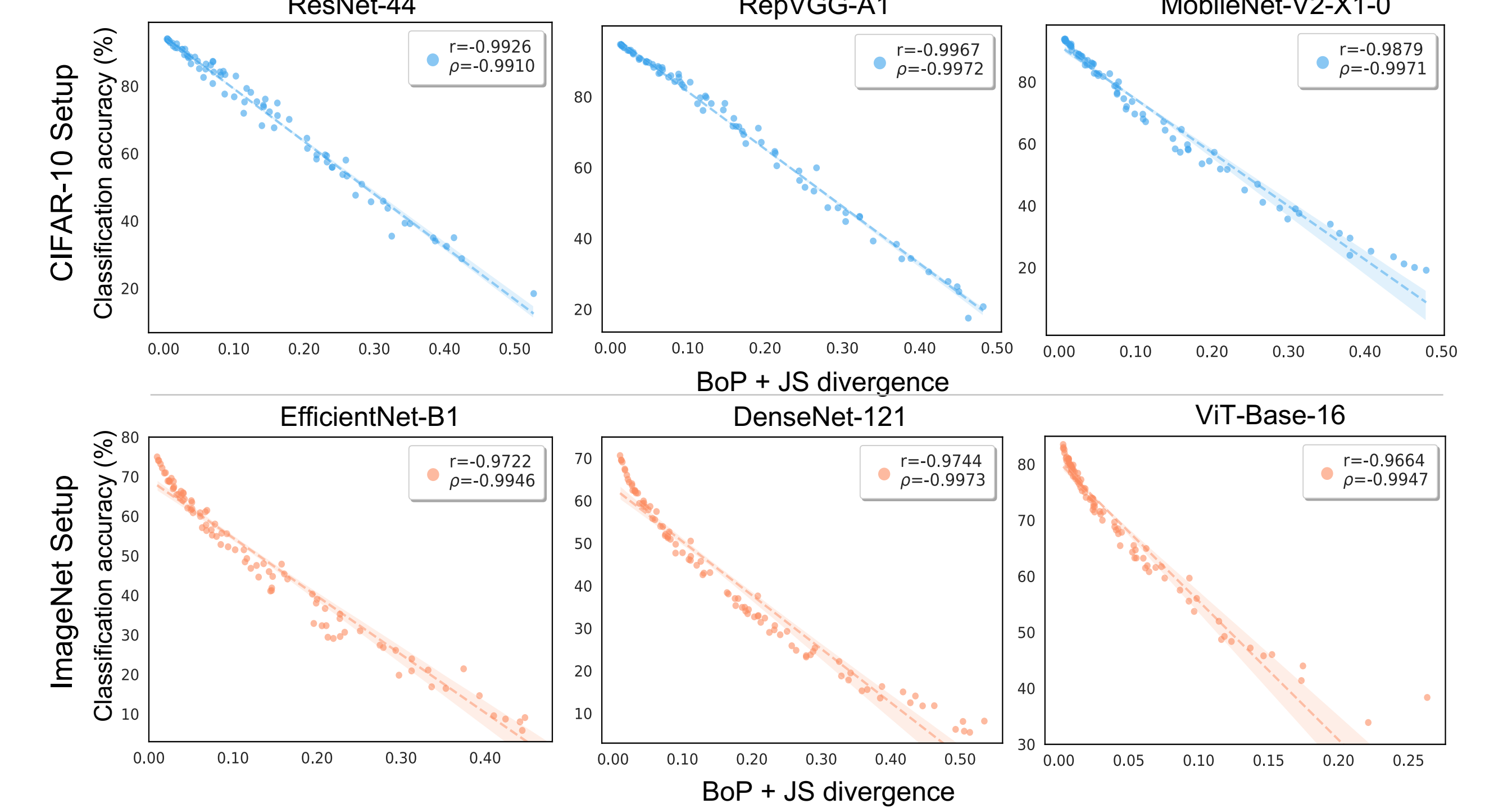
- Impacts of codebook size



- 1) BoP + JS well measures dataset-dataset similarity. It consistently exhibit **strong** correlations with classification accuracy on different test domains.
- 2) BoP is **stable and effective** using various codebook sizes. The correlation strength firstly increases and become stable when codebook ≥ 400 .

Experiment: Test set difficulty

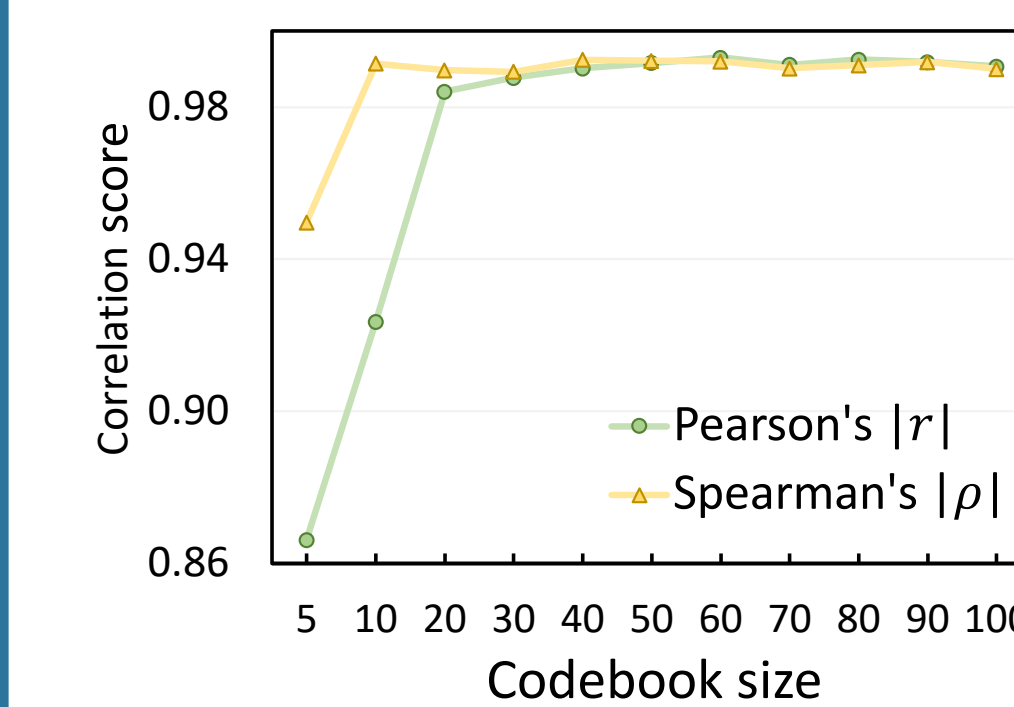
- Correlation study:



- Effectiveness of predicting accuracy (RSME):

Method	CIFAR-10.1	CIFAR-10.2	CIFAR-10.2- \bar{C} (50)					Overall
			Severity 1	Severity 2	Severity 3	Severity 4	Severity 5	
Prediction score ($\tau = 0.8$)	4.899	4.800	10.127	12.869	16.809	21.427	24.371	17.910
Prediction score ($\tau = 0.9$)	0.297	0.550	3.638	5.078	8.048	11.804	14.108	9.404
ATC-MC [30]	2.650	2.672	3.080	4.306	7.108	11.015	13.040	8.601
DoC [34]	0.490	0.263	2.247	2.916	5.117	9.012	6.637	5.744
$\mu + \sigma + FD$ [21]	0.455	0.561	5.875	5.823	4.724	4.908	6.486	5.602
BoP ($K = 80$)	0.218	0.122	2.458	2.818	3.730	5.836	6.451	4.551
BoP ($K = 100$)	0.186	0.124	2.849	2.808	3.548	4.025	4.777	3.663

- Impacts of codebook size:



- BoP + JS exhibits **strong** correlations for different models under CIFAR-10 and ImageNet setups.
- BoP is **superior** in predicting accuracy than others.
- BoP is **stably effective** using various codebook sizes.

Code and Data

<https://github.com/Klaus-Tu/Bag-of-Prototypes>



SCAN ME FOR CODES