Few-Class Arena: A Benchmark for Efficient Selection of Vision Models and Dataset Difficulty Measurement

Bryan Bo Cao, Lawrence O'Gorman, Michael Coss, Shubham Jain

Stony Brook University

NOKIA BELL LABS

ICLR

github.com/bryanbocao/fca

Code

SCAN ME

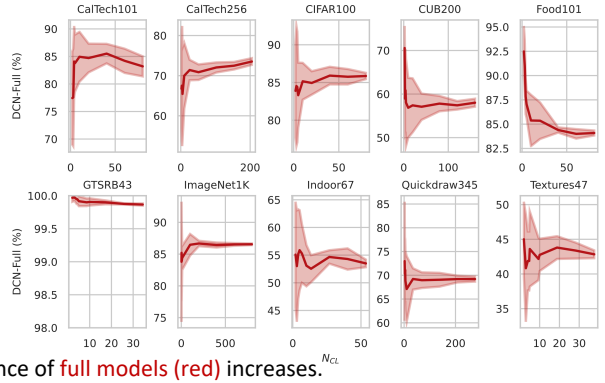
Goal

Few-Class Arena (FCA) is a benchmark tool designed to facilitate the selection of efficient neural network and research in the **few-class regime**, i.e., **number of classes < 10** (N_{CL} < 10).

- Propose a **one-time** sub-class dataset difficulty measurement.
- Unveil new insights in the **few-class regime.**

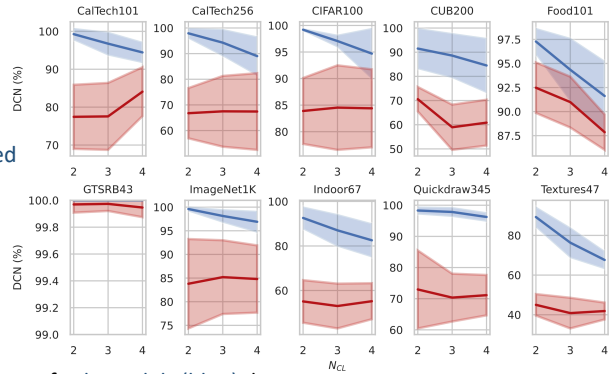
Motivation

Limitations of Existing many-class Benchmarks (red)



- The variance of full models (red) increases.
- This large variance indicates high uncertainty in efficient model selection.

Our proposed Few-Class Arena (FCA) (blue)

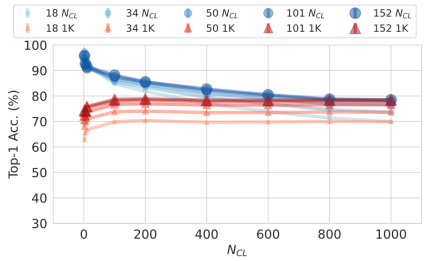


- The variance of sub-models (blue) decreases.
- Sub-model accuracy increases.

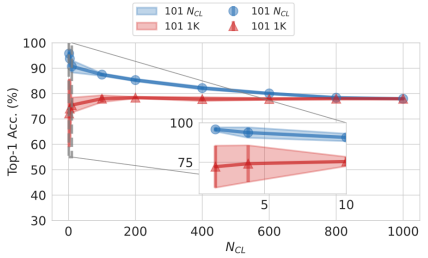
Lack of few-class benchmarks designed for sub-models.

DCN: dataset classification difficulty number (DCN) is the empirically highest accuracy achieved in a dataset. Acc.: Accuracy. Var.: Variance. Corr.: correlation.

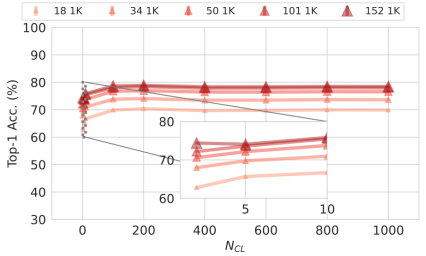
Higher Acc. & Lower Var. of Sub-models



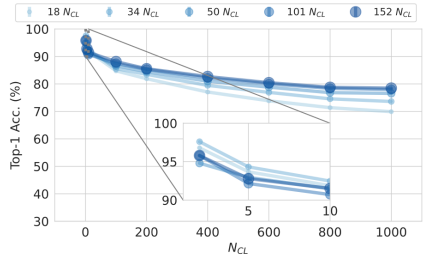
Sub-models attain higher upper-bound accuracy than full models.



Full model accuracy range widens. Sub-models narrow the accuracy range.



Full models follow the scaling law in model size.



The scaling law is violated for sub-models.

Proposed Similarity-Based Silhouette Score (SimSS)

$$SimSS^{(D)} = \frac{1}{|L| \times |C_l|} \sum_{i \in C_l} \frac{S_\alpha(i) - S'_\beta(i)}{max(S_\alpha(i), S'_\beta(i))}$$

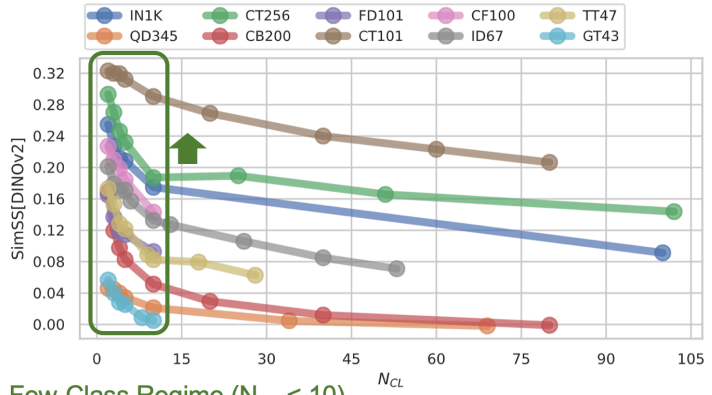
SimSS quantifies the cluster-like characteristics of a dataset to facilitate efficient vision model design and selection.

S_α : Intra-Class Similarity represents the **tightness** of the class

S'_β : Nearest Inter-Class Similarity represents the **closeness** to other classes.

MT: Model type. N_{CL}: Number of classes. Det: Object Detection. Seg: Segmentation.
F (Full model): models pre-trained on the full many classes. S (Sub-model): models trained on the sub classes (N_{CL} < 10).

SimSS in the Few-Class Regime



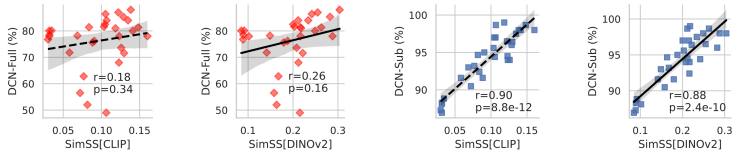
Few-Class Regime (N_{CL} < 10)

The importance of image similarity increases.

Extension to Detection & Segmentation

MT	N_{CL}	Det mAP@50↑	Det STDEV↓	Seg mAP@50↑	Seg STDEV↓
F	80	0.405	0.195	0.378	0.200
F	5	0.456	0.090	0.435	0.098
FT	5	0.488	**0.069**	0.465	0.082
S	5	**0.503**	**0.069**	**0.474**	**0.084**
F	2	0.488	0.161	0.475	0.180
FT	2	0.505	0.127	0.457	0.159
S	2	**0.538**	**0.106**	**0.482**	**0.152**

Summary — High Corr. SimSS & Acc.



Our Similarity-based score (SimSS) displays **high correlation** (r ≥ 0.88) with model accuracy for the purpose of a practitioner quickly comparing models.