Scalable Evaluation and Neural Models for Compositional Generalization

Giacomo Camposampiero
IBM Research – Zurich, ETH Zurich
giacomo.camposampiero1@ibm.com

Pietro Barbiero
IBM Research – Zurich
pietro.barbiero@ibm.com

Michael Hersche
IBM Research – Zurich
michael.hersche@ibm.com

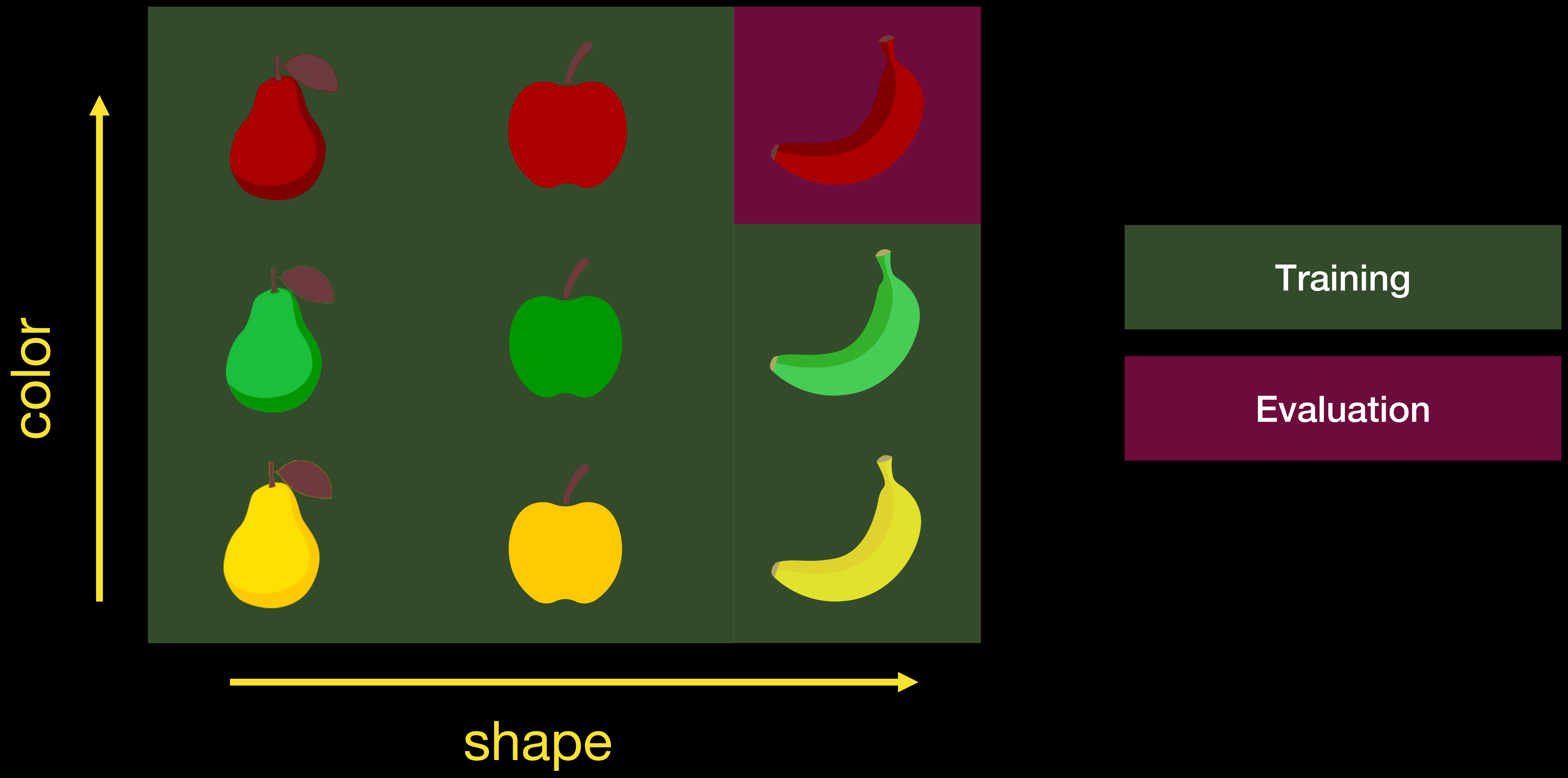
Roger Wattenhofer
ETH Zurich
wattenhofer@ethz.ch

Abbas Rahimi
IBM Research – Zurich
abr@zurich.ibm.com

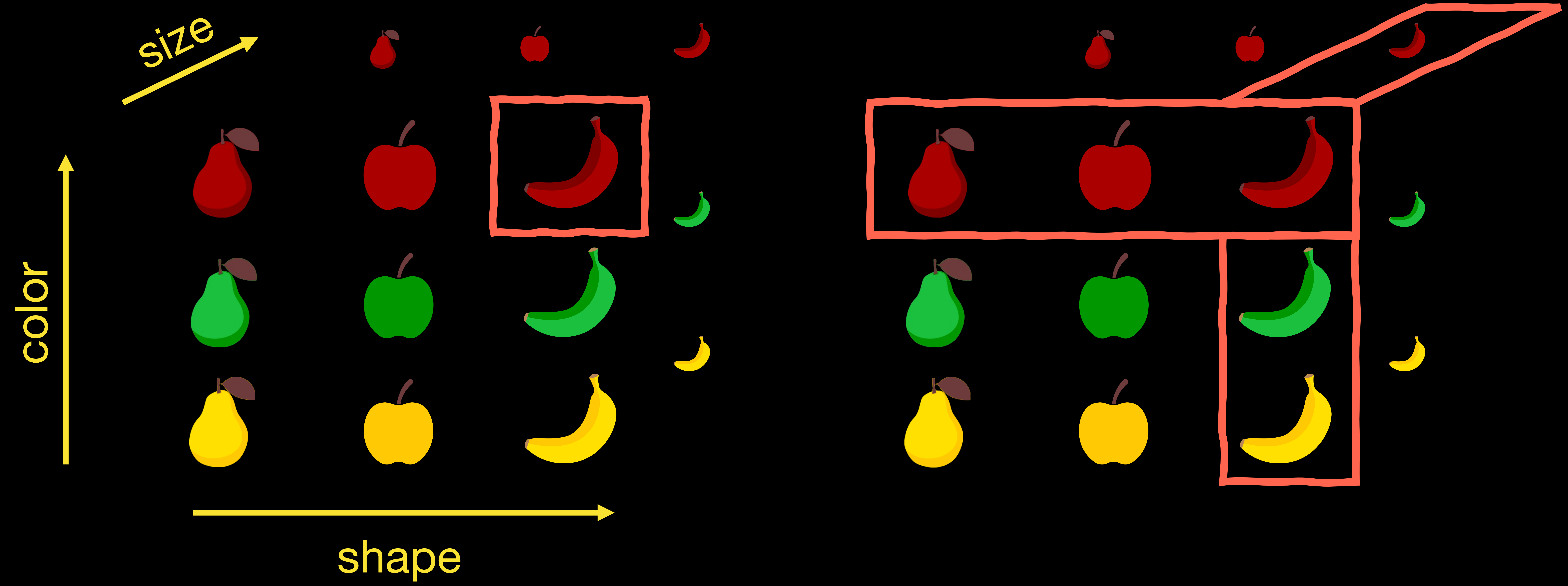
IBM **Research** ETH zürich

NEURAL INFORMATION PROCESSING SYSTEMS

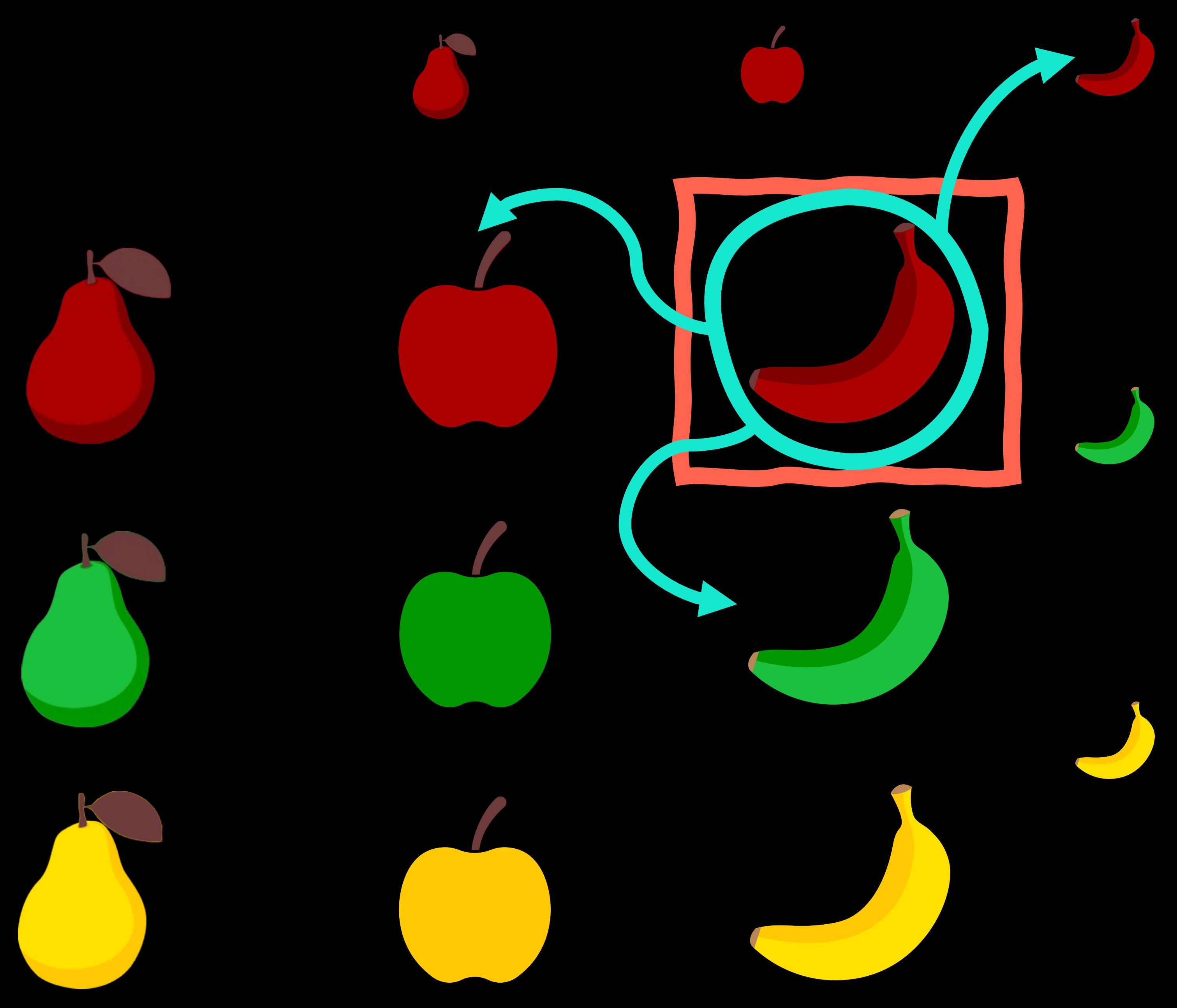
We consider the problem of compositional evaluation.



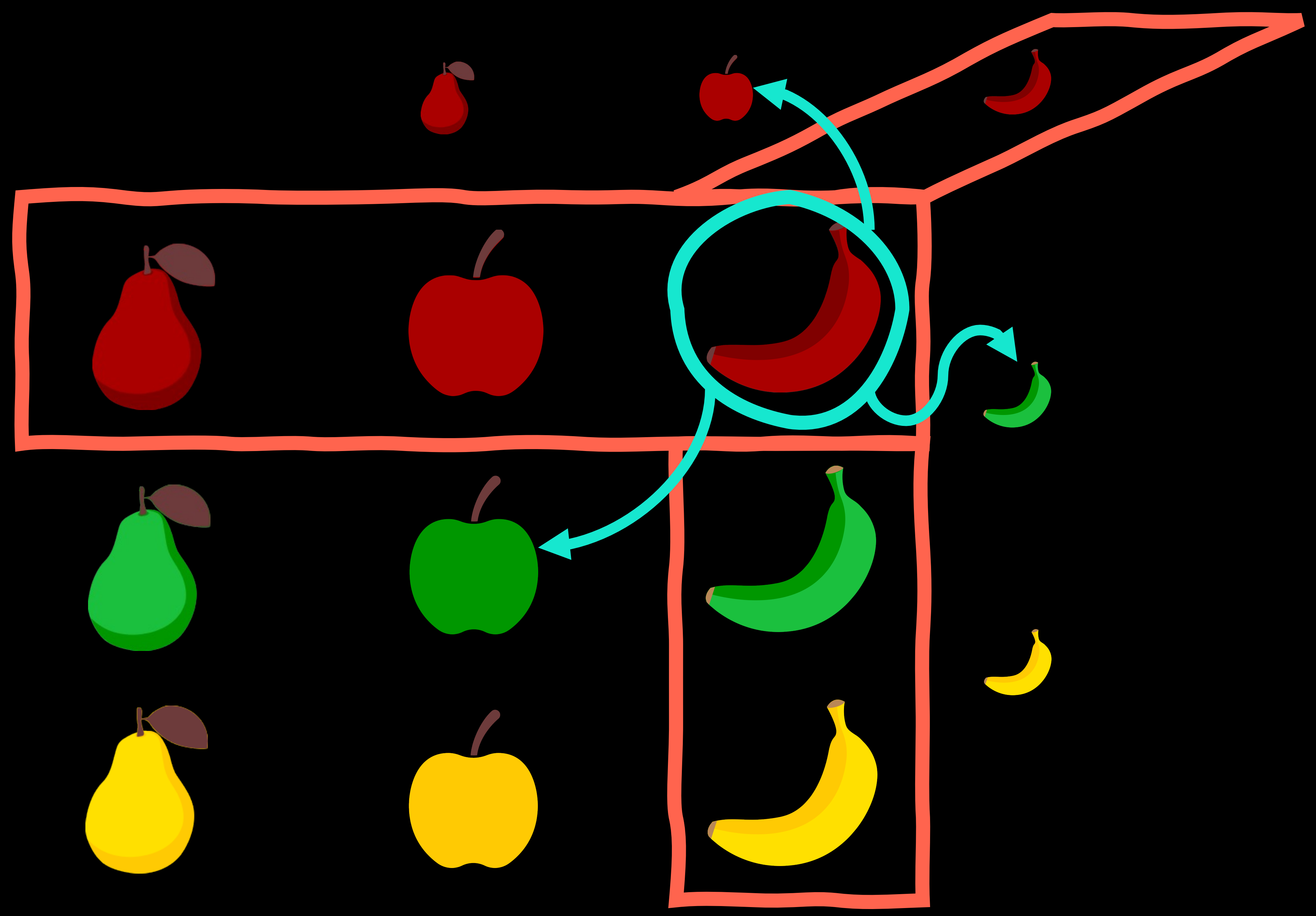
Previous works evaluated compositional generalization under significantly heterogeneous assumptions, making comparability hard.



Previous works evaluated compositional generalization under significantly heterogeneous assumptions, making comparability hard.

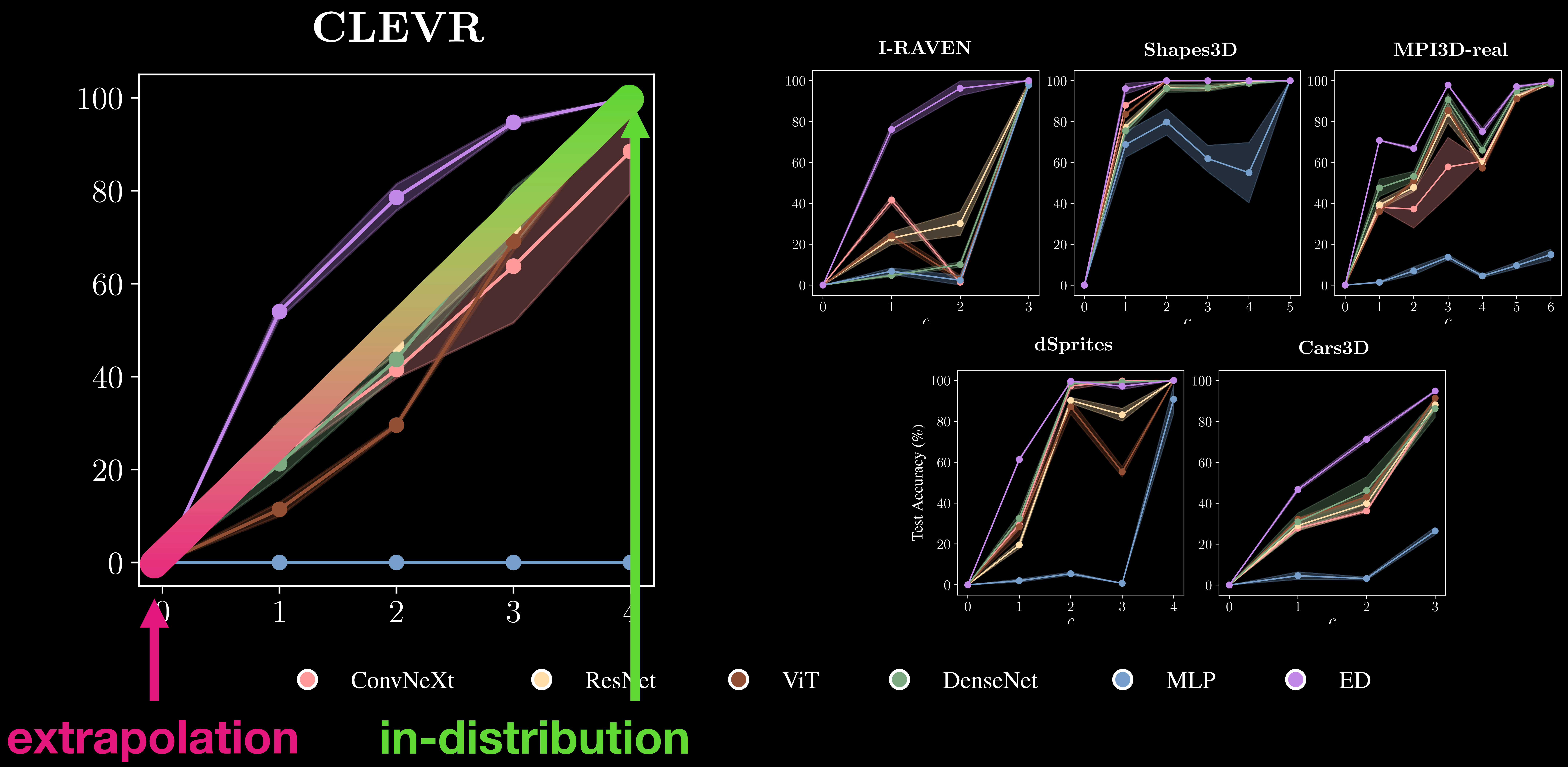


semantic similarity $(c) = 2$

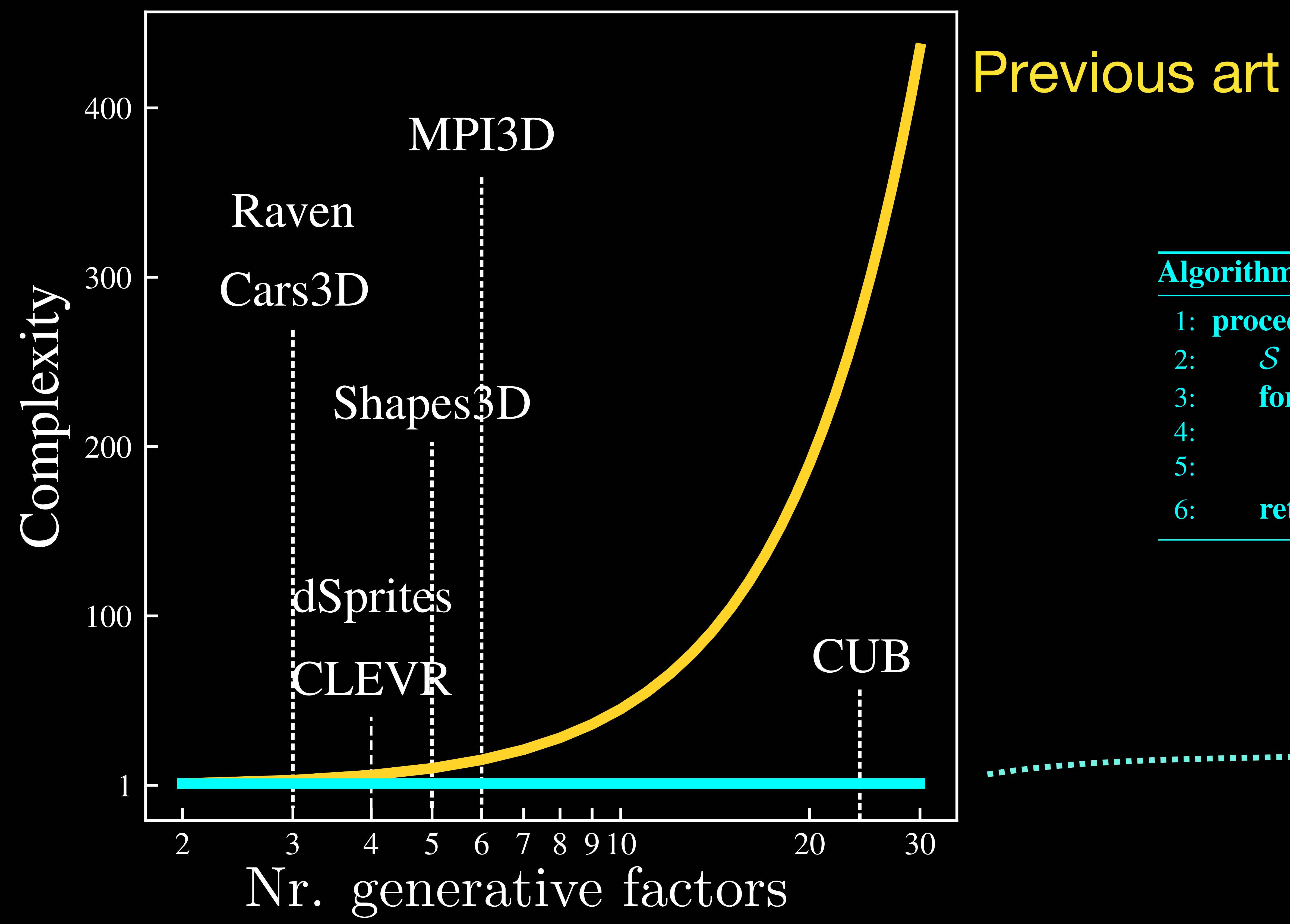


semantic similarity $(c) = 1$

Orthotopic evaluation framework allows rigorous compositional generalization testing.



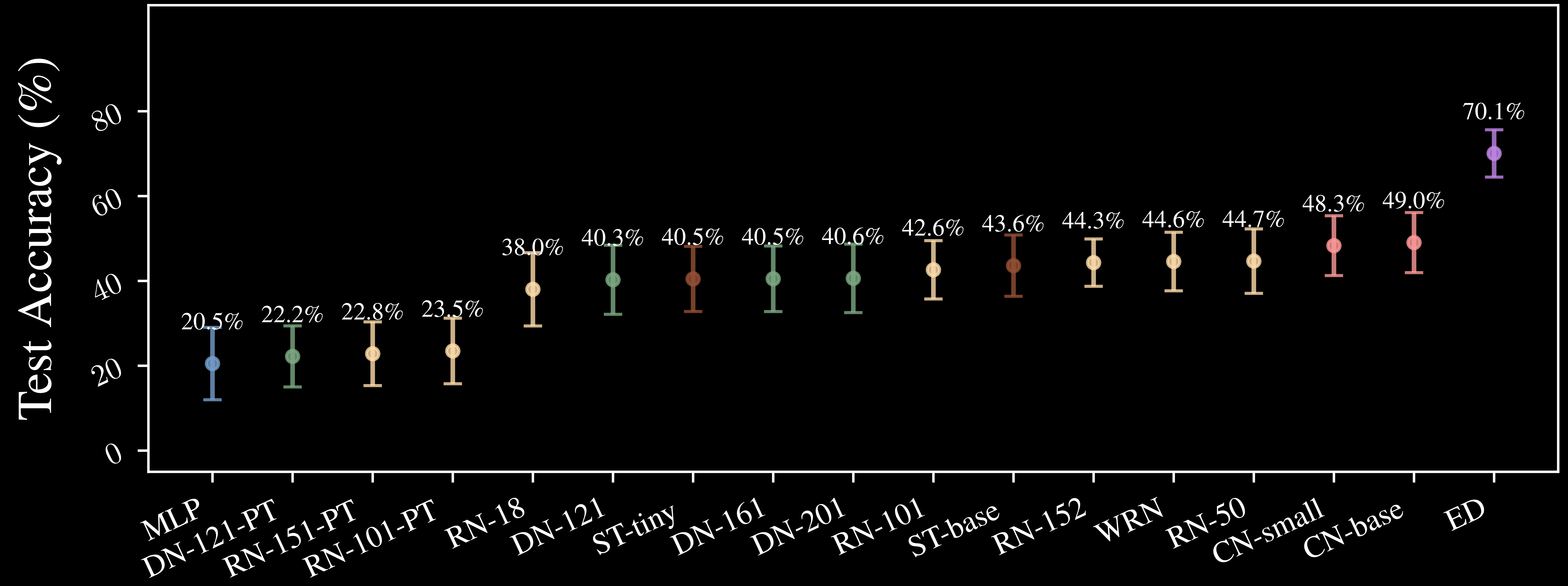
Orthotopic evaluation framework allows efficient compositional generalization testing.



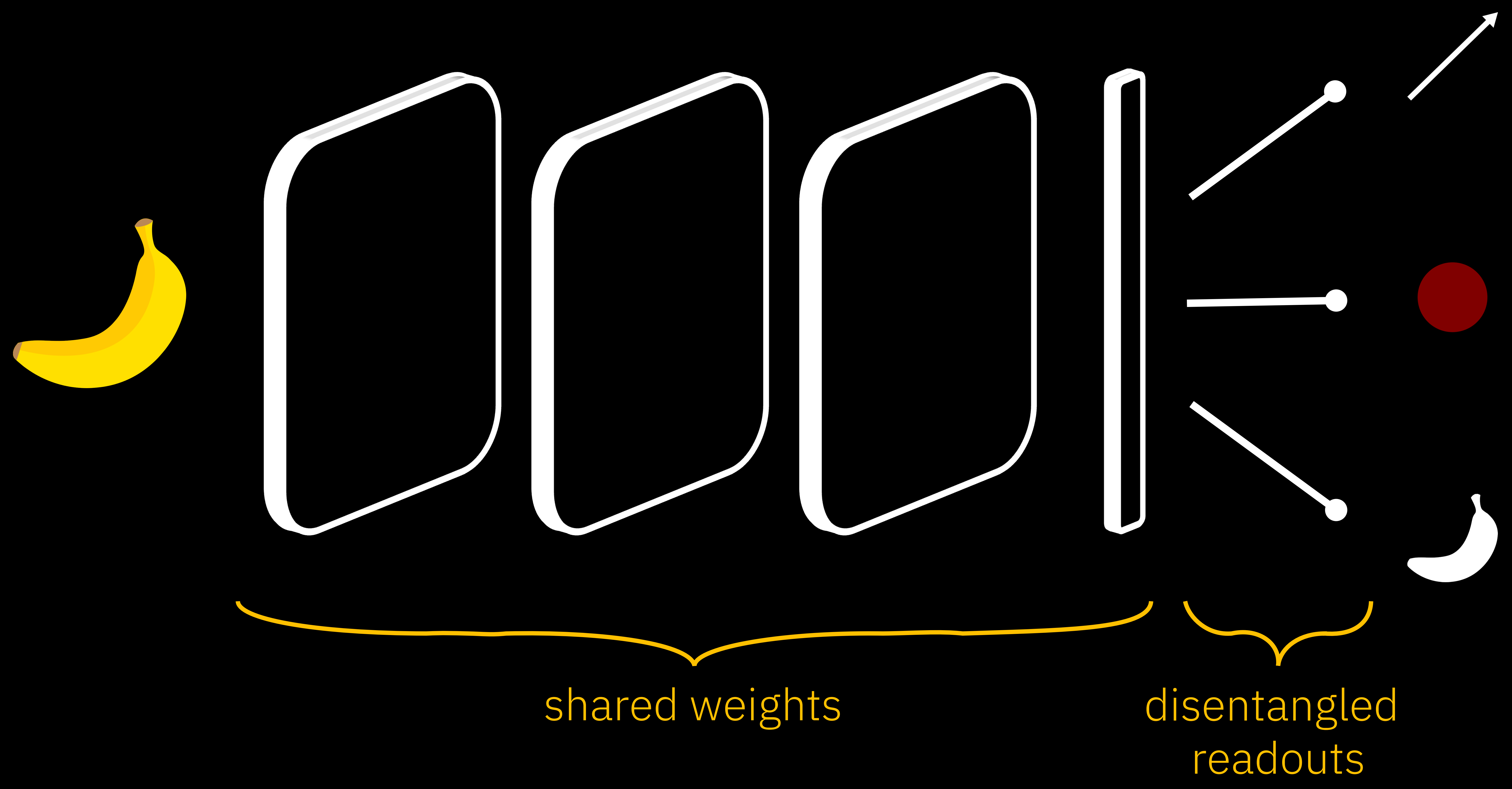
Algorithm 1 Orthotopic split generation.

procedure SPLITDATASET($\mathbf{X}$, $\mathbf{G}$, c)
    $\mathcal{S} = \binom{\mathbf{G}}{c+1} = \{\mathbf{S} \subseteq \mathbf{G} : |\mathbf{S}| = c+1\}$ ▷ Get all combinations of c factors in $\mathbf{G}$
    for each $s \in \mathcal{S}$ **do** ▷ Iterative orthotope pruning
        $\mathbf{X}_{proj} = \mathbf{proj}_s \mathbf{X}_{train}$ ▷ Project $\mathbf{X}$ onto the c-dimensional subspace s
        $\mathbf{X}_{train} = \mathbf{X}_{train} \setminus (\mathbf{X}_{proj} \cap \text{exclusion}(s))$ ▷ Pruning operation in the subspace s
    return $\mathbf{X}_{train}, \mathbf{X}^{\complement}_{train}$

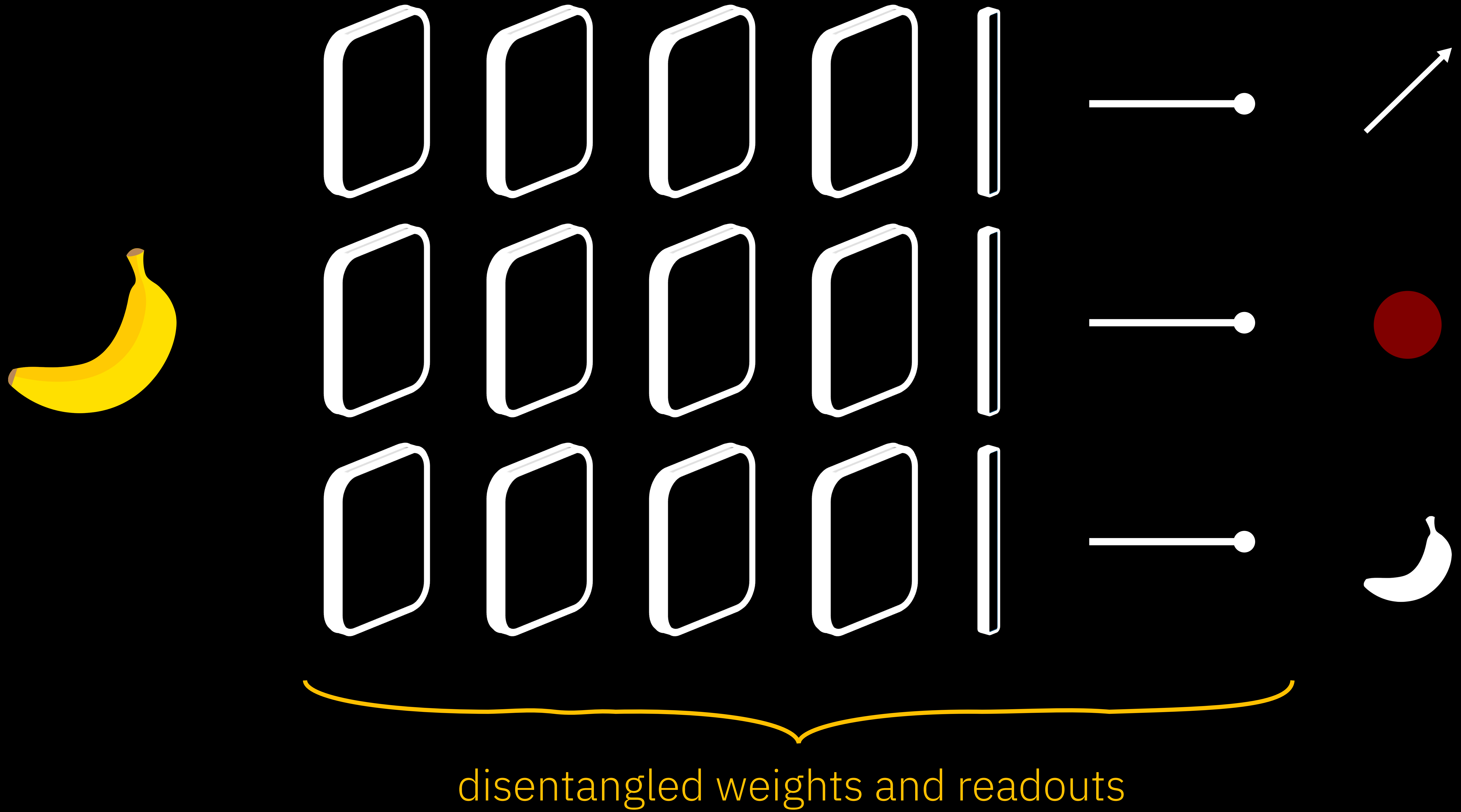
We provide an updated picture of compositional generalization in SOTA models, training more than 5000 models in total.



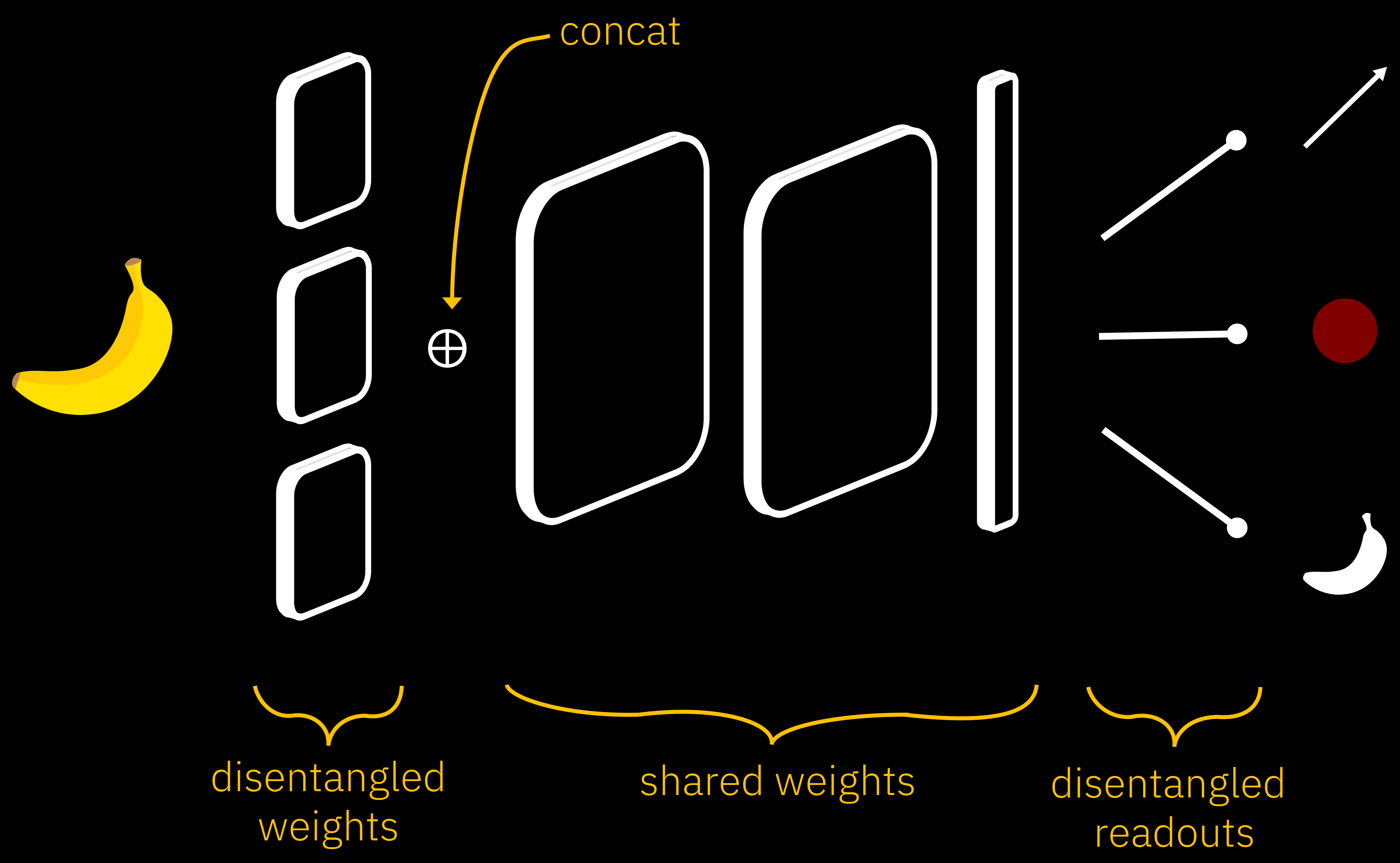
Monolithic networks are parameter-efficient but do not generalize well.



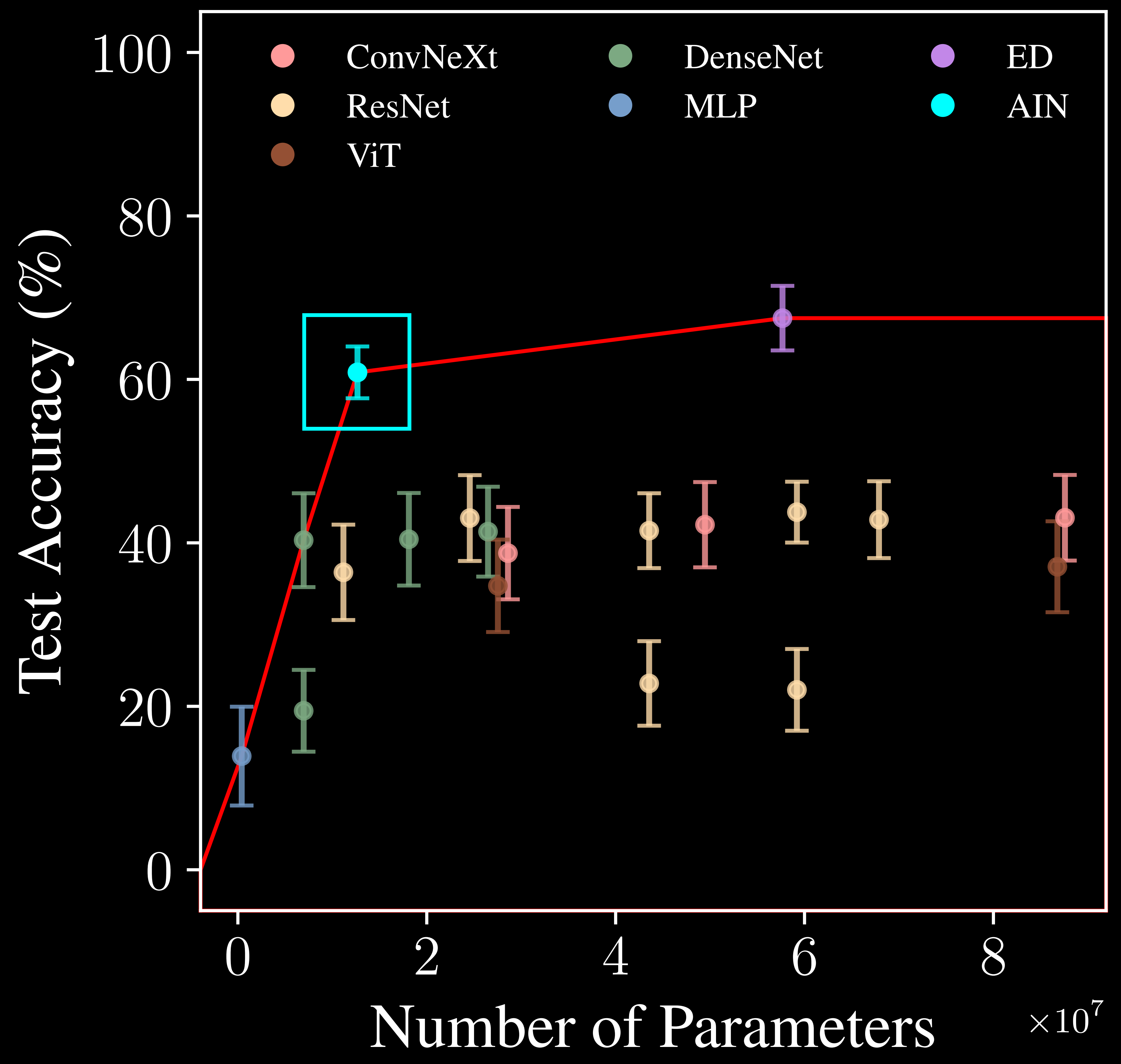
Disentangled networks generalize well but are very parameter-inefficient.



The solution? Attribute Invariant Networks generalize well and are parameter efficient!



AIN achieve Pareto optimality in compositional generalization.



TL;DR: Orthotopic evaluation makes measuring compositional generalization faster and more rigorous. Attribute Invariant Networks achieved the Pareto-optimality in this task.

github.com/IBM/scalable-compositional-generalization