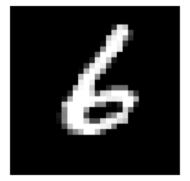
Are Labels Necessary for Neural Architecture Search?

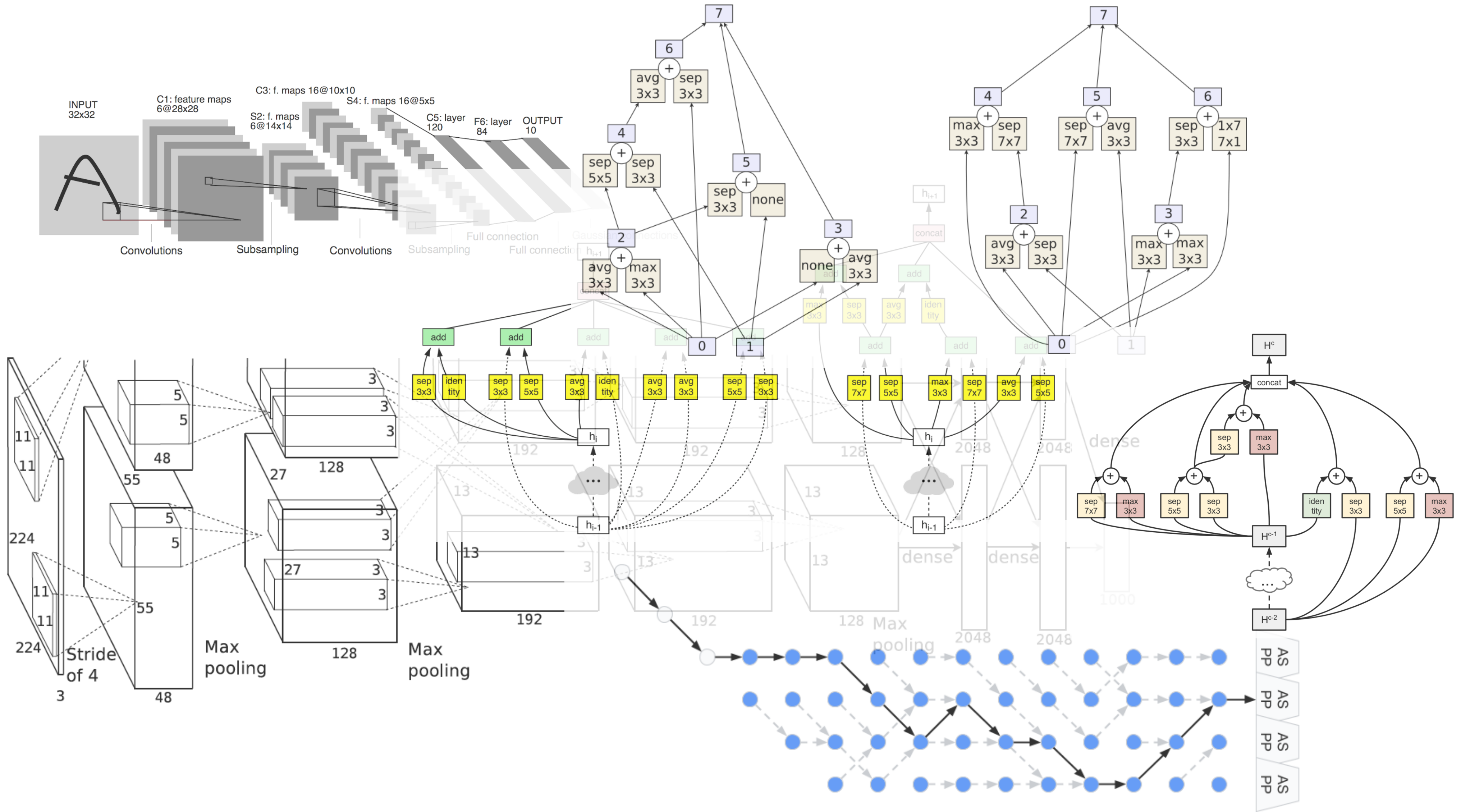
Chenxi Liu, Piotr Dollár, Kaiming He, Ross Girshick, Alan Yuille, Saining Xie

Spotlight @ECCV 2020

(Long Video)

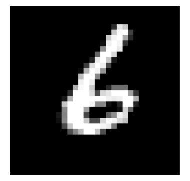
Designing neural architectures

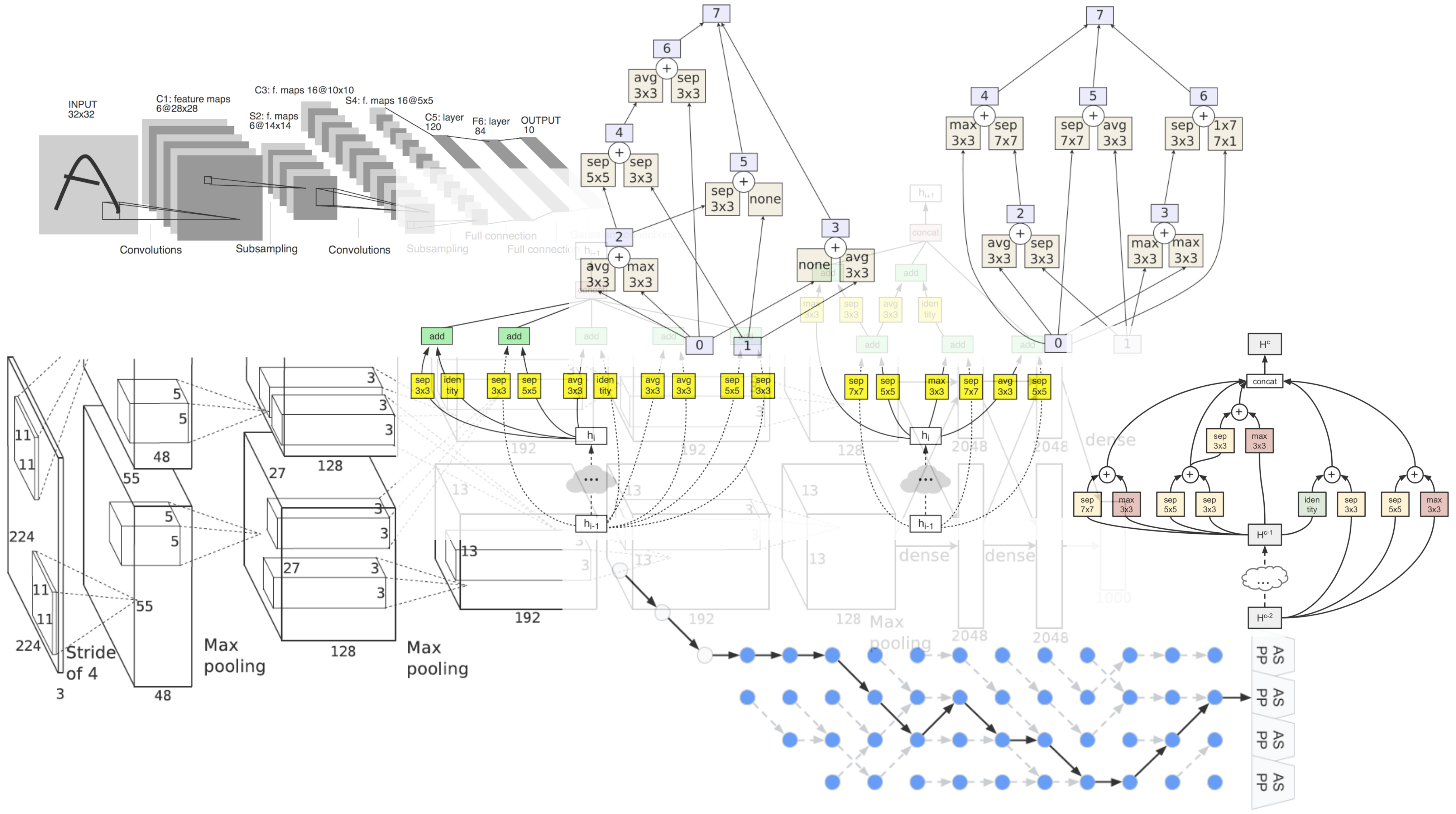
( , 6) ( , ship) ( , panda)



In Artificial Intelligence, Neural Architecture Search has always been supervised…

Designing neural architectures

( , 6) ( , ship) ( , panda)

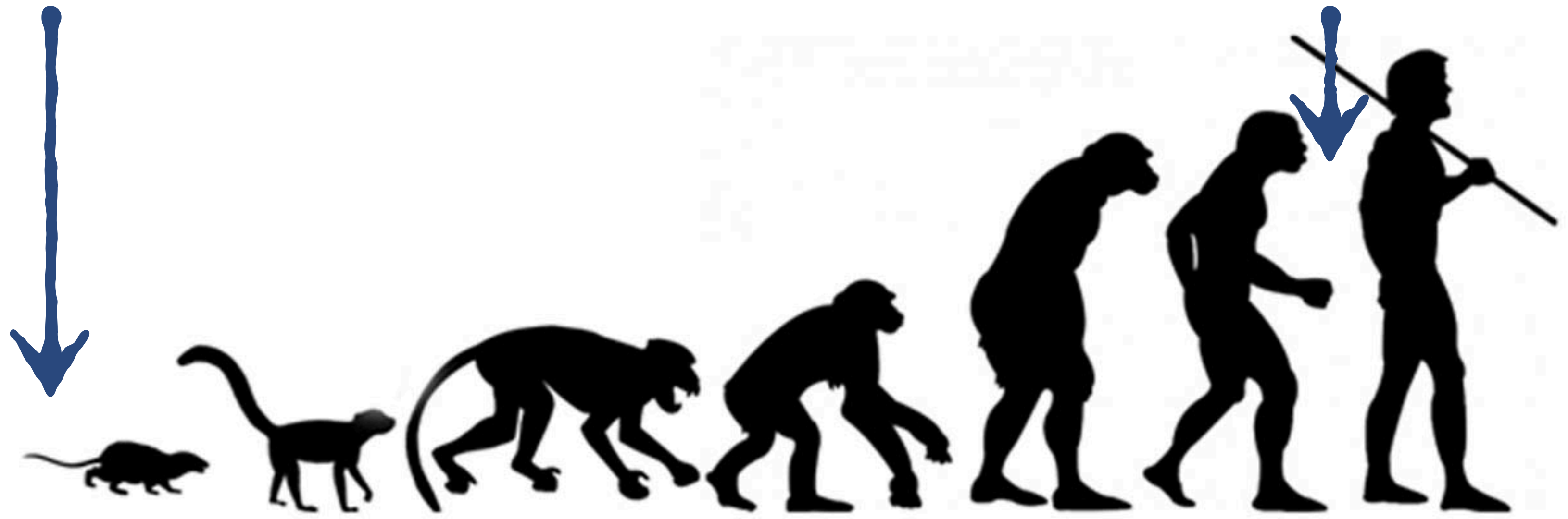


Artificial Intelligence

NATURAL INTELLIGENCE

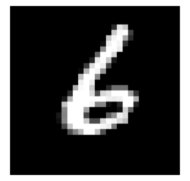
"Nature Architecture Search" started

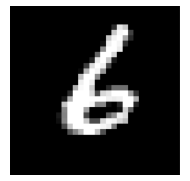
Semantic labels emerged



But in Natural Intelligence, "Nature Architecture Search" started long before semantic labels

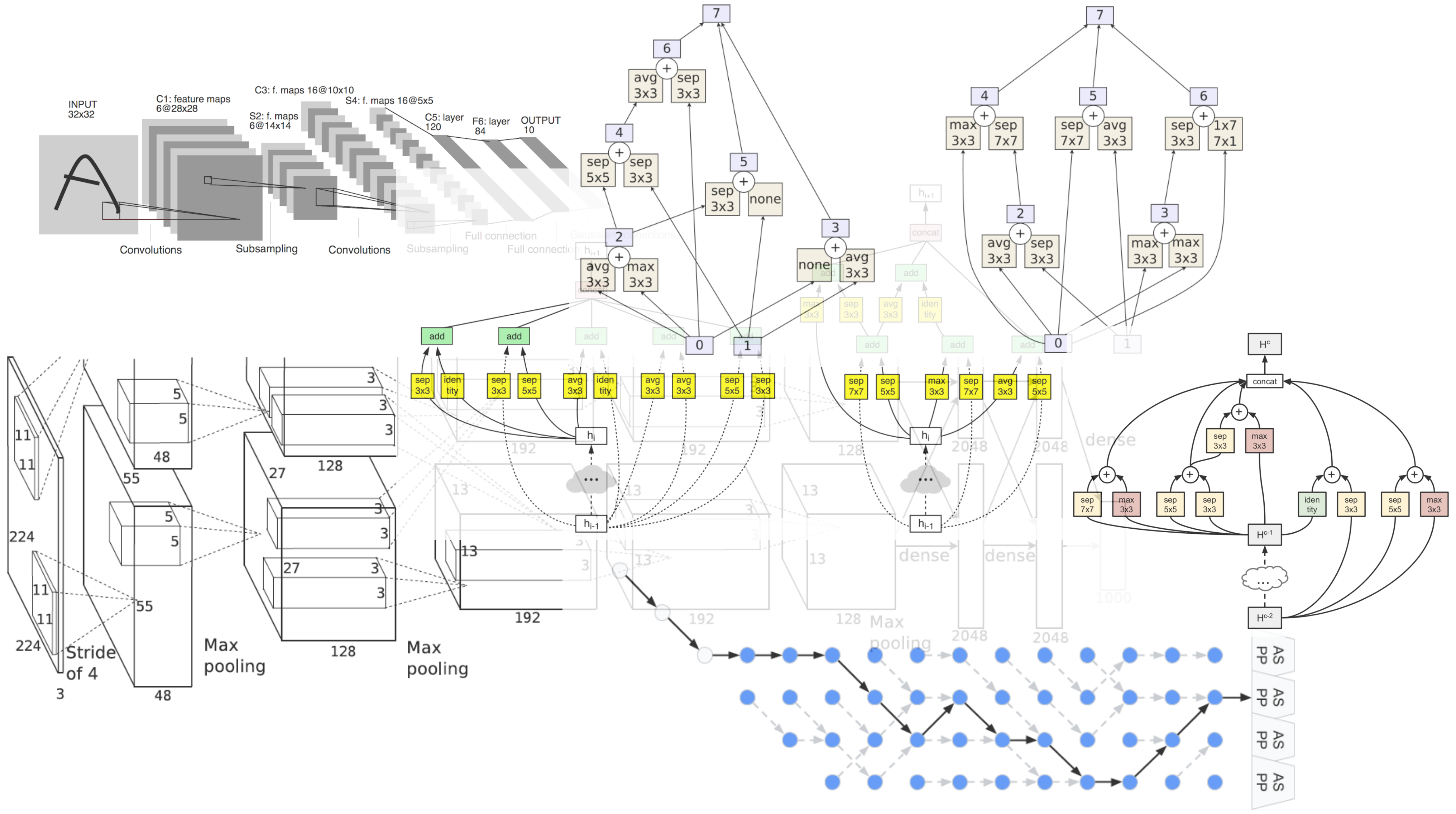
Designing neural architectures

( , 6) ( , ship) ( , panda)

( , ~~6~~) ( , ~~ship~~) ( , ~~panda~~)

Status quo

OURS



What neural architectures will we find?

Will they *look similar* to those on the left?

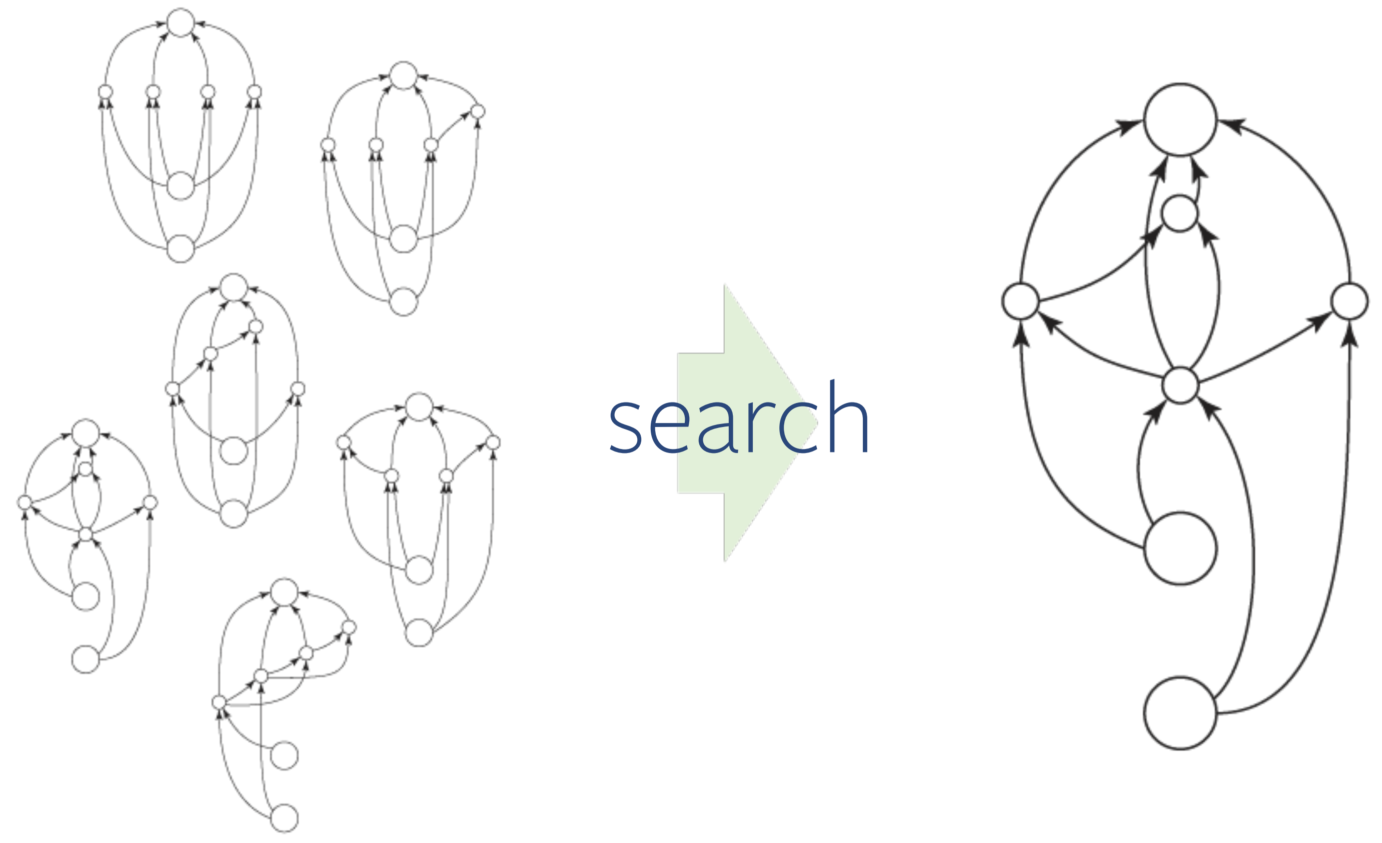
Will they *work as well* as those on the left?

: Supervised

Defining Unsupervised NAS (UnNAS)

Search Phase

(Supervised) NAS



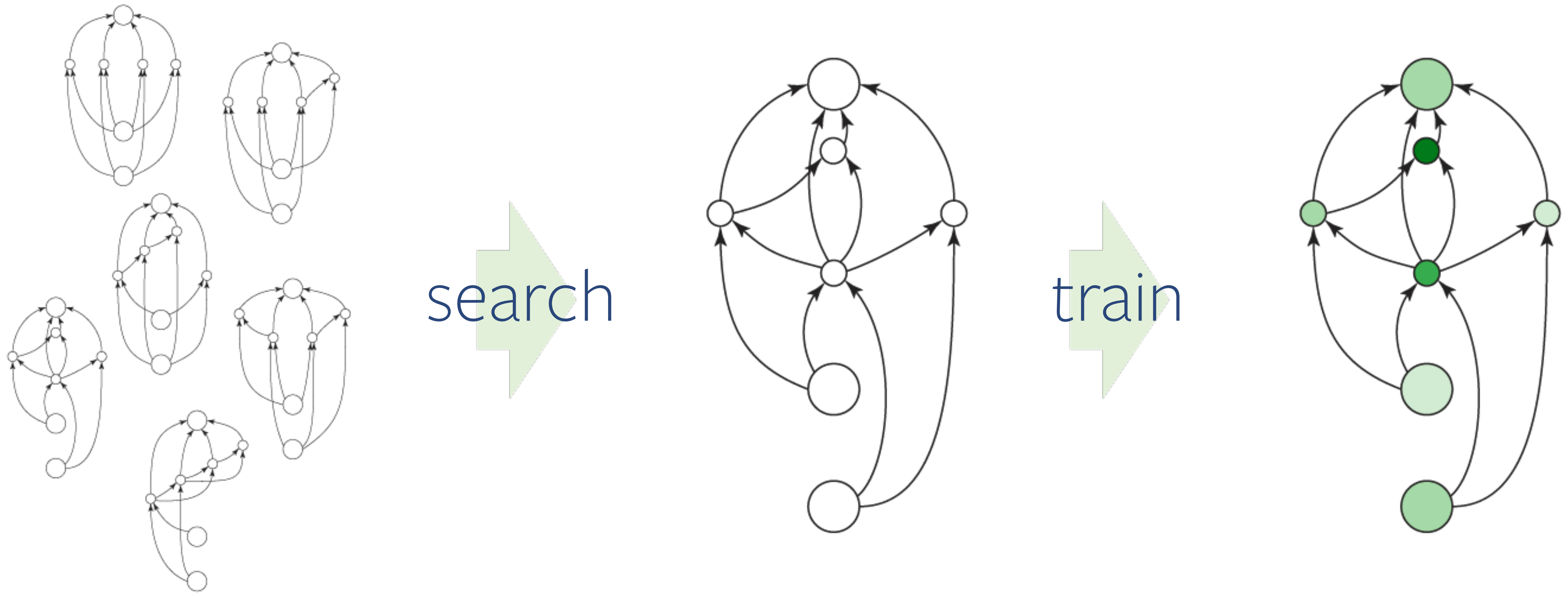
: Supervised

Defining Unsupervised NAS (UnNAS)

Search Phase

Eval Phase

(Supervised) NAS



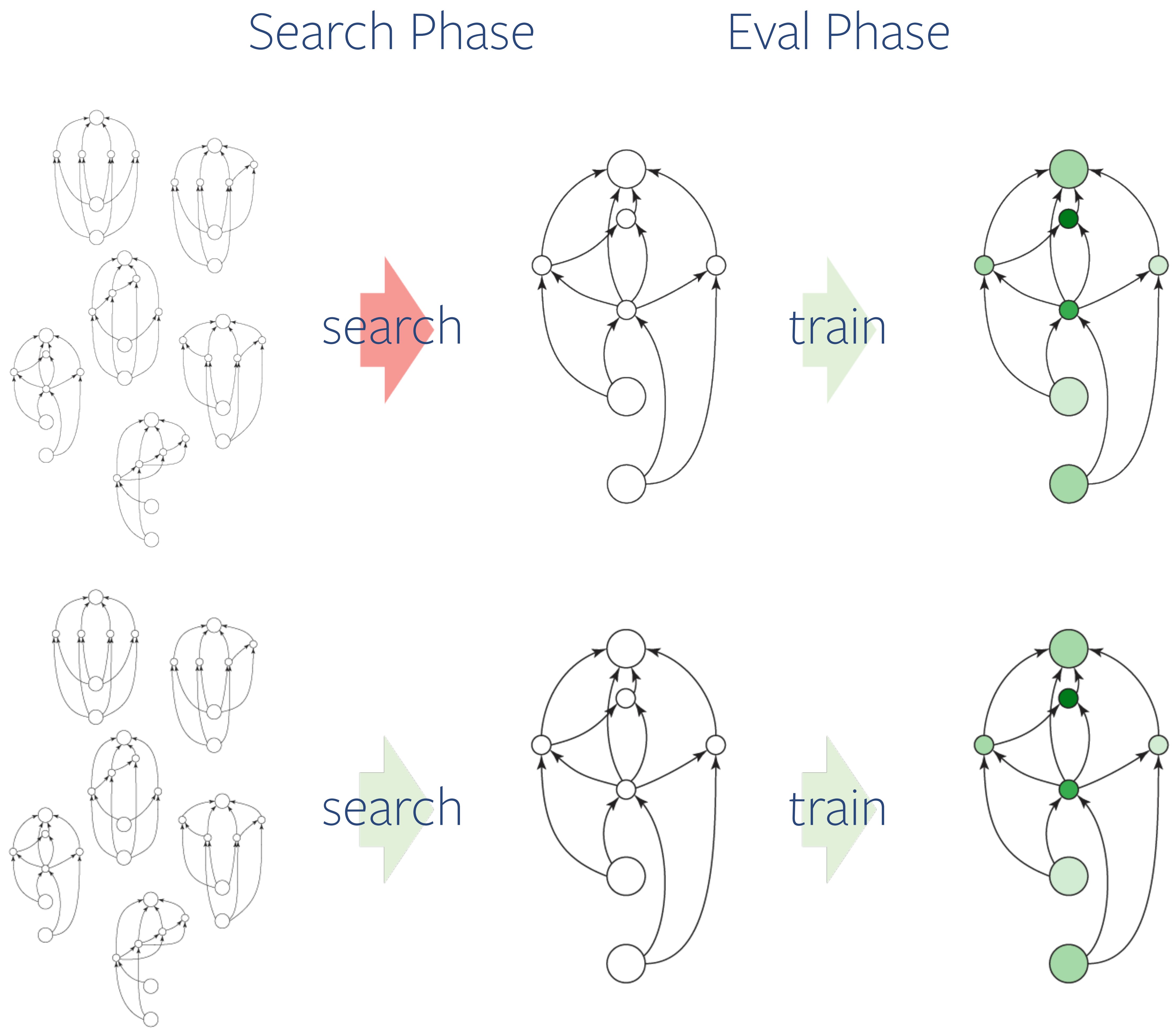
Defining Unsupervised NAS (UnNAS)

: Supervised

: Unsupervised

Unsupervised NAS (ours)

(Supervised) NAS



Defining Unsupervised NAS (UnNAS)

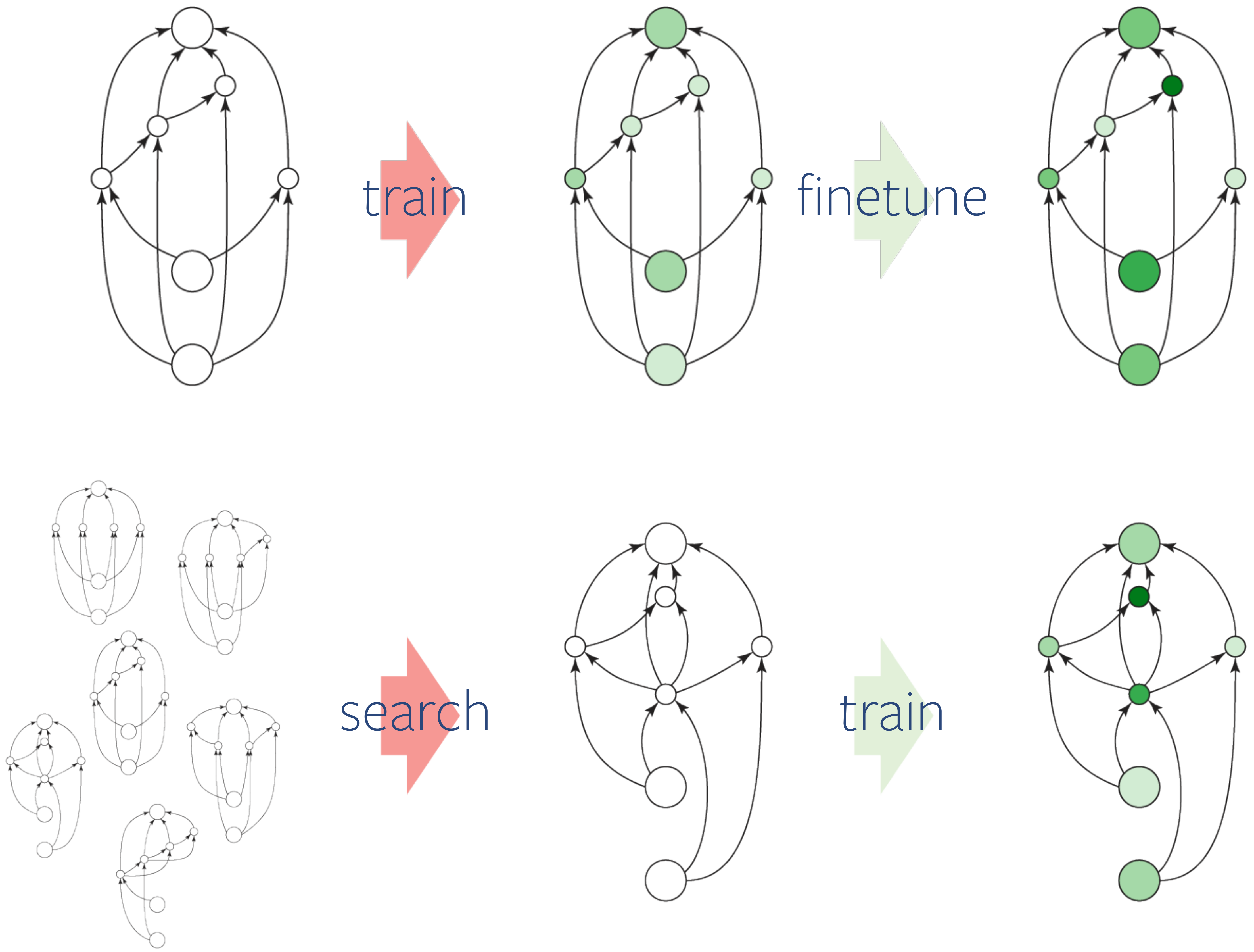
: Supervised

: Unsupervised

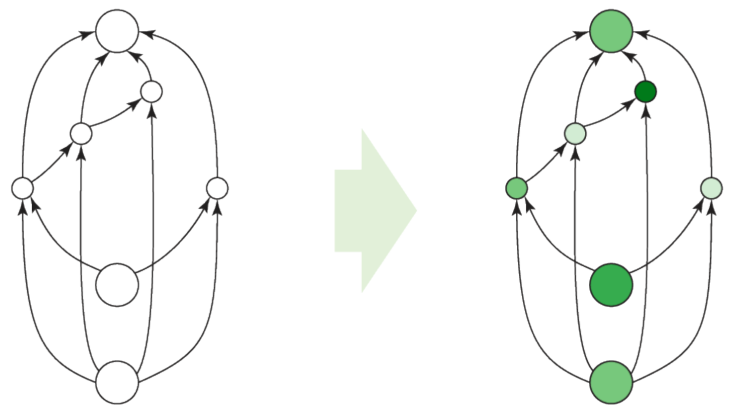
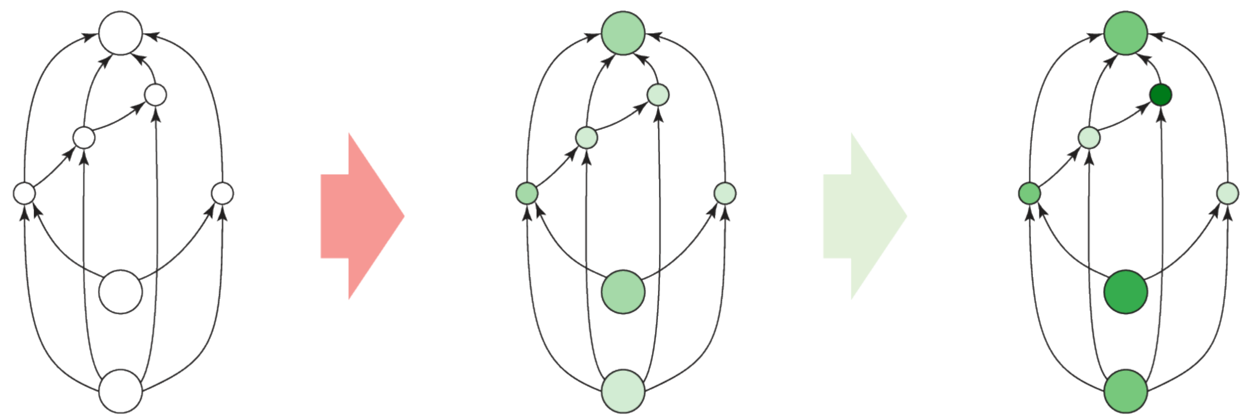
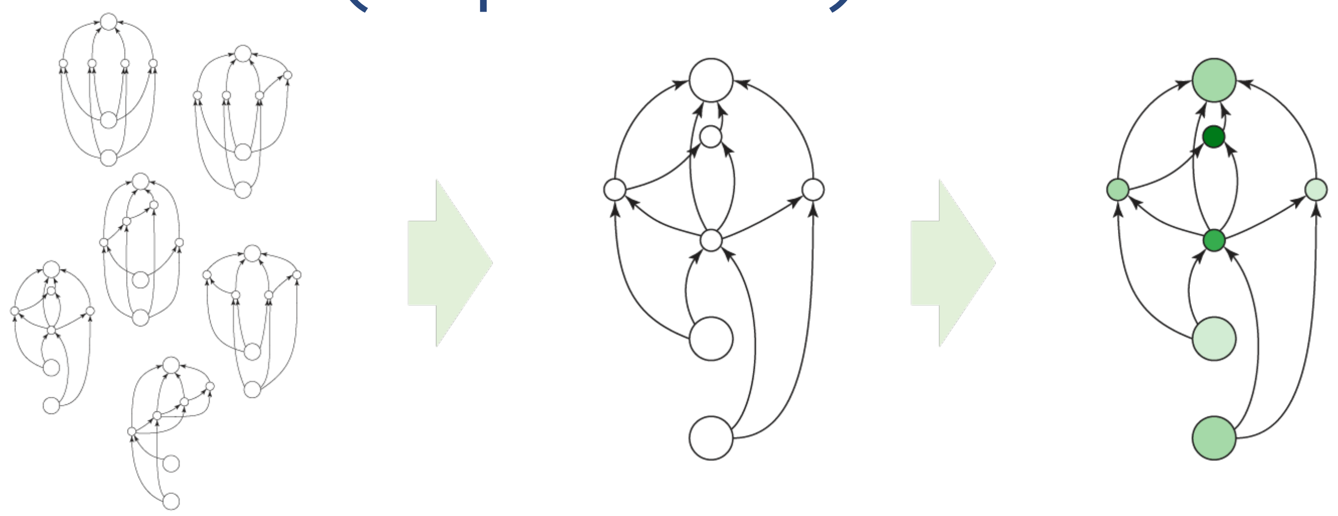
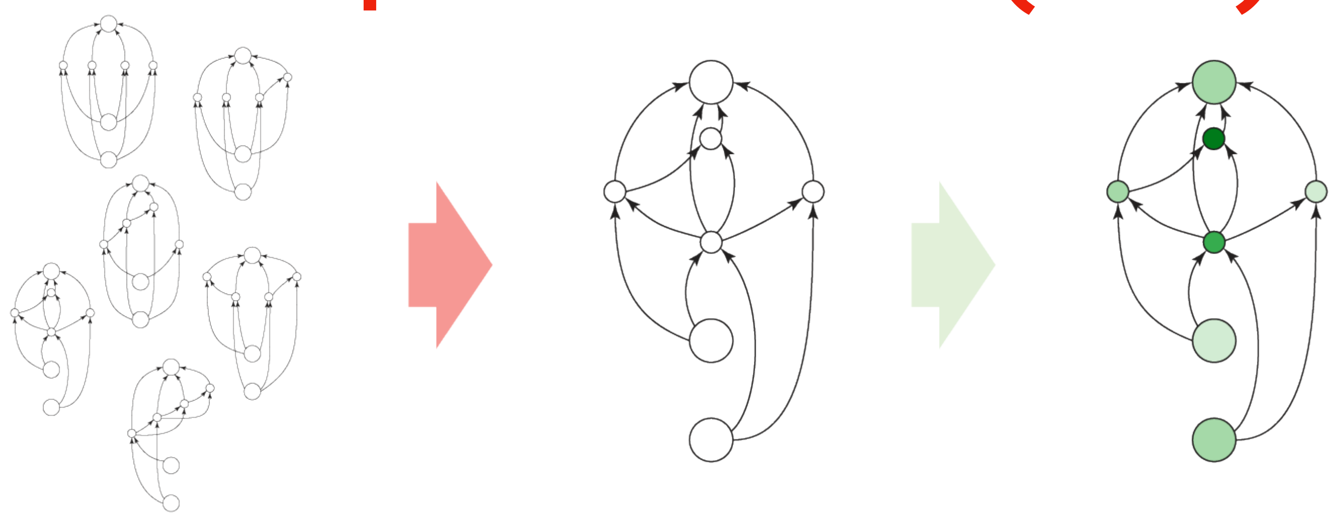
Search/Training Phase   Eval Phase

Unsupervised (feature) learning

Unsupervised NAS (ours)



Defining Unsupervised NAS (UnNAS)

	Supervised ( )	*Unsupervised* ( )
Learning weight	Supervised learning 	Unsupervised (feature) learning 
Learning architecture	(Supervised) NAS 	**Unsupervised NAS (ours)** 

Signals to exploit

Signals to exploit

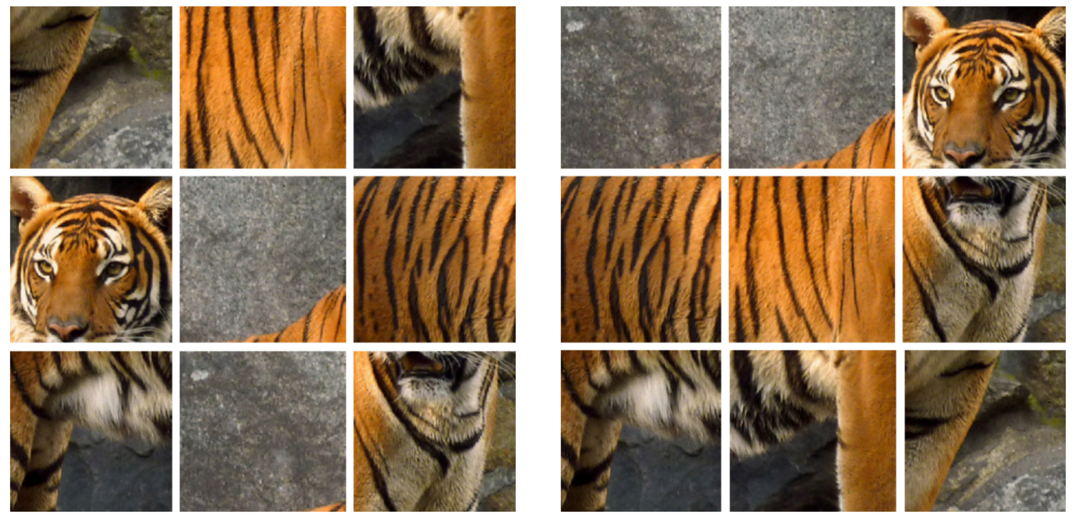
In this project, we rely on **self-supervised objectives**

- We will use “unsupervised” and “self-supervised” interchangeably
- These objectives were originally developed to transfer **learned weights**
- We study their ability to transfer **learned architectures** instead

Rotation

Colorization

Jigsaw



Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. ”Unsupervised representation learning by predicting image rotations.” In ICLR. 2018.
Zhang, Richard, Phillip Isola, and Alexei A. Efros. ”Colorful image colorization.” In ECCV. 2016.
Noroozi, Mehdi, and Paolo Favaro. ”Unsupervised learning of visual representations by solving jigsaw puzzles.” In ECCV. 2016.

Signals to exploit

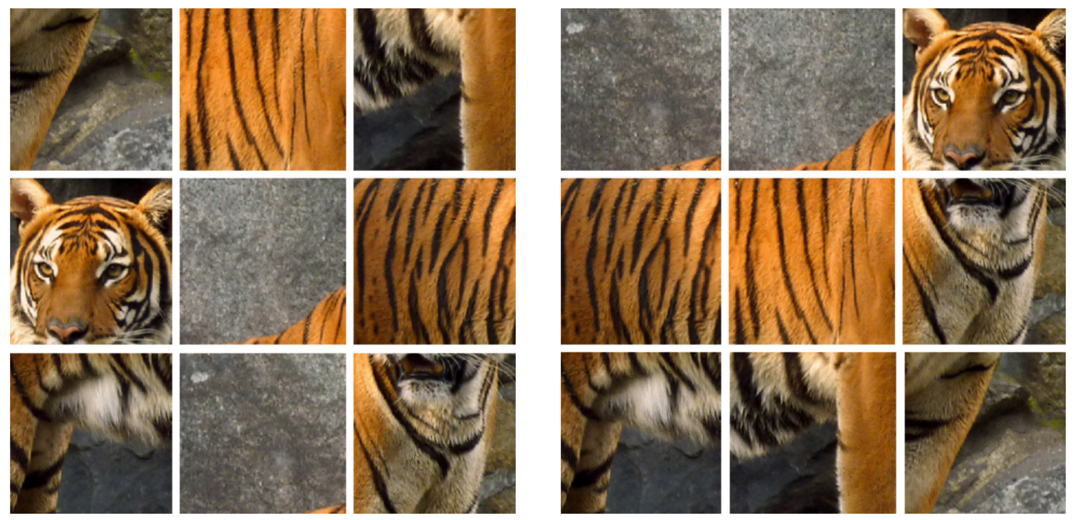
Using these 3 self-supervised objectives, we conduct **two sets of experiments** of complementary nature:

- Sample-based experiments
- Search-based experiments

Rotation

Colorization

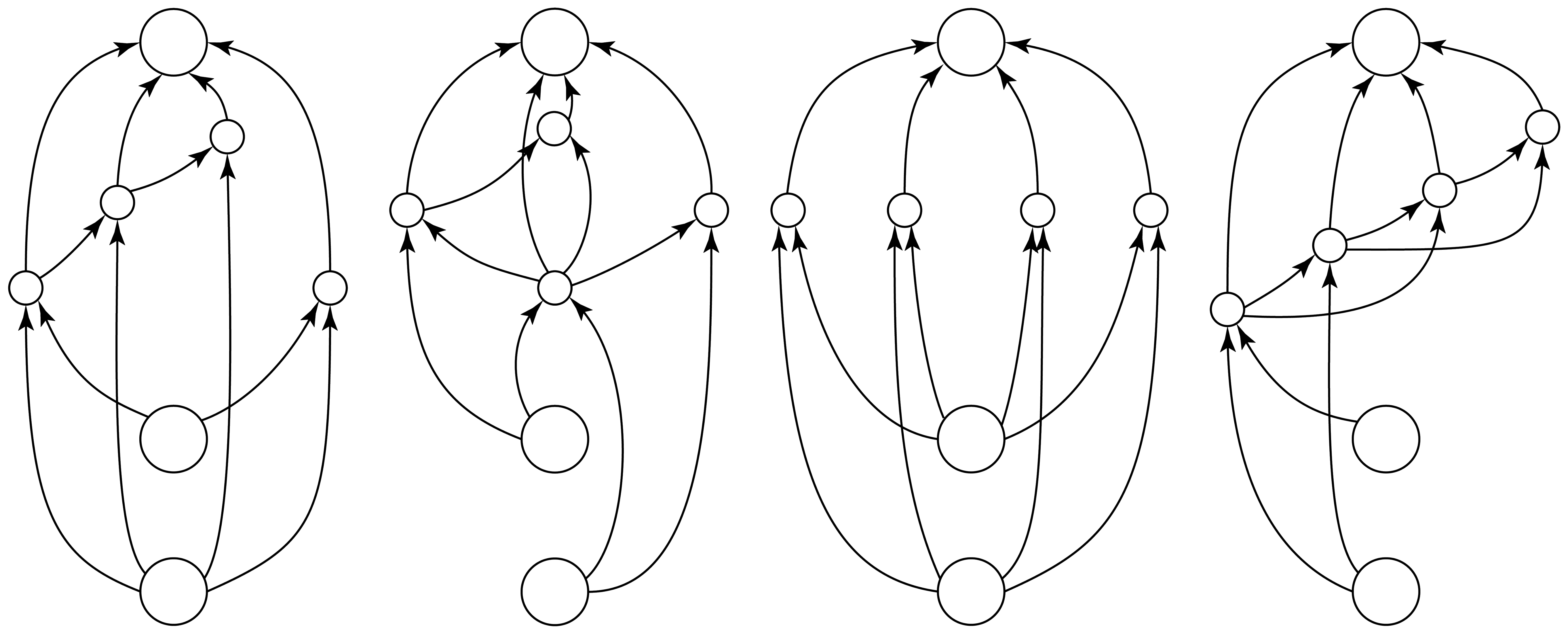
Jigsaw



Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations." In ICLR. 2018.
Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." In ECCV. 2016.
Noroozi, Mehdi, and Paolo Favaro. "Unsupervised learning of visual representations by solving jigsaw puzzles." In ECCV. 2016.

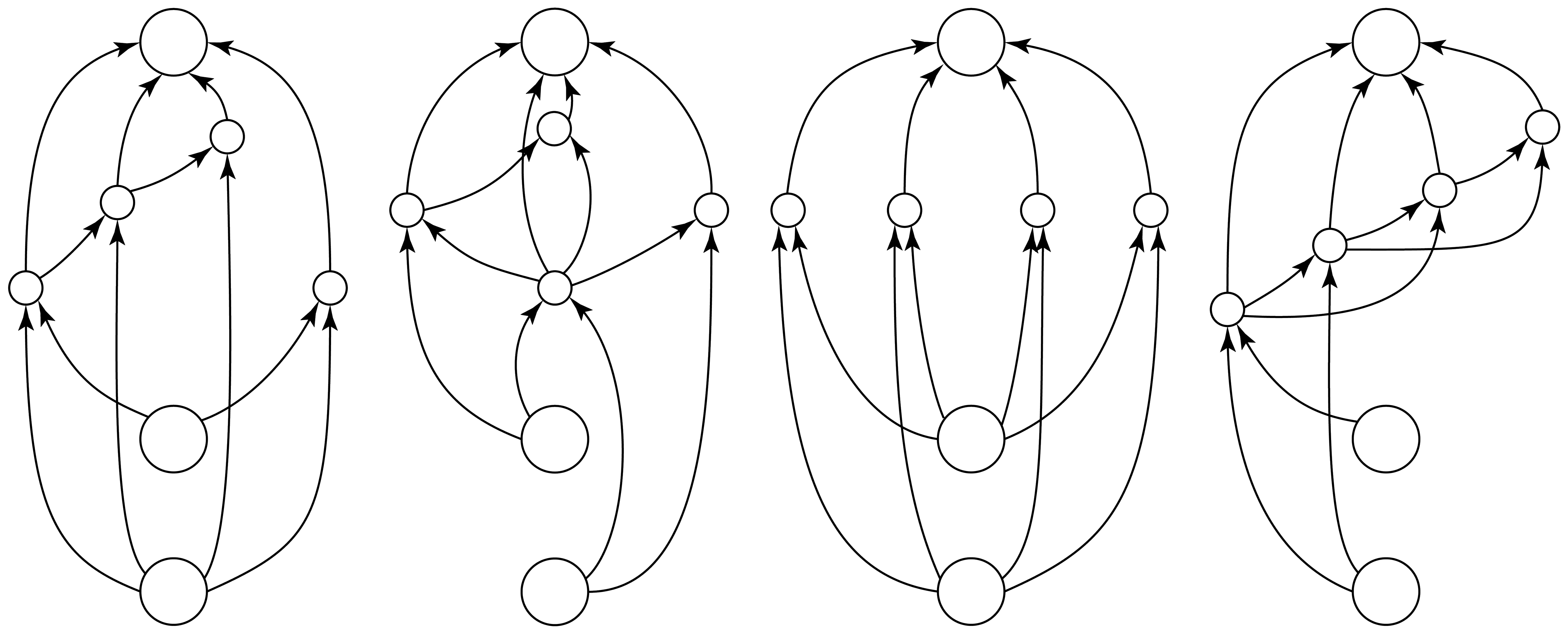
Sample-based experiments

Sample-based experiments



1. Sample 500 architectures

Sample-based experiments



1. Sample 500 architectures
2. Train them from scratch on different tasks; get accuracy

Rot:	93.0	93.1	93.2	93.3	
Col:	30.5	30.6	30.7	30.8	unsupervised
Jig:	76.0	76.1	76.2	76.3	
Cls:	92.1	92.2	92.3	92.4	supervised

Sample-based experiments

Rot: [93.0, 93.1, 93.2, 93.3]

ρ

Cls: [92.1, 92.2, 92.3, 92.4]

ρ ρ

Col: [30.5, 30.6, 30.7, 30.8]

Jig: [76.0, 76.1, 76.2, 76.3]

1. Sample 500 architectures
2. Train them from scratch on different tasks; get accuracy
3. Measure rank correlation between unsupervised and supervised

Do above on 2 datasets (CIFAR-10, ImageNet) and 2 search spaces (DARTS, NAS-Bench-101)

Sample-based experiments

Rot: [93.0, 93.1, 93.2, 93.3]

$\updownarrow \rho$

Cls: [92.1, 92.2, 92.3, 92.4]

ρ $\qquad$ ρ

Col: [30.5, 30.6, 30.7, 30.8] $\qquad$ Jig: [76.0, 76.1, 76.2, 76.3]

1. Sample 500 architectures
2. Train them from scratch on different tasks; get accuracy
3. Measure rank correlation between unsupervised and supervised

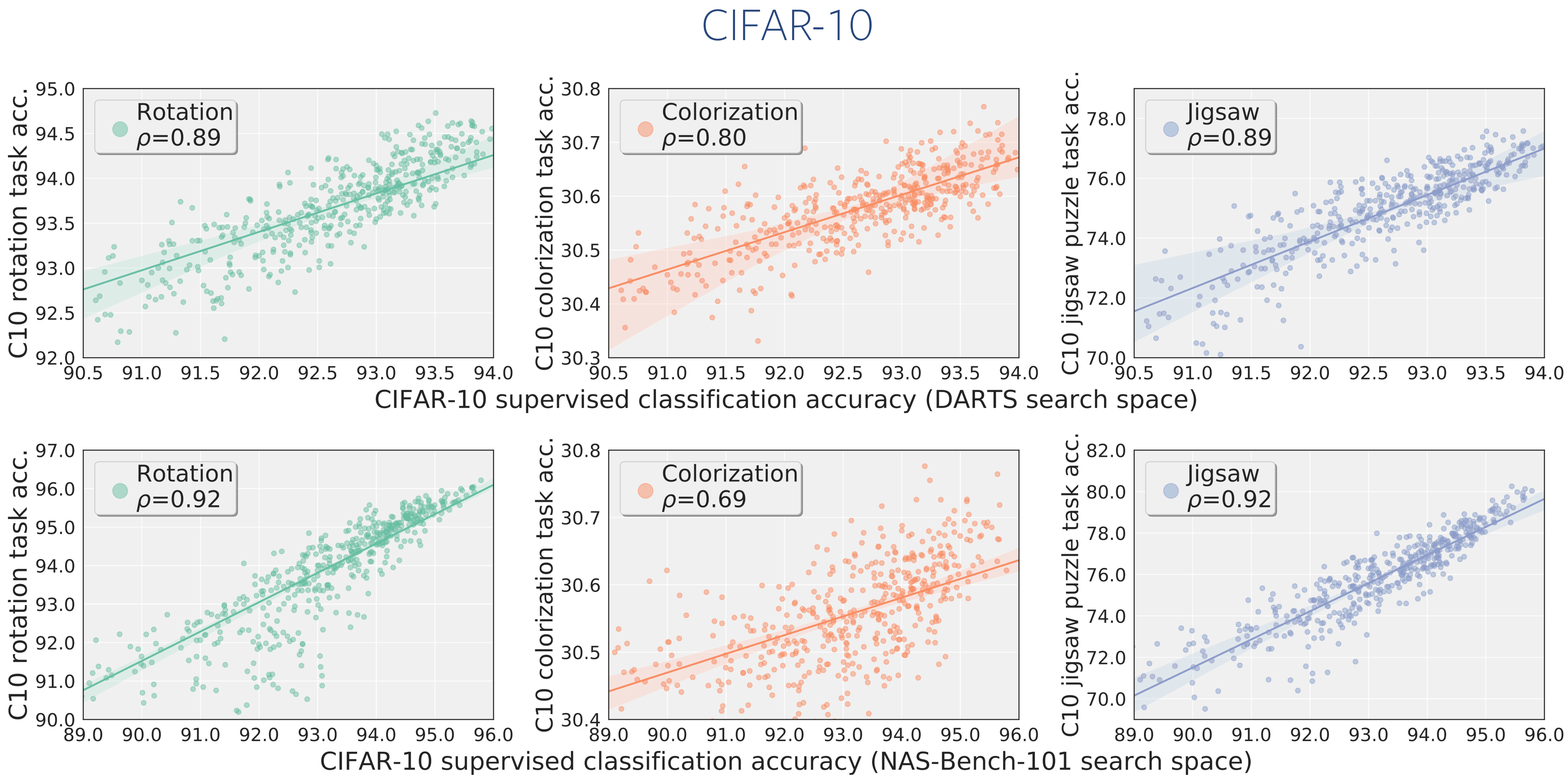
Do above on 2 datasets (CIFAR-10, ImageNet) and 2 search spaces (DARTS, NAS-Bench-101)

👍 : Each network trained and evaluated **individually**

👎 : Can only afford a **small, random subset** of entire search space

Sample-based experiments

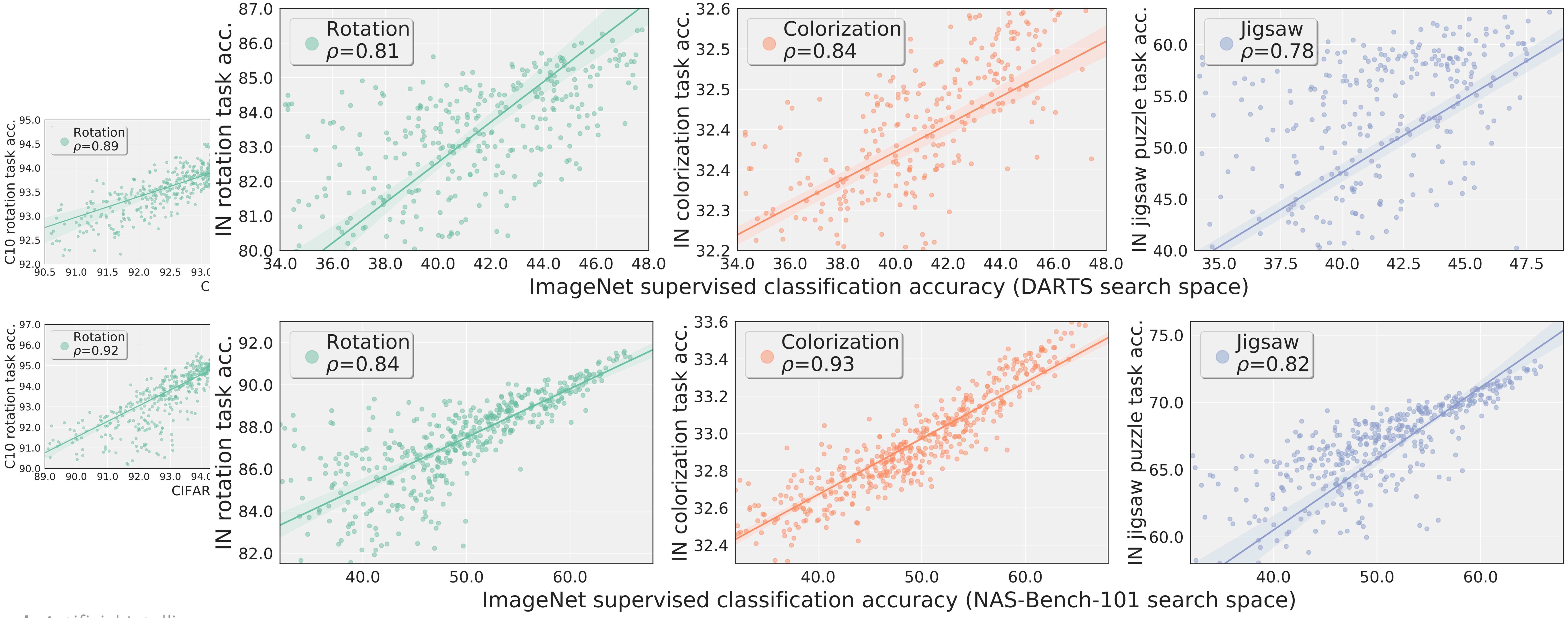
Architecture rankings produced with and without labels are **highly correlated** on the **same dataset**



Sample-based experiments

Architecture rankings produced with and without labels are **highly correlated** on the **same dataset**

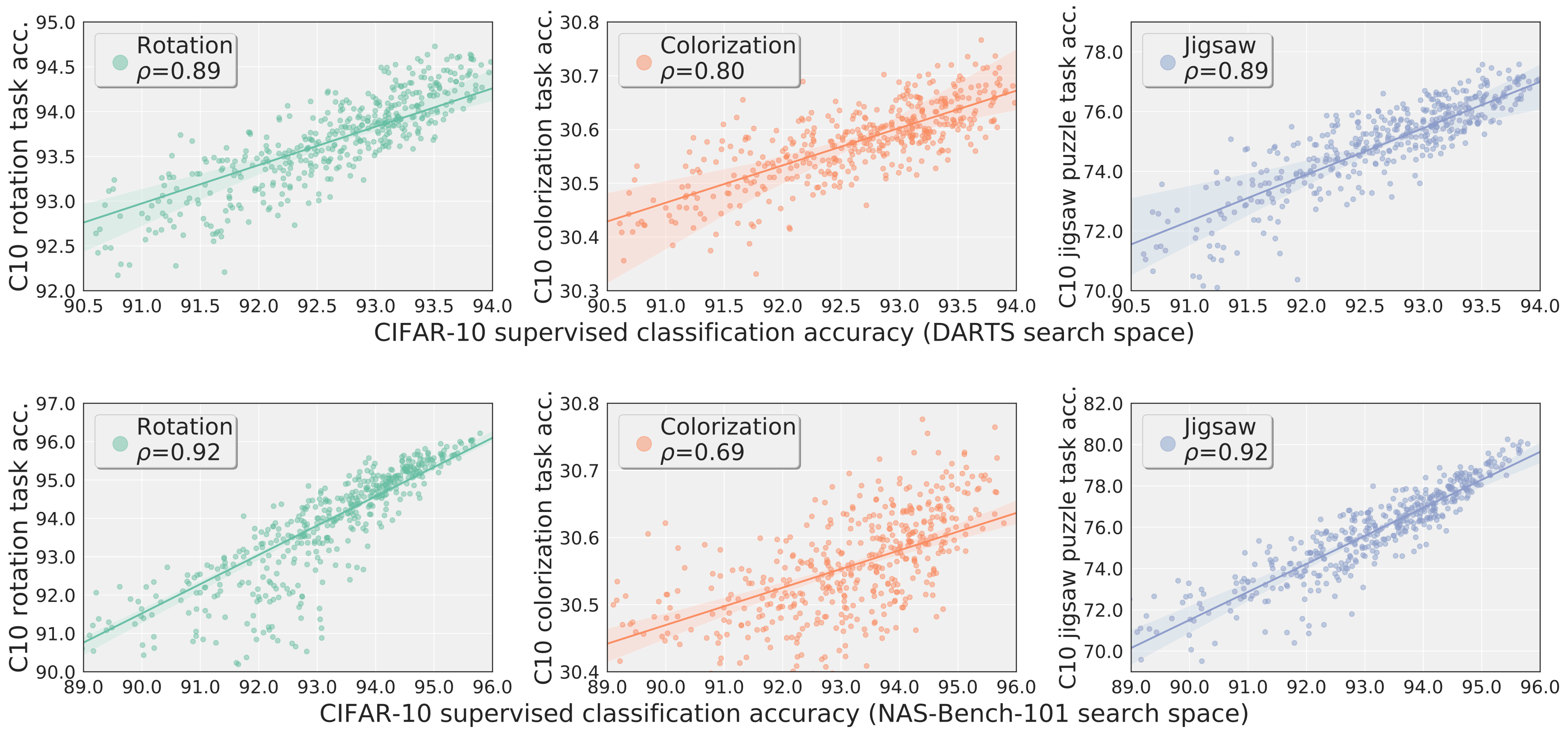
ImageNet-1K



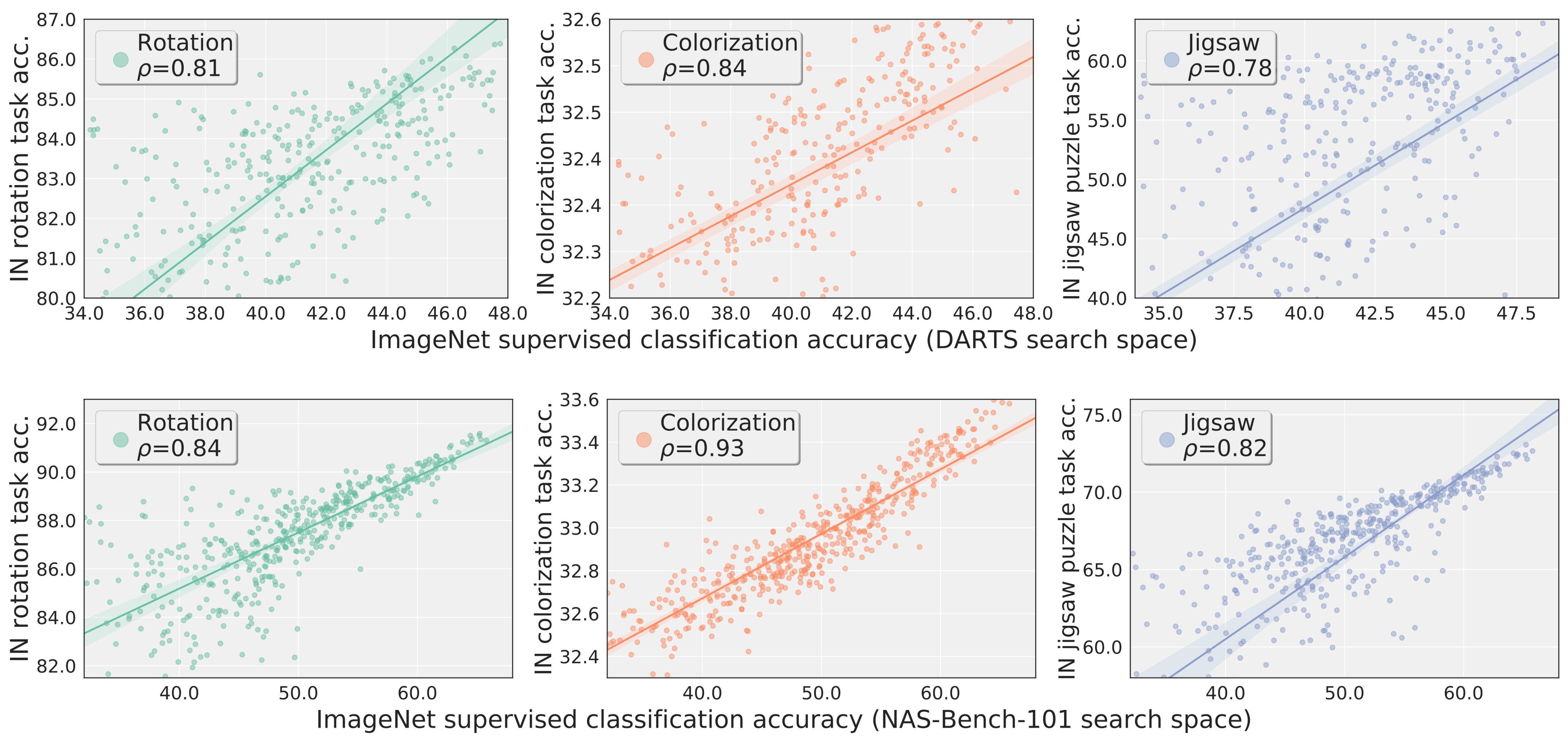
Sample-based experiments

Architecture rankings produced with and without labels are **highly correlated** on the **same dataset**

CIFAR-10

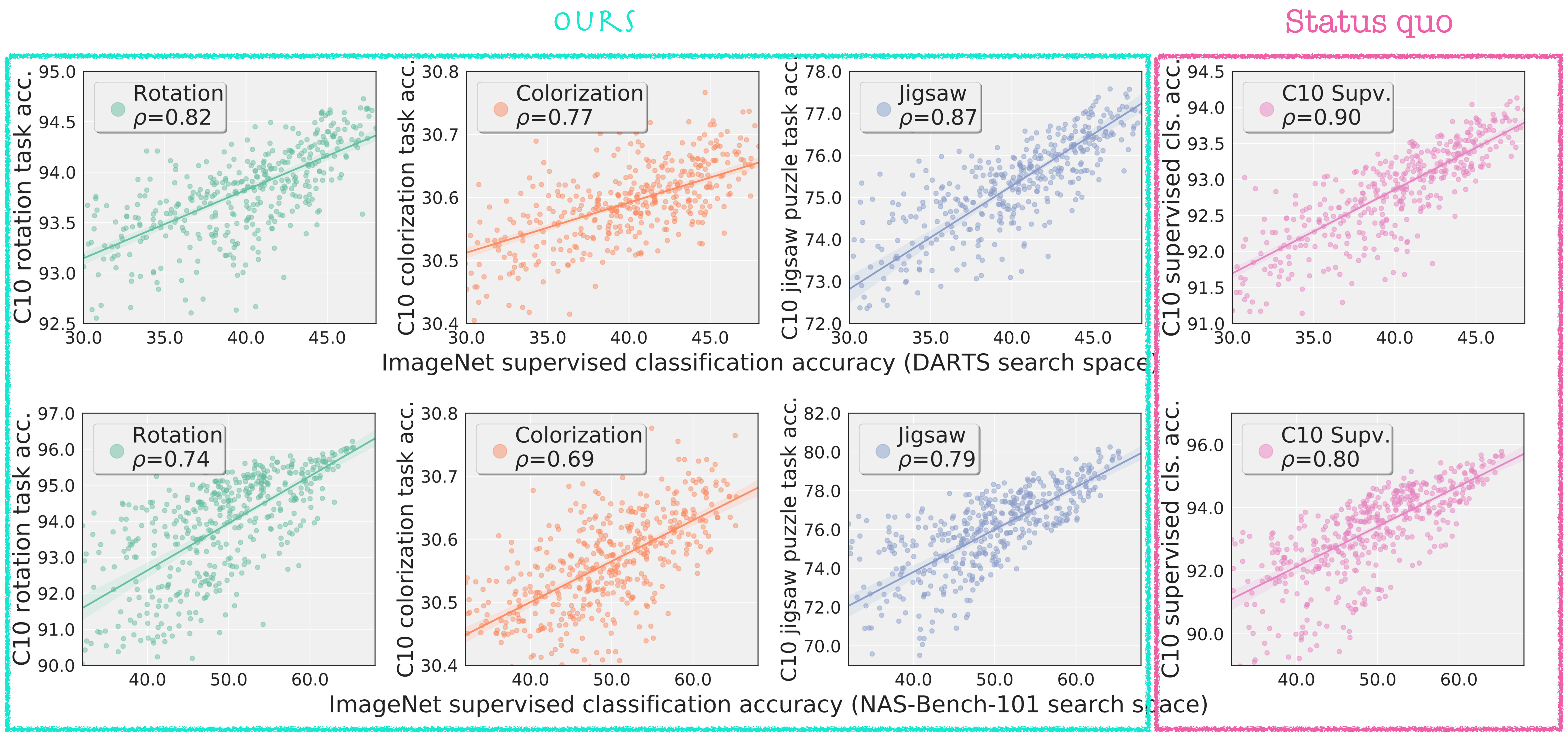


ImageNet-1K



Sample-based experiments

Architecture rankings produced with and without labels are **highly correlated** even **across datasets**

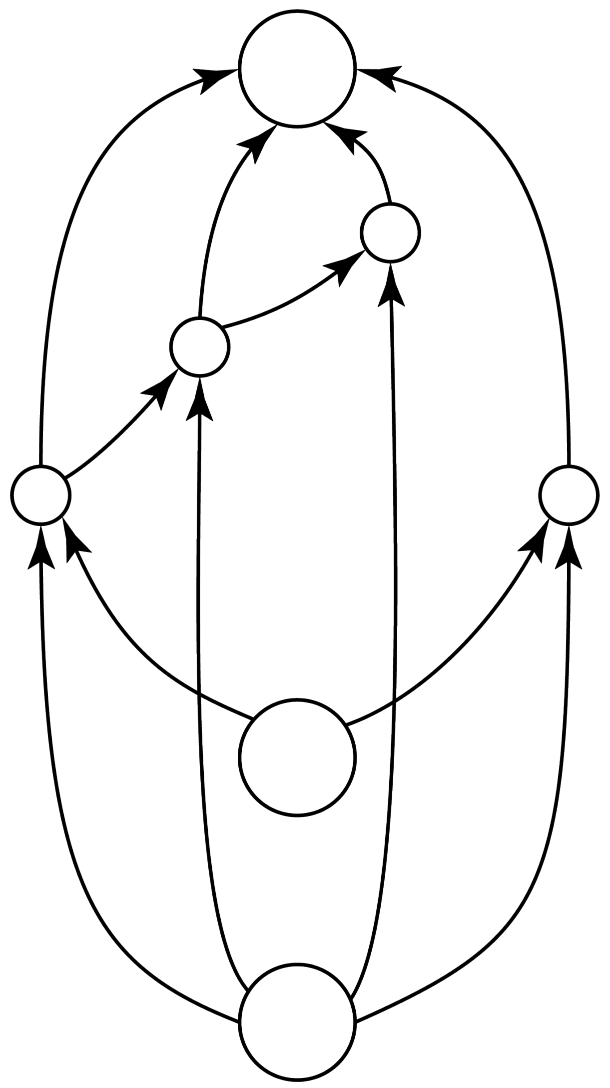


Search-based experiments

Search-based experiments

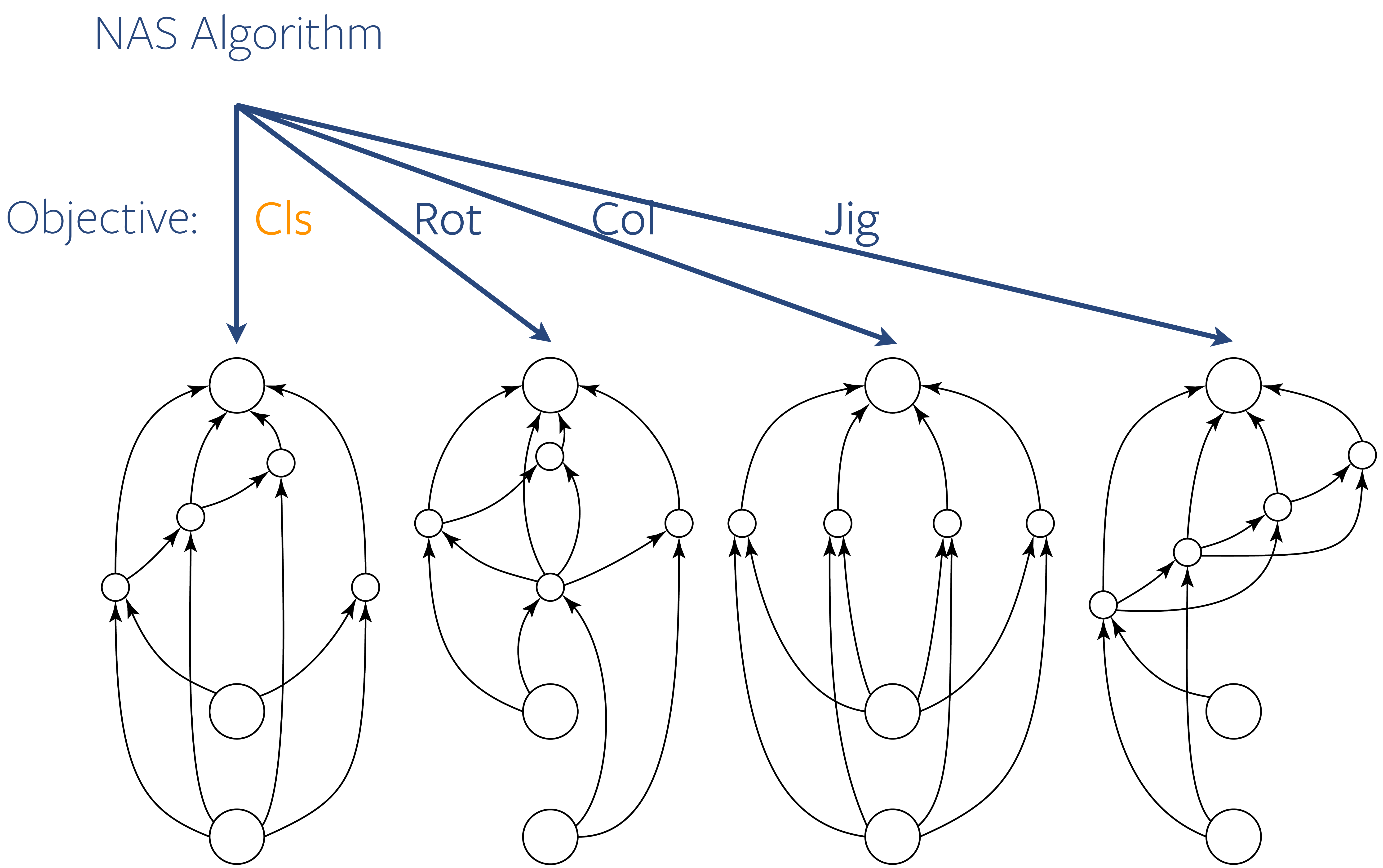
NAS Algorithm

Objective: Cls



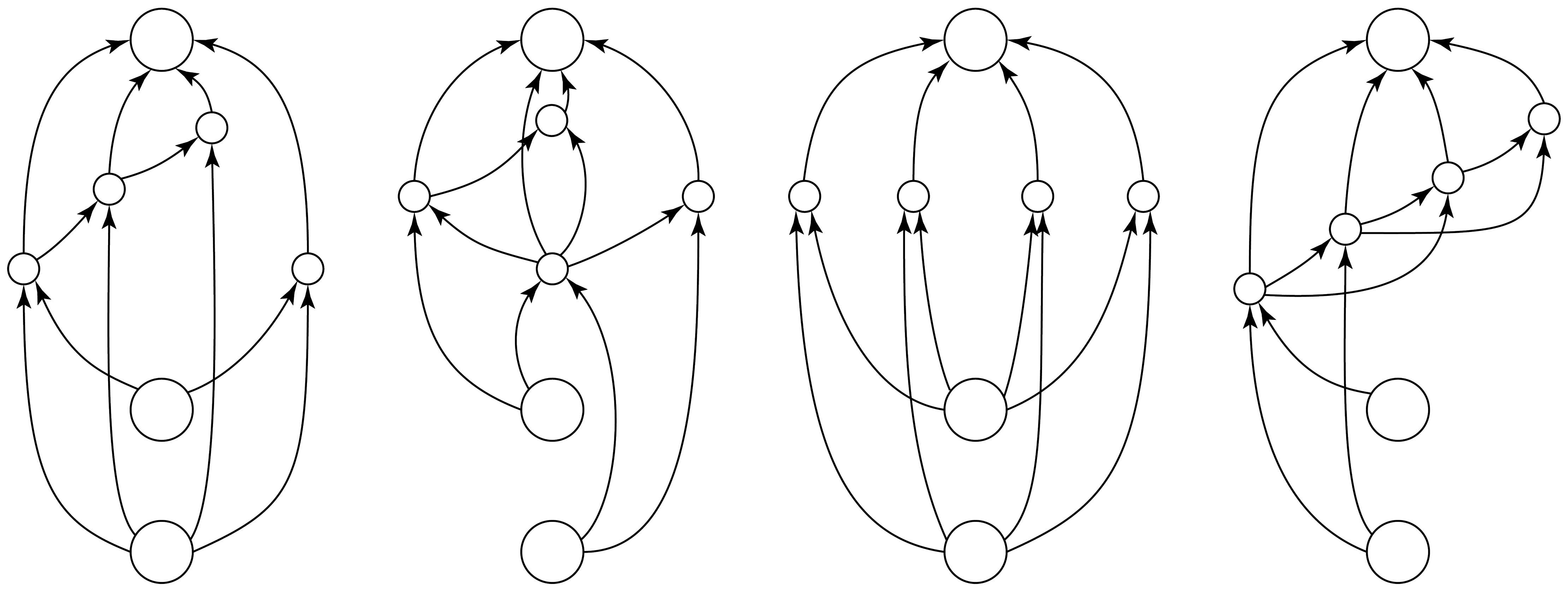
1. Take a NAS algorithm (DARTS)

Search-based experiments



1. Take a NAS algorithm (DARTS)
2. Run it with an unsupervised search objective

Search-based experiments

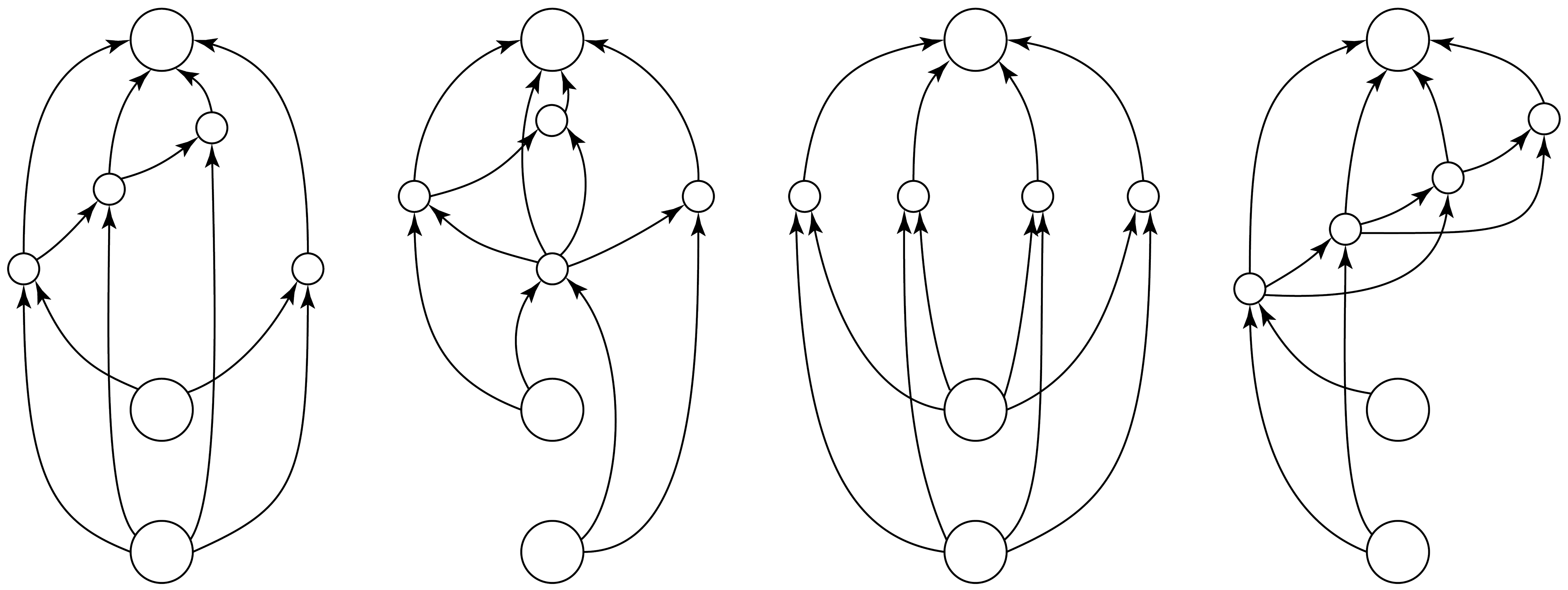


	Cls	Rot	Col	Jig
ImageNet-1K accuracy:	75.9	75.7	75.9	75.9
Cityscapes mIoU:	72.4	72.9	73.6	73.1

1. Take a NAS algorithm (DARTS)
2. Run it with an unsupervised search objective
3. Train and evaluate the searched architecture; compare supervised vs unsupervised

Do above on 3 search datasets (ImageNet-1K, ImageNet-22K, Cityscapes) and 2 target datasets + tasks (ImageNet-1K classification, Cityscapes semantic segmentation)

Search-based experiments



	Cls	Rot	Col	Jig
ImageNet-1K accuracy:	75.9	75.7	75.9	75.9
Cityscapes mIoU:	72.4	72.9	73.6	73.1

1. Take a NAS algorithm (DARTS)
2. Run it with an unsupervised search objective
3. Train and evaluate the searched architecture; compare supervised vs unsupervised

Do above on 3 search datasets (ImageNet-1K, ImageNet-22K, Cityscapes) and 2 target datasets + tasks (ImageNet-1K classification, Cityscapes semantic segmentation)

👍: Explore the **entire** search space

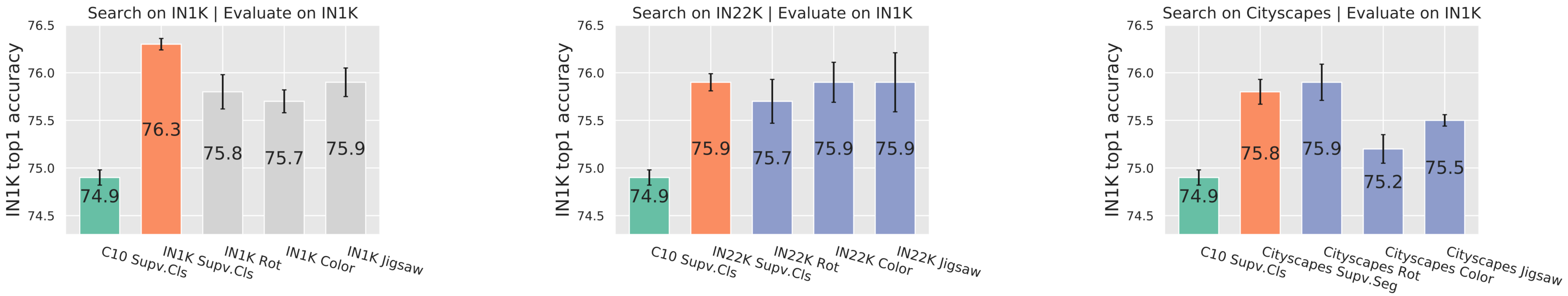
👎: Training dynamics **mismatch** between search phase and eval phase

Search-based experiments

UnNAS:

- is better than the commonly used CIFAR-10 supervised proxy
- is comparable to (supervised) NAS across search tasks and datasets
- can even outperform the state-of-the-art (75.8) which uses a more sophisticated algorithm

ImageNet classification



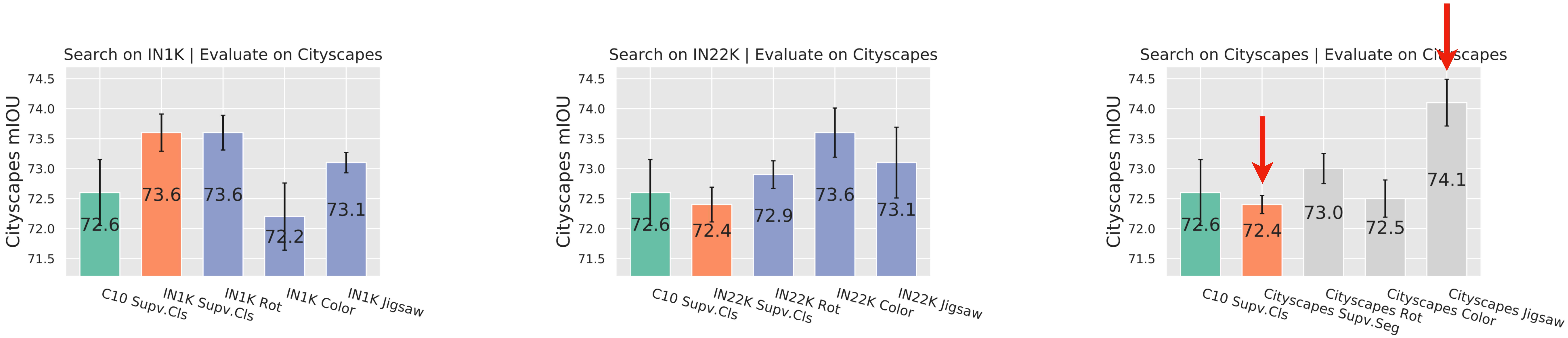
Xu, Yuhui, et al. "Pc-darts: Partial channel connections for memory-efficient differentiable architecture search." In ICLR. 2020.

Search-based experiments

UnNAS:

- is better than the commonly used CIFAR-10 supervised proxy
- is comparable to (supervised) NAS across search tasks and datasets
- can even be clearly better than supervised NAS

Cityscapes semantic segmentation

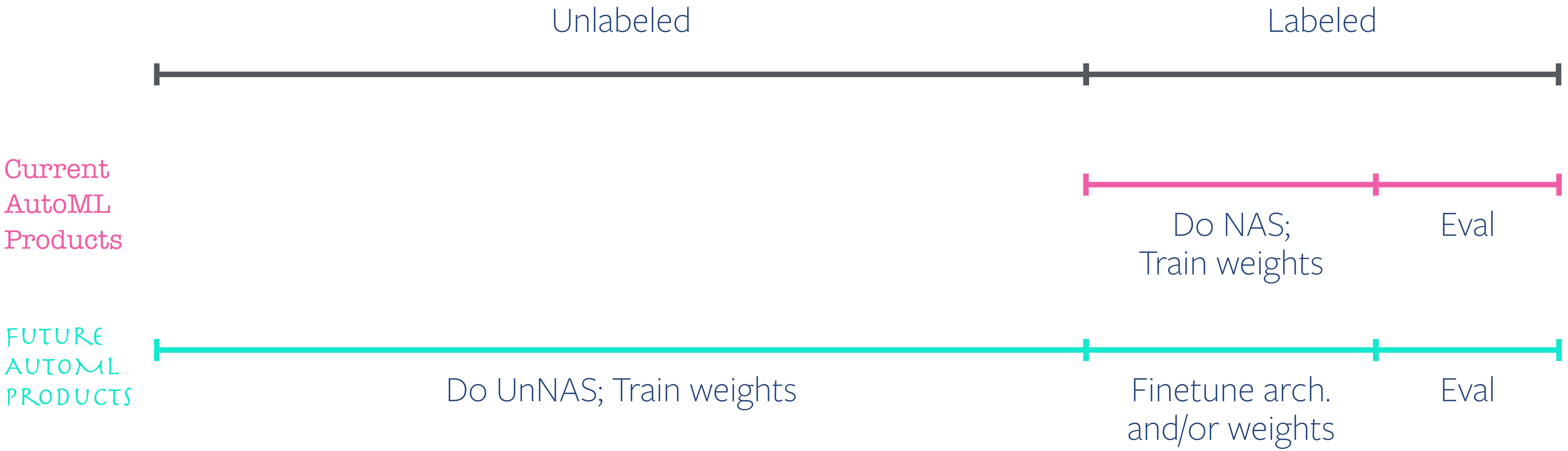


To perform NAS *successfully*,
labels are ***not*** *necessary*

Implications

Reduce the labeling requirement in existing AutoML products

Enable the possibility of searching for architectures on datasets too large to label



Thank you!