

#82



Adaptive Neural Trees

Ryutaro Tanno, Kai Arulkumaran,
Daniel C. Alexander, Antonio Criminisi, Aditya Nori



Imperial College
London

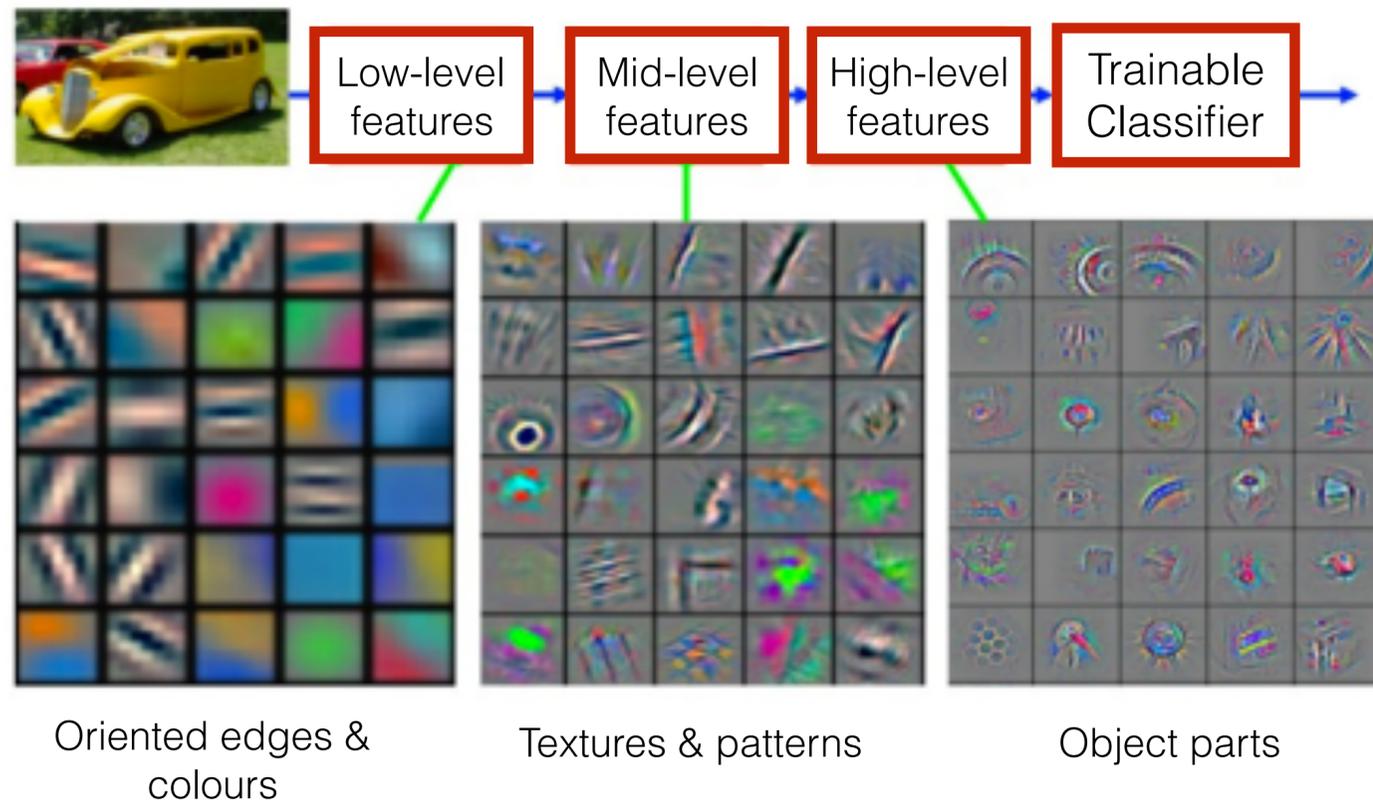


Two Paradigms of Machine Learning

Deep Neural Networks

『hierarchical *representation* of data』

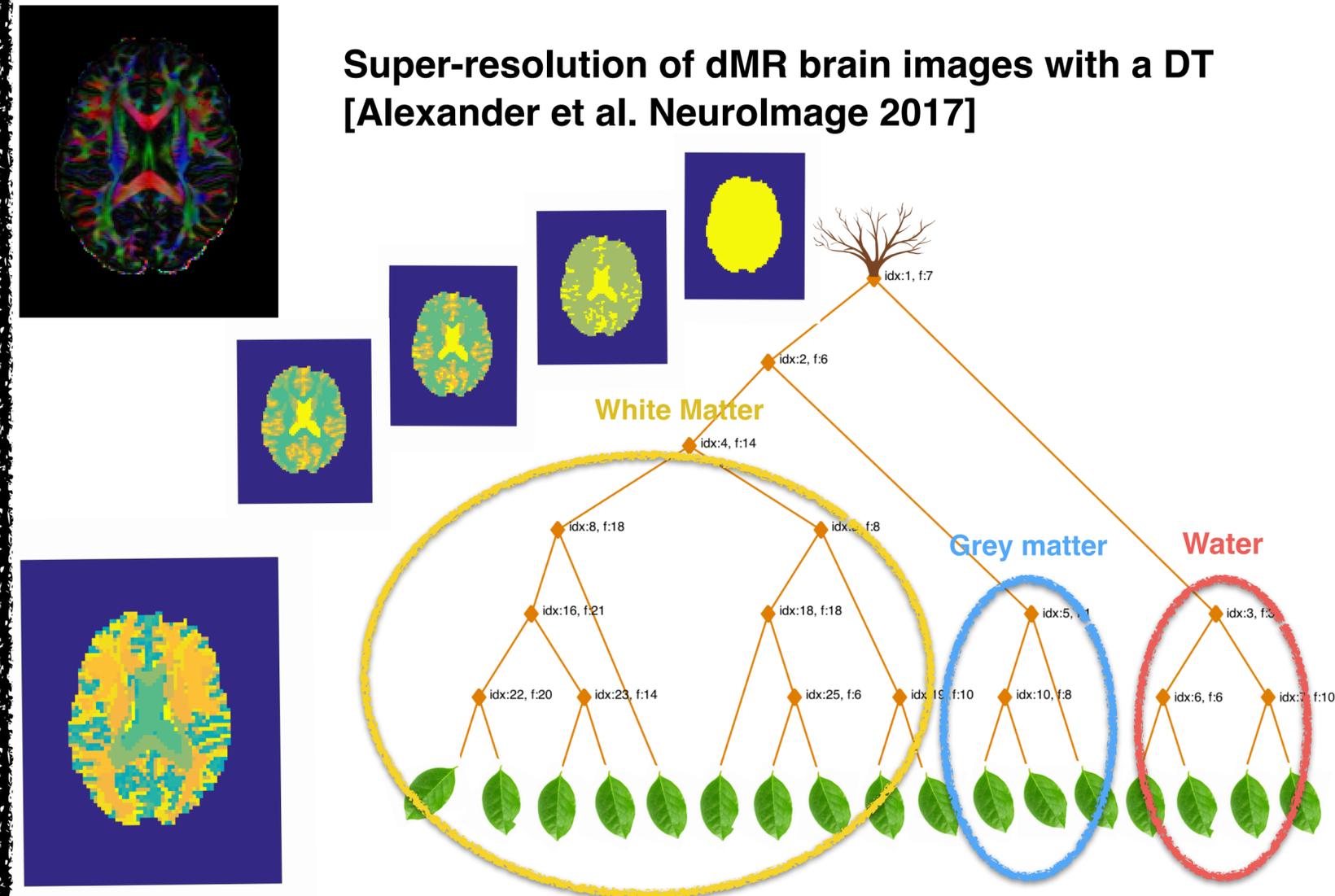
ImageNet classifiers with CNNs
[Zeiler and Fergus, ECCV 2014]



Decision Trees

『hierarchical *clustering* of data』

Super-resolution of dMR brain images with a DT
[Alexander et al. NeuroImage 2017]



Two Paradigms of Machine Learning

Deep Neural Networks

『hierarchical *representation* of data』

Decision Trees

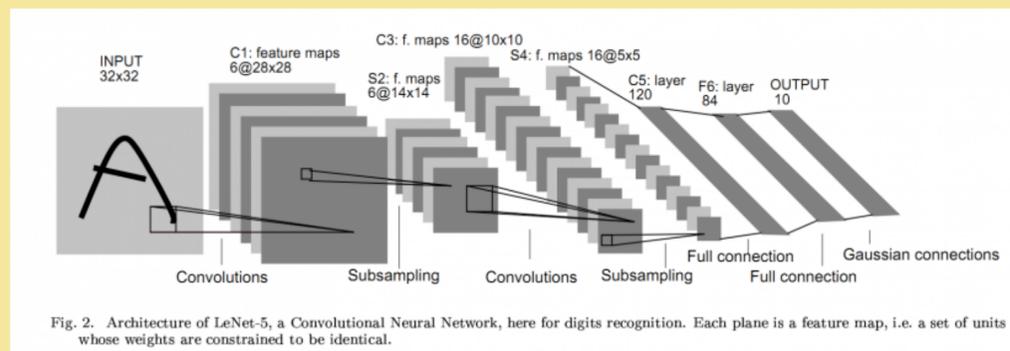
『hierarchical *clustering* of data』

Two Paradigms of Machine Learning

Deep Neural Networks

『hierarchical *representation* of data』

- + learn features of data
- + scalable learning with stochastic optimisation
- architectures are hand-designed
- heavy-weight inference, engaging every parameter of the model for each input



Decision Trees

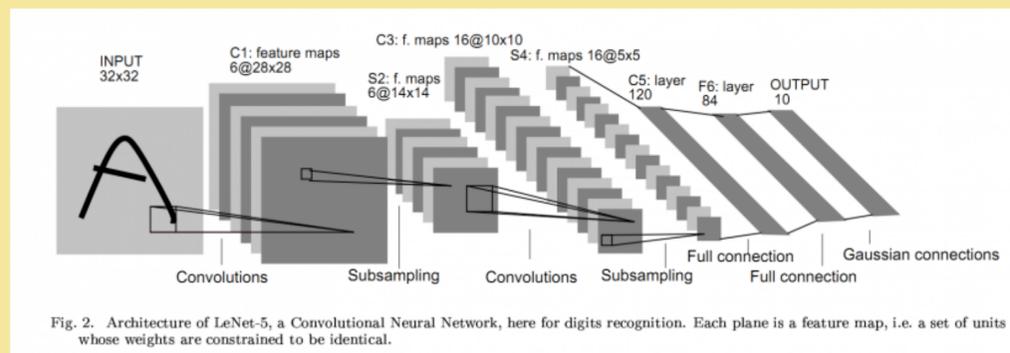
『hierarchical *clustering* of data』

Two Paradigms of Machine Learning

Deep Neural Networks

『hierarchical *representation* of data』

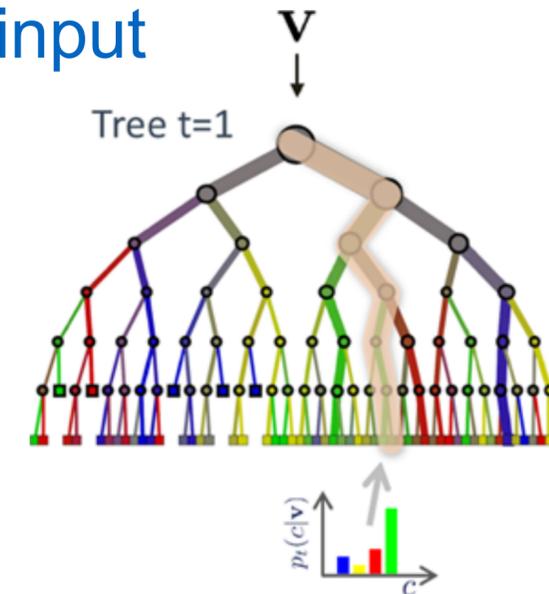
- + learn features of data
- + scalable learning with stochastic optimisation
- architectures are hand-designed
- heavy-weight inference, engaging every parameter of the model for each input



Decision Trees

『hierarchical *clustering* of data』

- operate on hand-designed features
- limited expressivity with simple splitting functions
- + architectures are learned from data
- + lightweight inference, activating only a fraction of the model per input



Joining the Paradigms

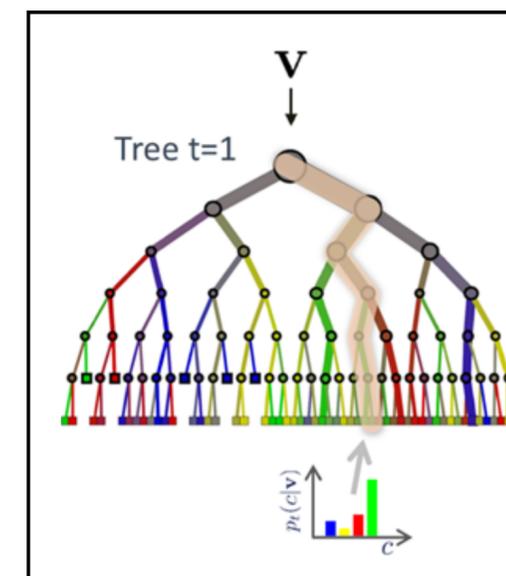
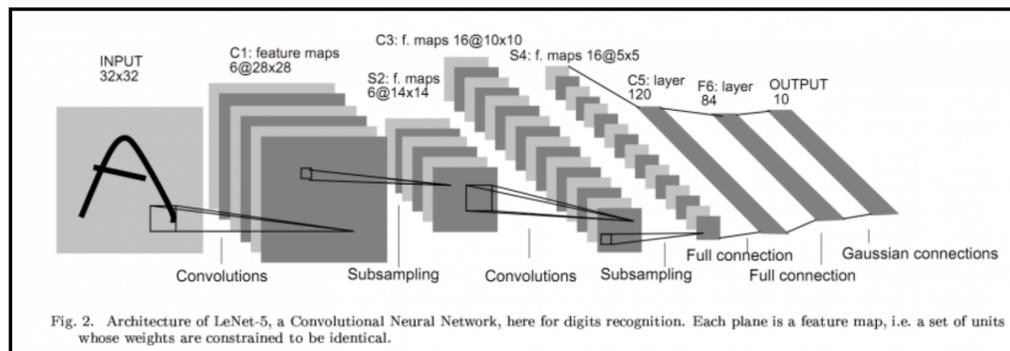
Adaptive Neural Trees

『hierarchical *representation* of data』

- + learn features of data
- + scalable learning with stochastic optimisation

『hierarchical *clustering* of data』

- + architectures are learned from data
- + lightweight inference, activating only a fraction of the model per input



ANTs unify the two paradigms and generalise previous work

Joining the Paradigms

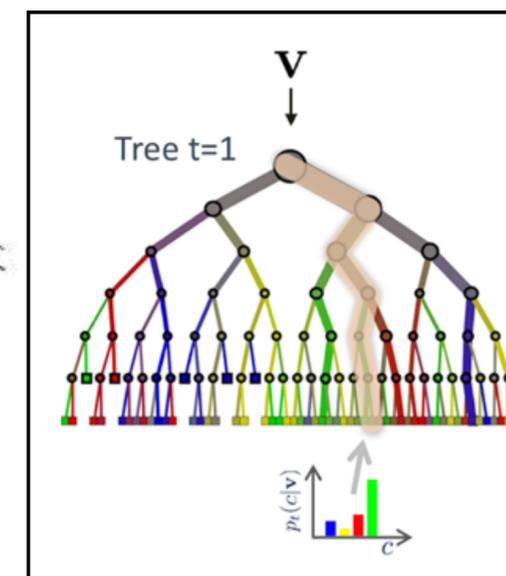
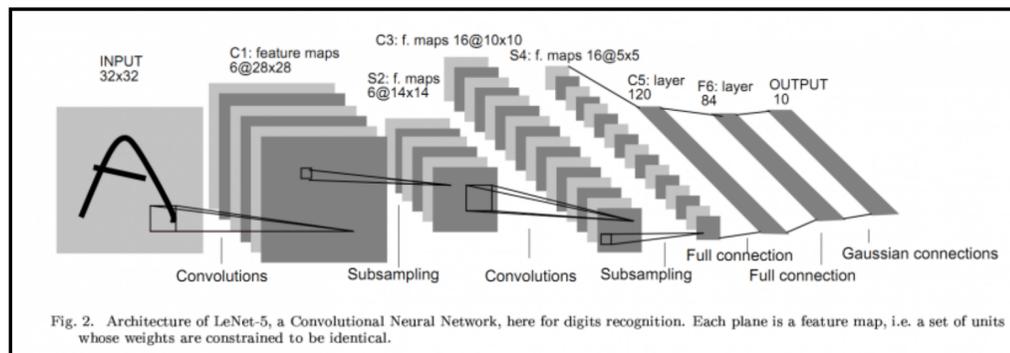
Adaptive Neural Trees

『hierarchical *representation* of data』

- + learn features of data
- + scalable learning with stochastic optimisation

『hierarchical *clustering* of data』

- + architectures are learned from data
- + lightweight inference, activating only a fraction of the model per input



ANTs unify the two paradigms and generalise previous work

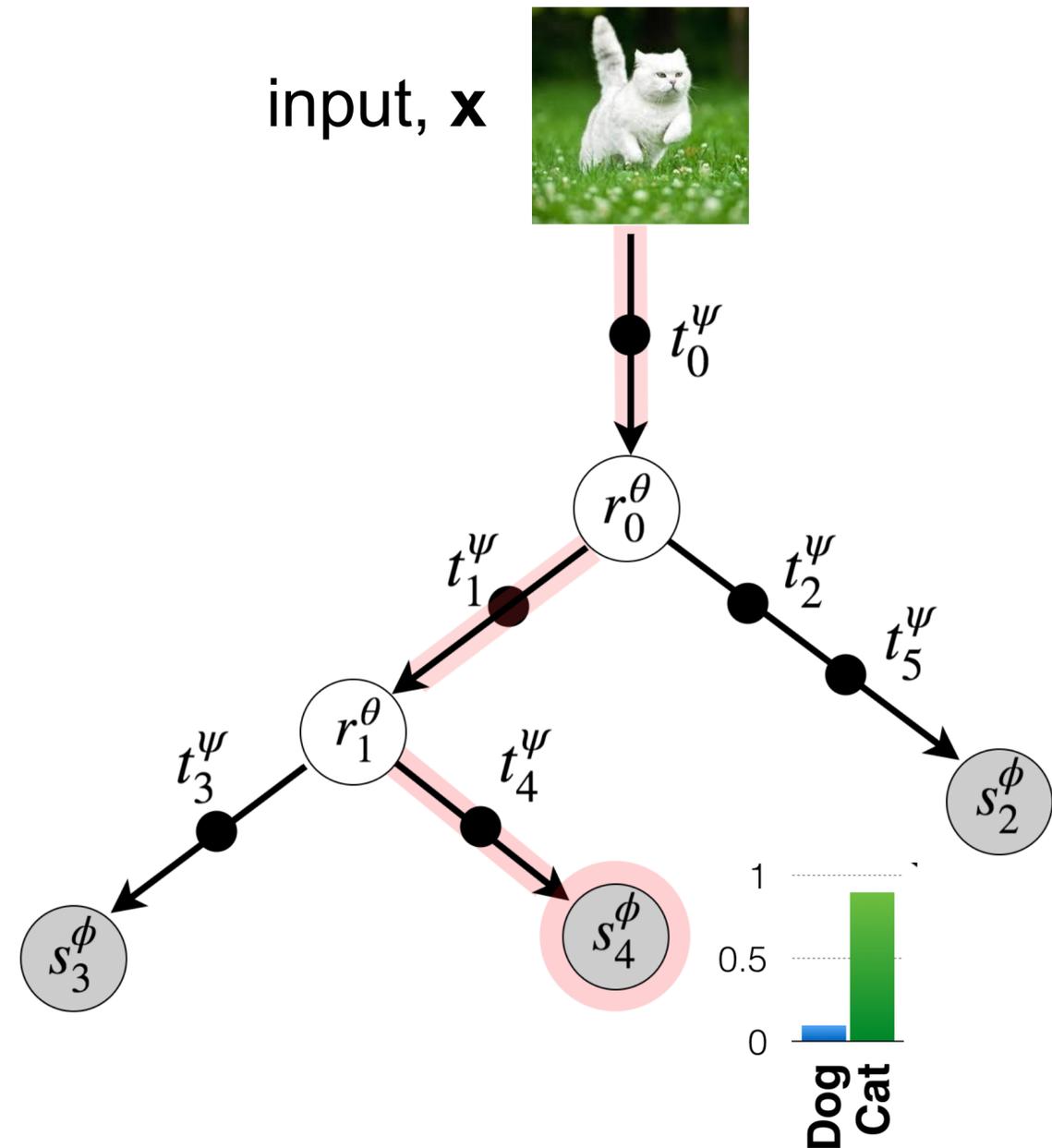
What are ANTs?

- ANTs consist of two key designs:

What are ANTs?

- ANTs consist of two key designs:

(1). DTs which uses NNs in every path and routing decisions.

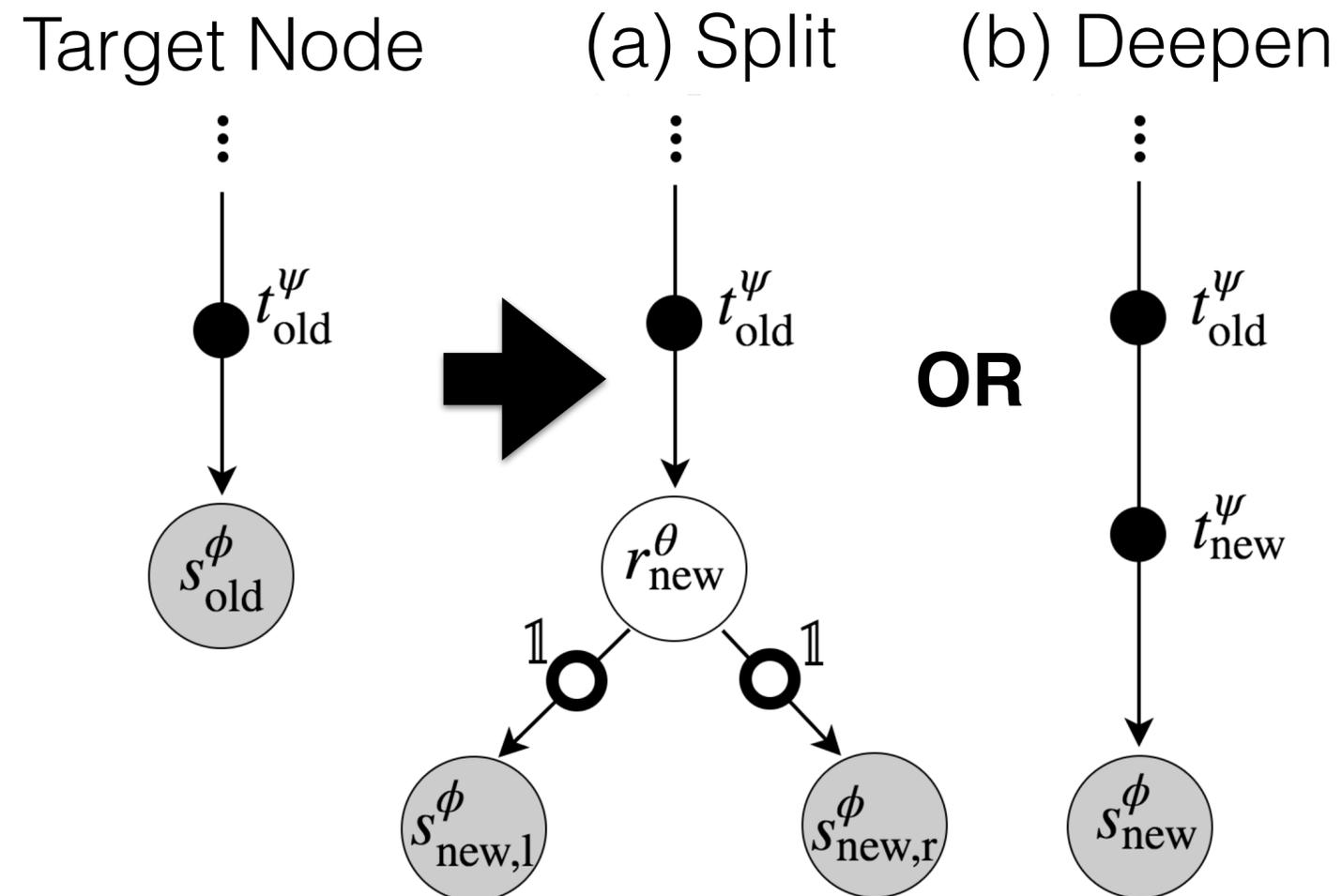


What are ANTs?

• ANTs consist of two key designs:

(1). DTs which uses NNs in every path and routing decisions.

(2). DT-like architecture growth using SGD

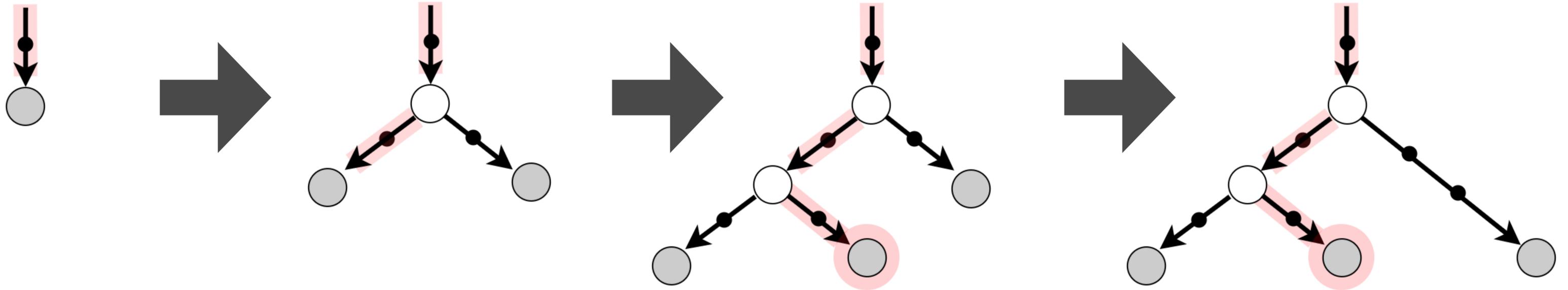


What are ANTs?

- ANTs consist of two key designs:

- (1). DTs which uses NNs in every path and routing decisions.

- (2). DT-like architecture growth using SGD

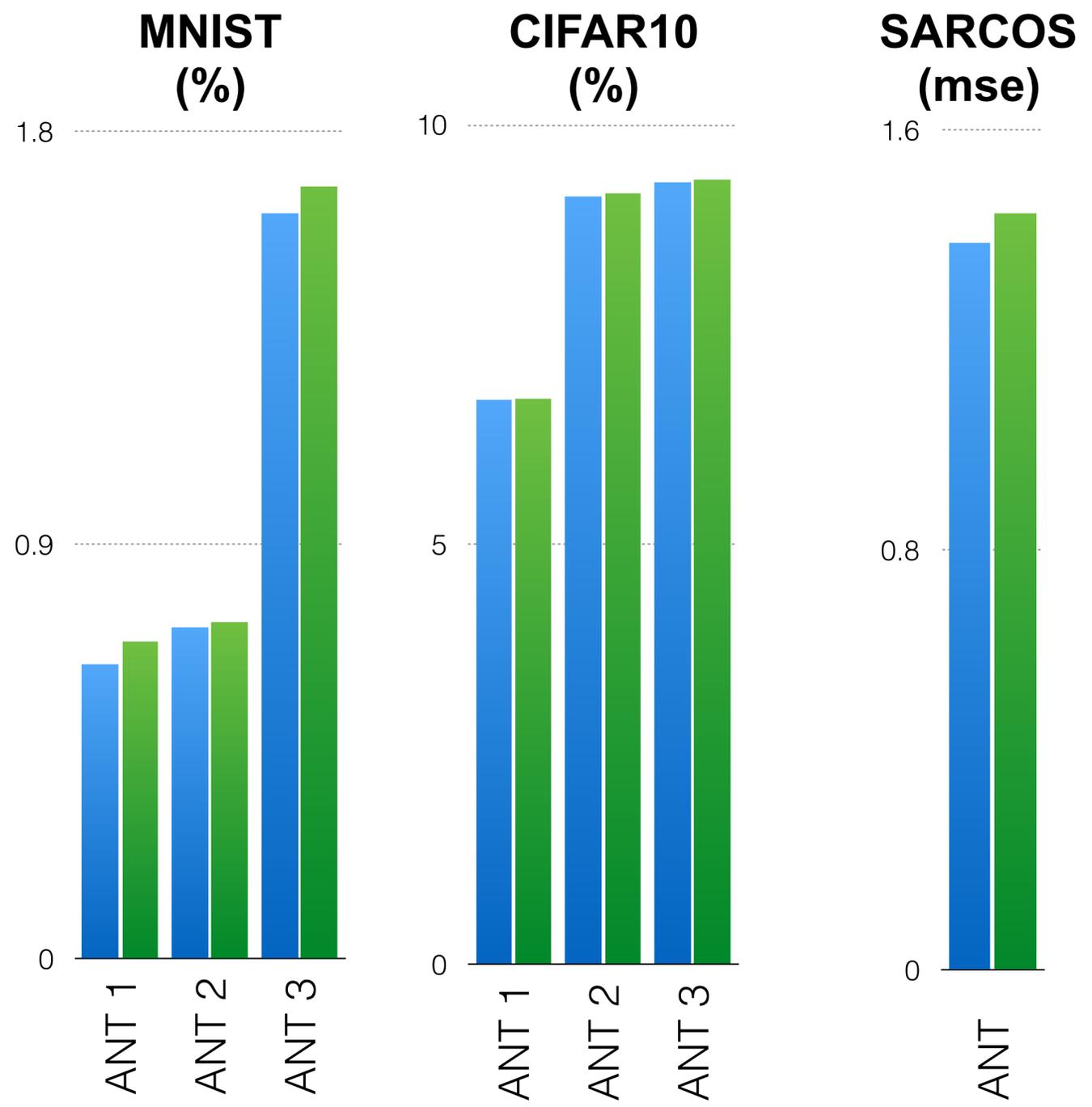


Conditional Computation

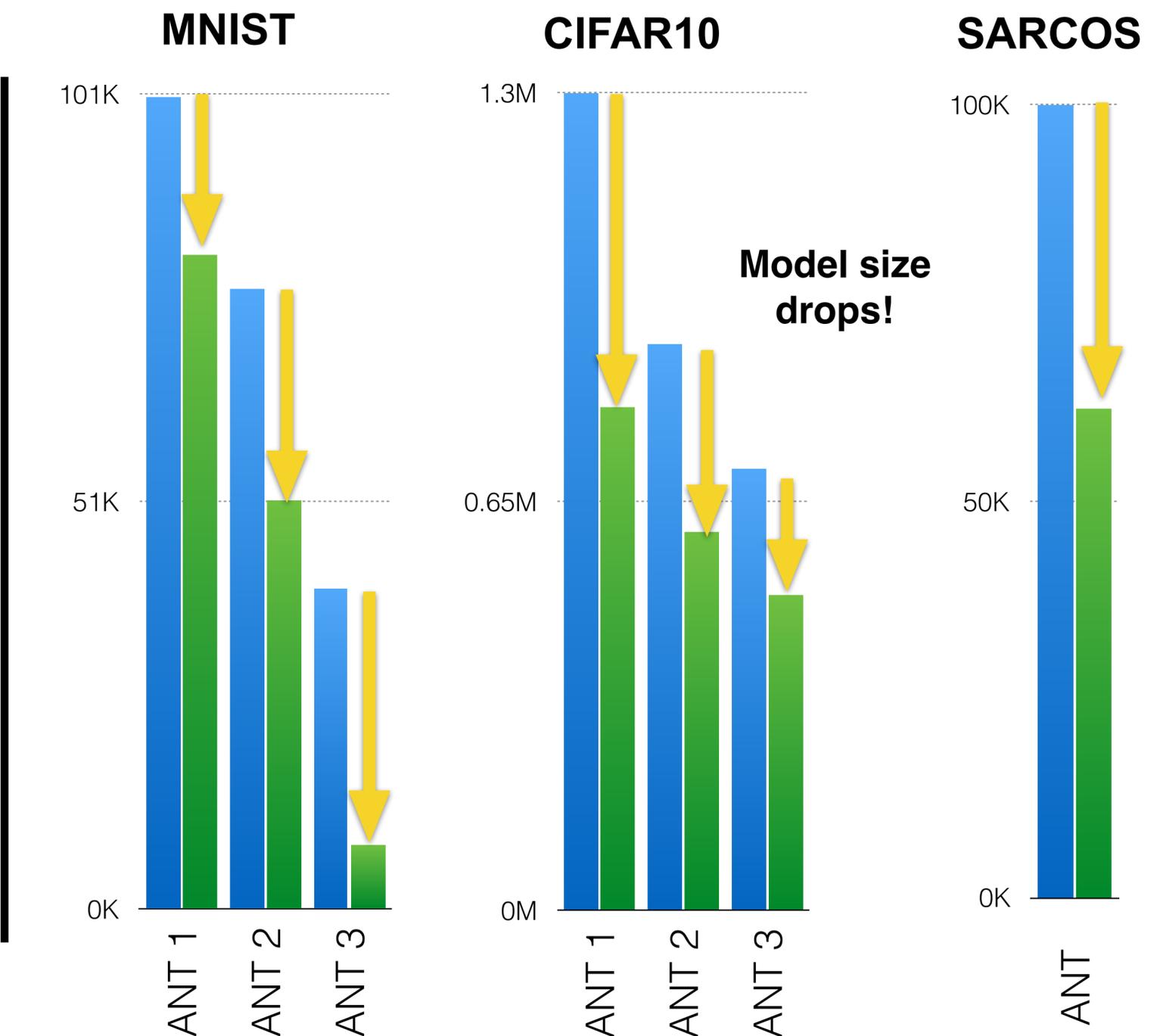
- Multi-path inference
- Single-path inference

• Single-path inference enables efficient inference without compromising accuracy.

Errors

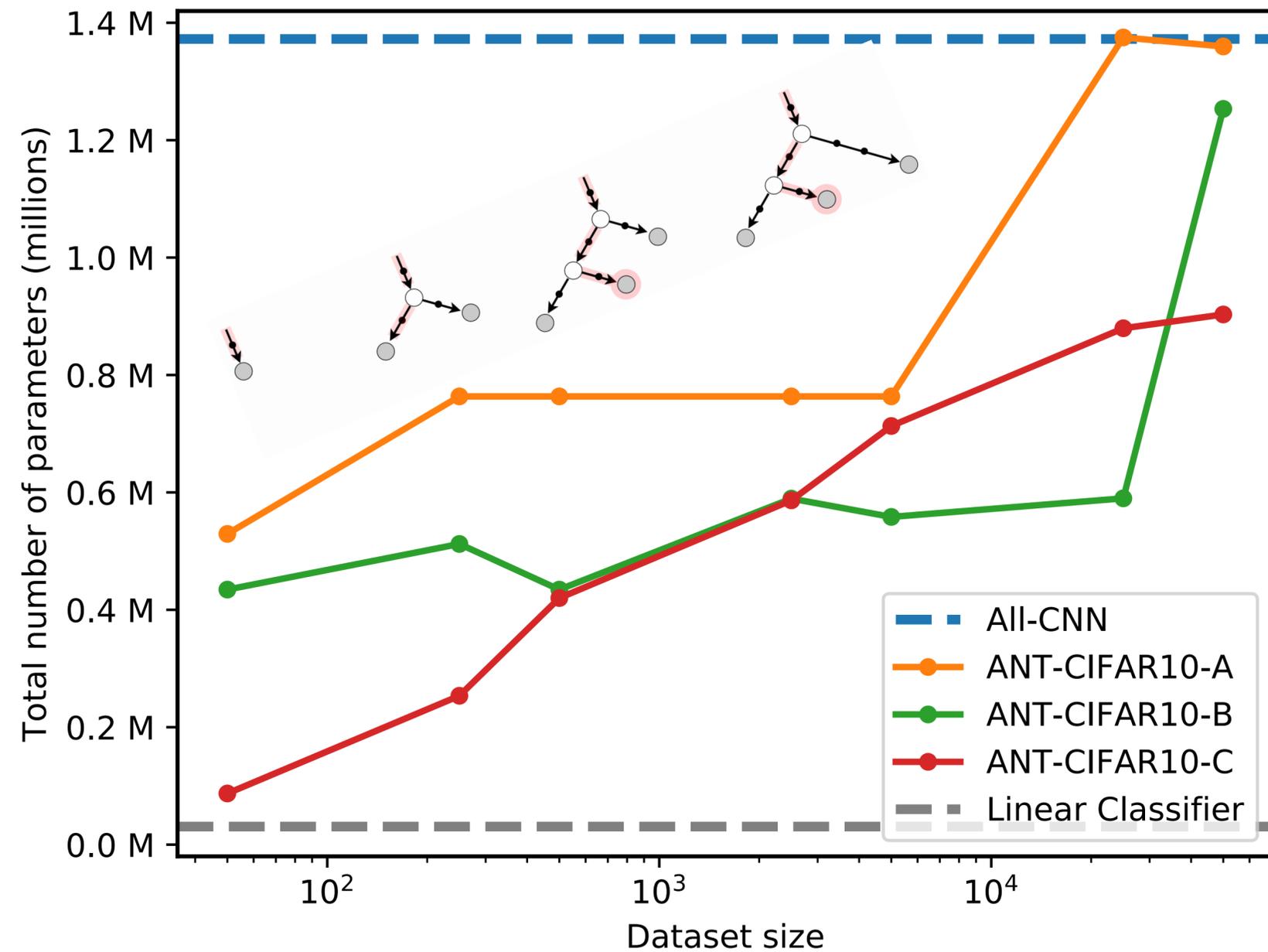
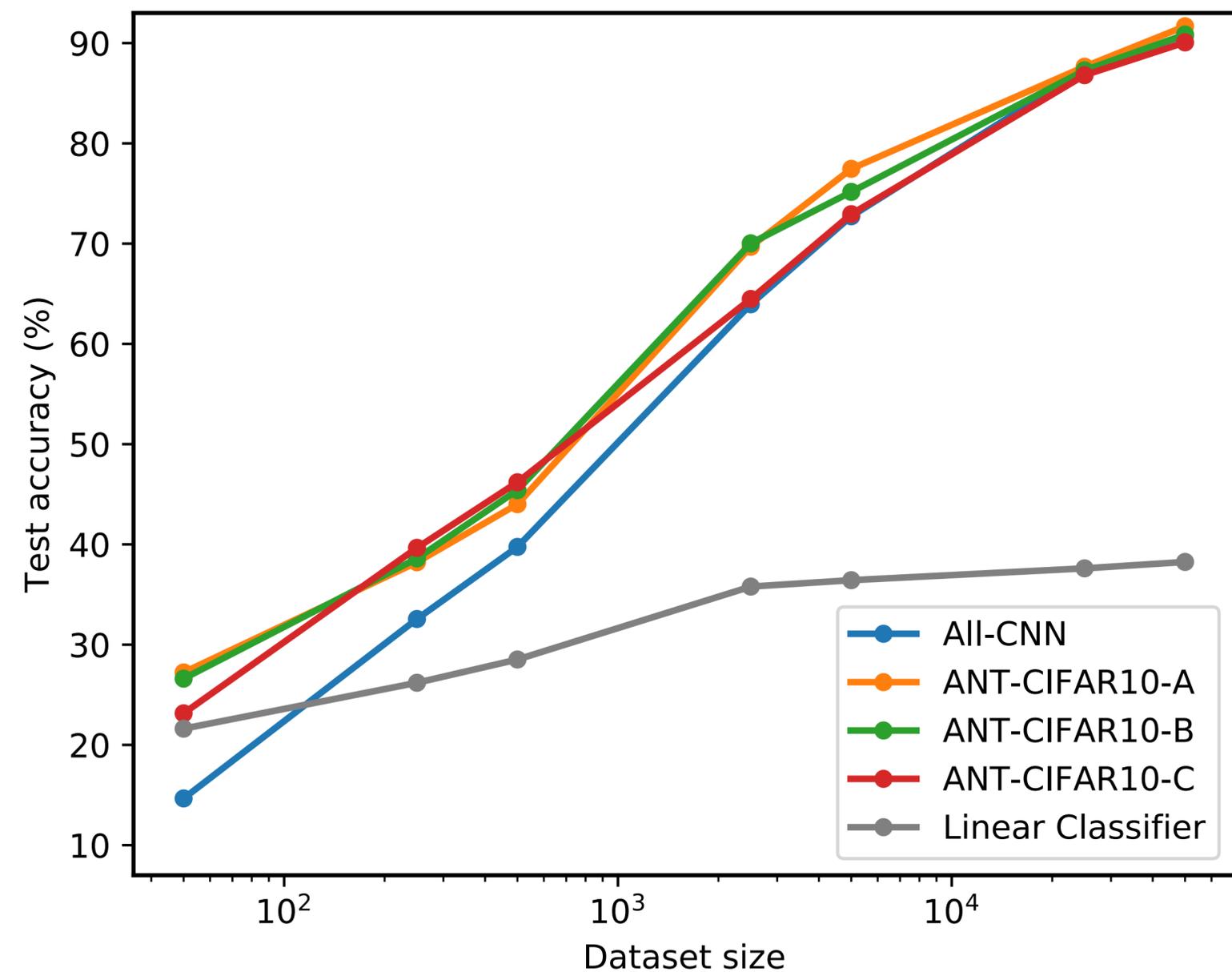


Number of Parameters



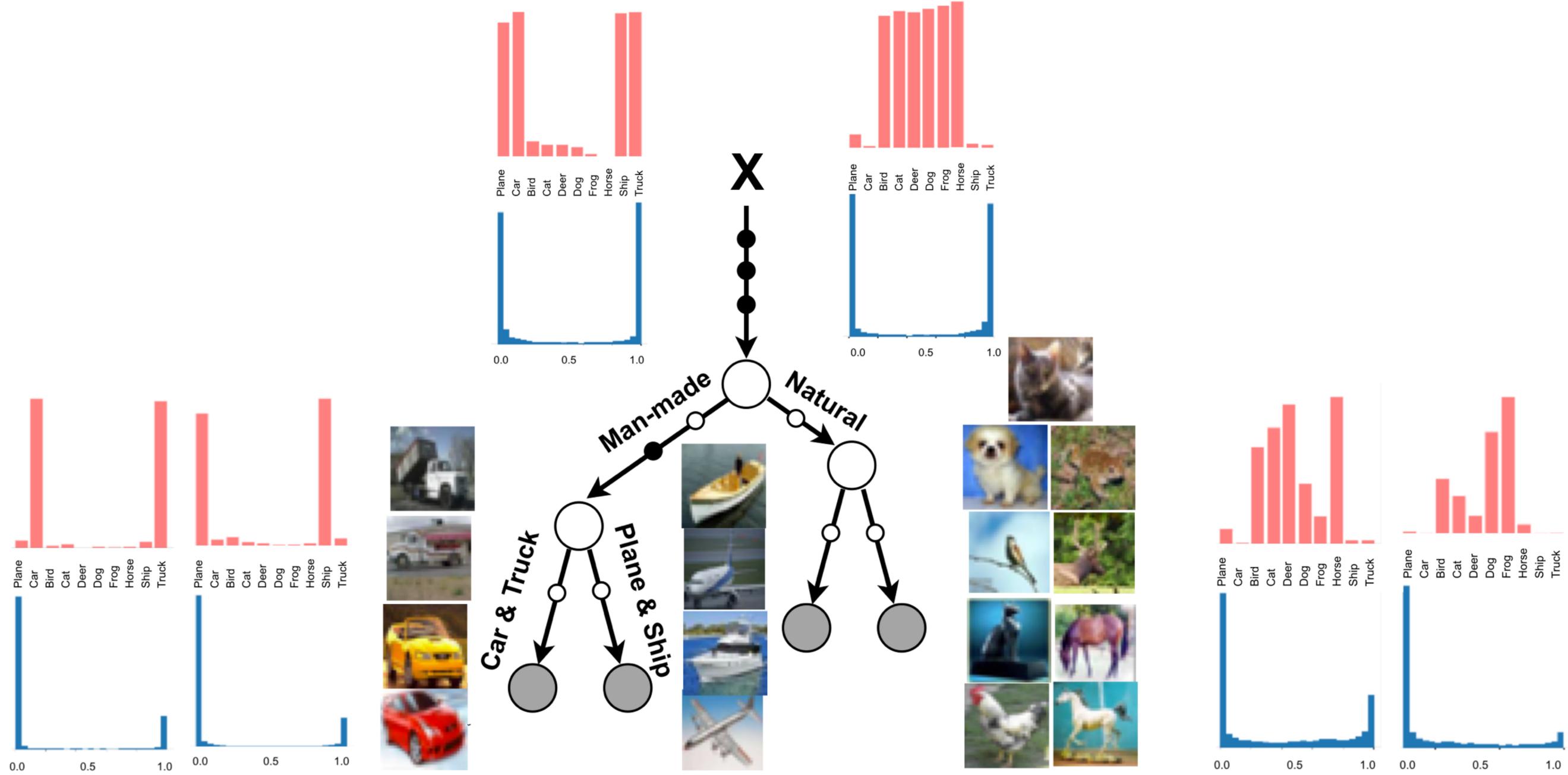
Adaptive Model Complexity

- ANTs can tune the architecture to the availability of training data.



Models are trained on subsets of size 50, 250, 500, 2.5k, 5k, 25k, 45k examples.

Unsupervised Hierarchical Clustering



Please come & see me at poster #82 for details!

