Wiring Up Vision: Minimizing Supervised Synaptic Updates Needed to Produce a Primate Ventral Stream

Franziska Geiger*, Martin Schrimpf*, Tiago Marquess, and James J. DiCarlo

Current top models of the primate ventral stream (ANN's) rely on a huge amount of supervised synaptic updates (number of labeled images x number of updated weights x number oftraining epochs)

Zador 2019: A child would have to ask a question every second of her life to receive a comparable amount of labeled data

Seibert 2018: No more than 4 months - 10 million seconds - of waking visual experience is needed to reach adult-level primate inferior-temporal cortex
> Current best model of the ventral stream would require 5.5 image-label pairs per second which is unrealistic

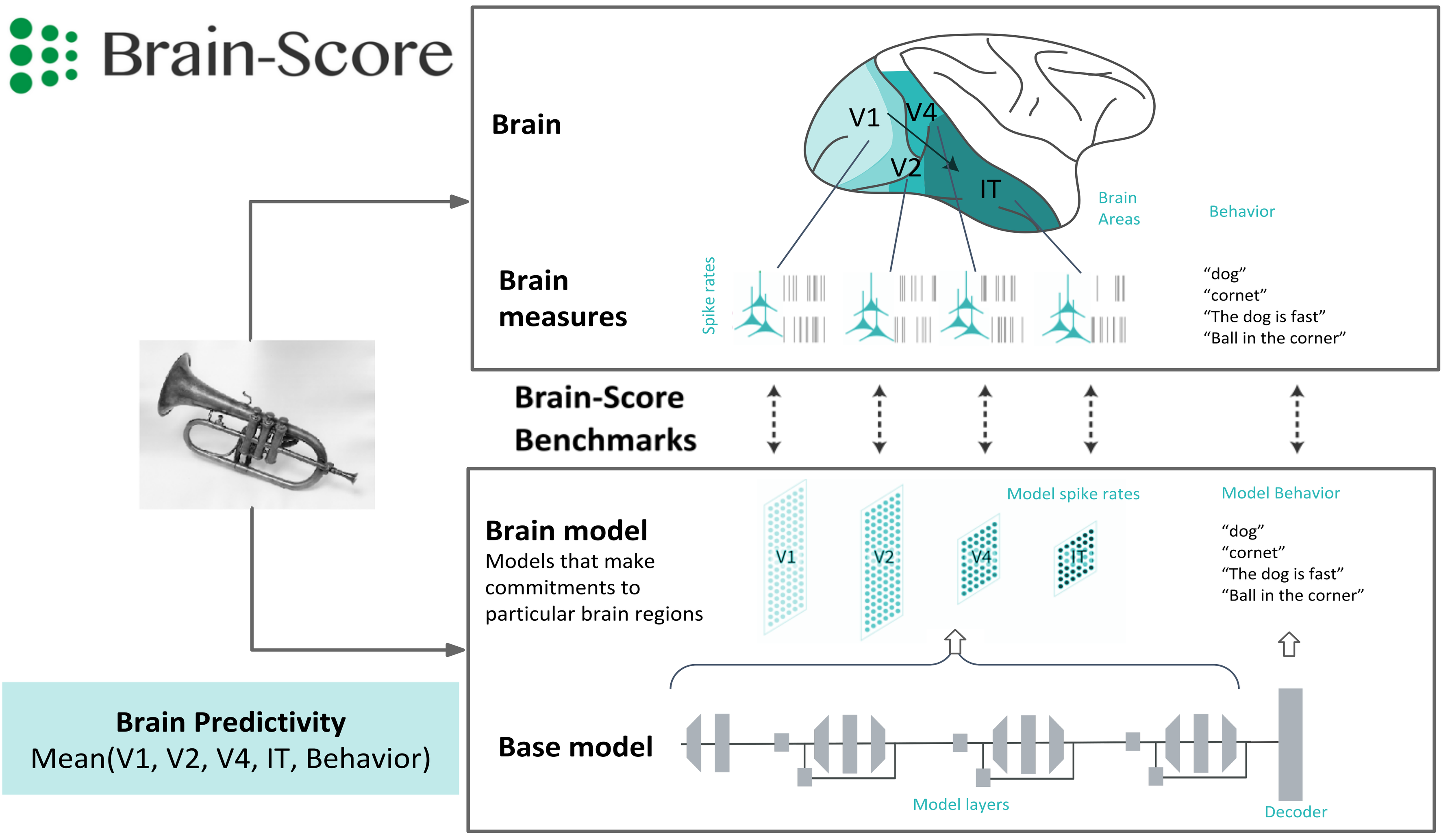
Goal:
Build models, that provide a basic understanding of how the ventral stream 'wires itself up'

Contribution
We propose a framework modelling the development of the visual system as a combination of:
- Evolution as the model's initial 'at-birth' state (1) and
- Developmental learning as model updates based on visual experience (2 + 3).

We fix the architecture to CORnet-S (Kubilius, Schimpf, et al. NeurIPS 2019), the currently most accurate model of adult primate visual processing

Benchmark artificial neural networks as models of the ventral stream

A framework mapping artifical model's activations to brain recodings for all ventral stream regions resulting in a brain predictivity score per region. In addition the model's outputs are mapped to monkey object recognition behavior to produce a behavior predictivity score.

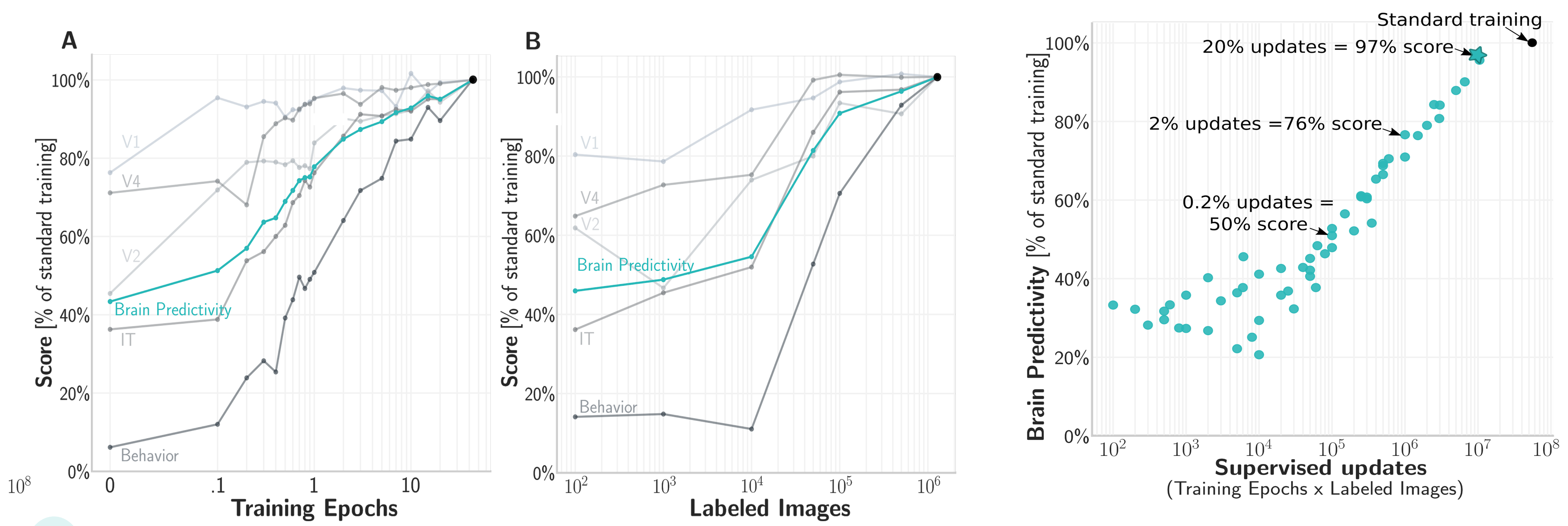


Schrimpf, Kubilius et al. 2020 (bioRxiv preprint)

1 Minimizing Images and Epochs

We explore how many supervised updates are needed to produce a primate ventral stream by reducing the number of images and/or epochs.

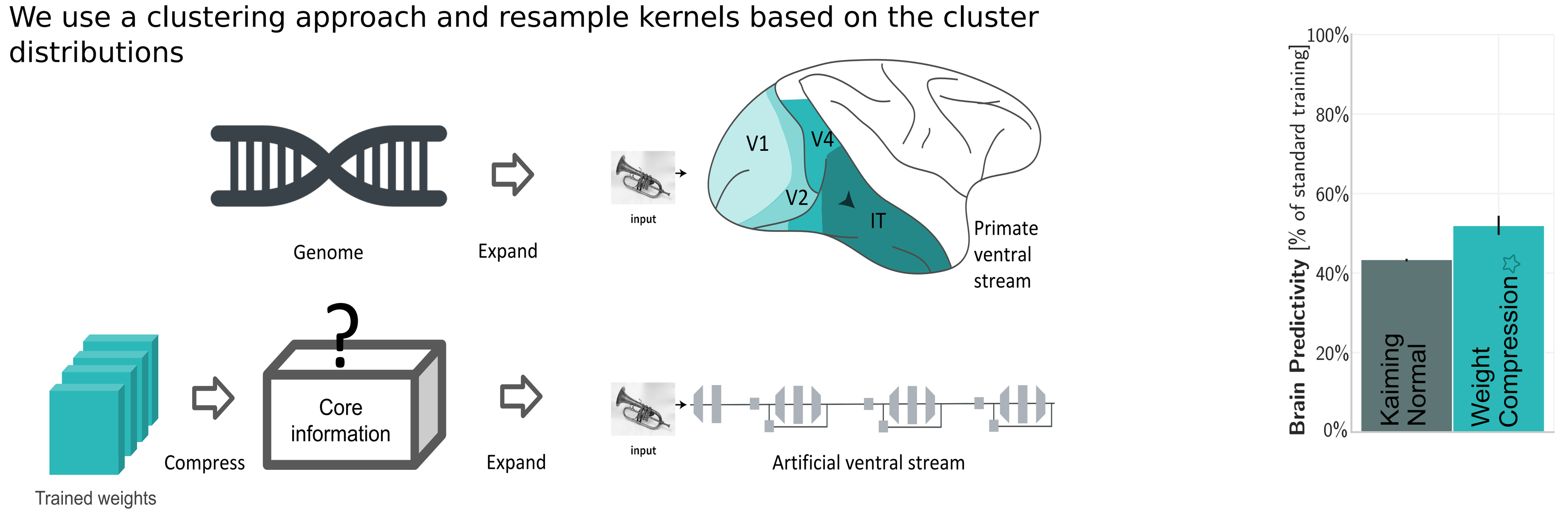
Close to 80% Brain Predictivity when training for 1 epoch or with 50,000 images



2 Advanced Initialization - Weight Compression (WC)

We compress trained weights by extracting distributions, which are then used to sample new weights (WC). The compressed weights represent the information encoded in the genome, used to set the synaptic connectivity of the ventra stream at its birth state.
We use a clustering approach and resample kernels based on the cluster distributions

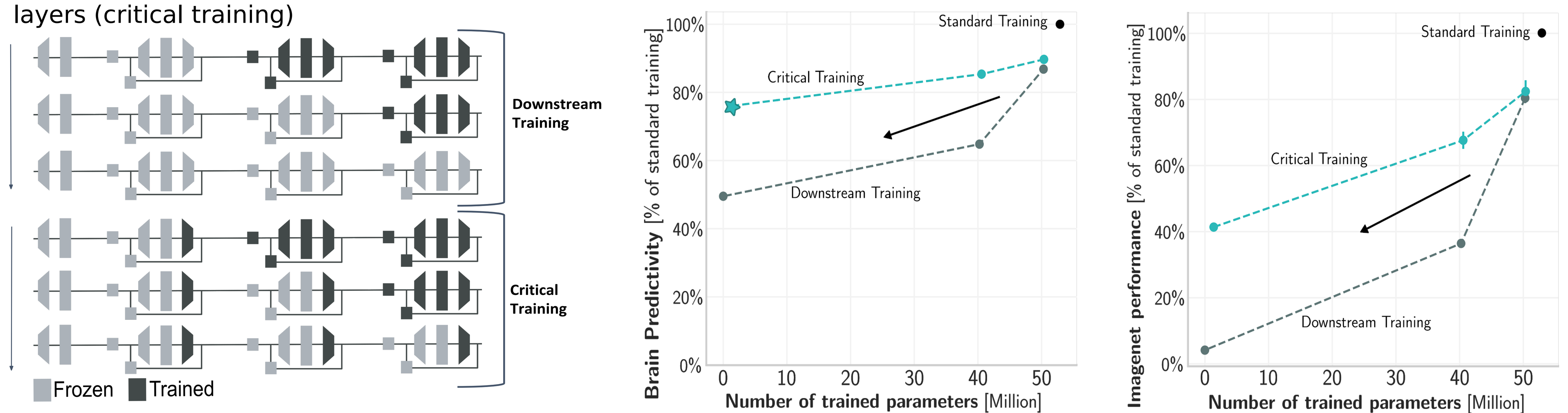
Already >40% brain predictivity with a birth-state model, which we can improve further with a Weight Compression technique



3 Training Reduction - Critical Training (CT)

Reducing the amount of updated parameters by incrementally freezing model submodules (downstream training) but continue training thin downstream layers (critical training)

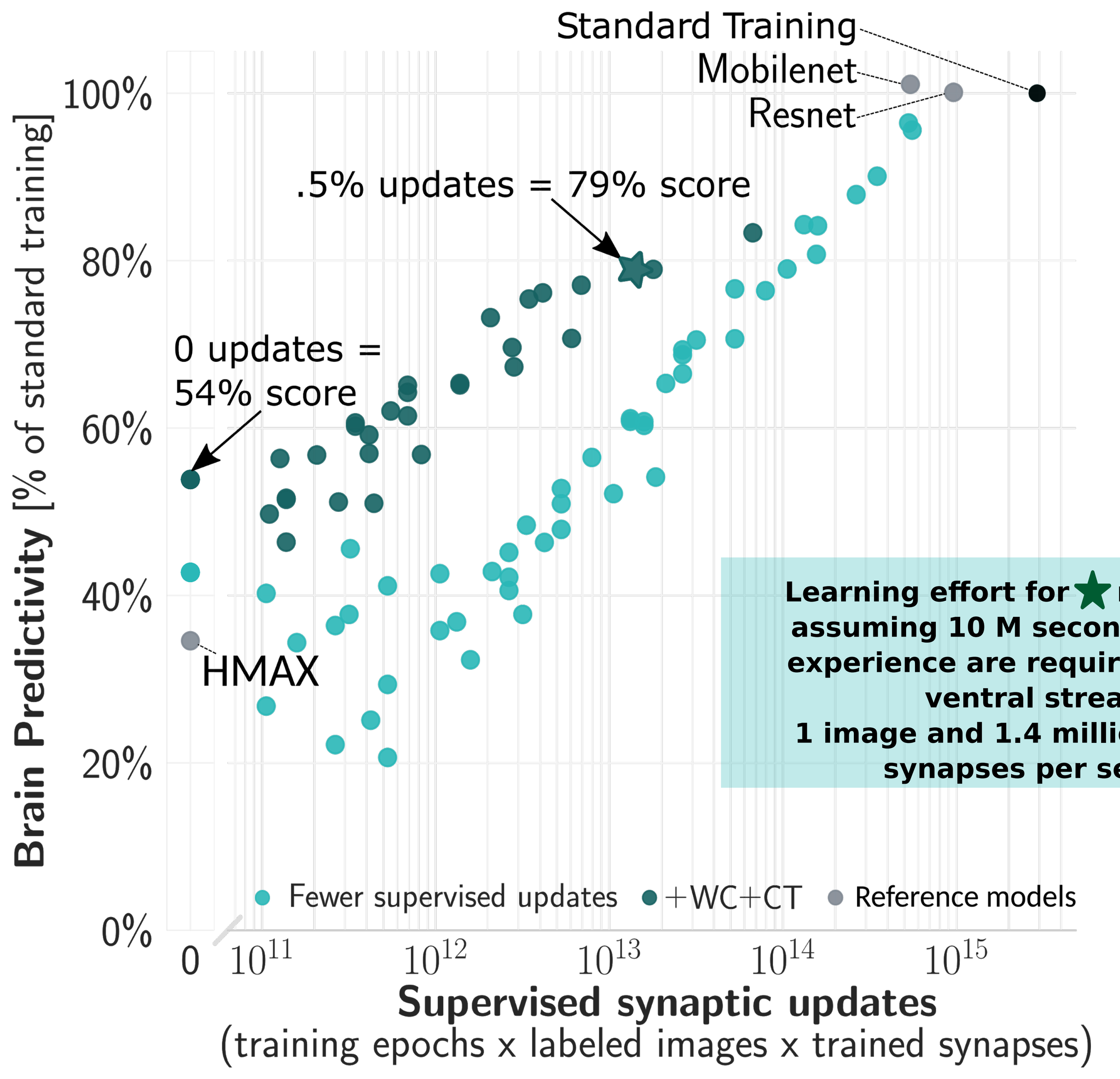
We achieve 78% Brain Predictivity when only training 3% of all parameters



1 + 2 + 3

Small number of supervised synaptic updates needed to produce a primate ventral stream

We combine the approaches by initializing from the WC distributions (2), updating only downstream layers (3) and training on a varying amount of labeled images and epochs (1): WC + CT. This yields higher brain predictivity on supervised synaptic updates than standard training with fewer supervised updates.



Learning effort for ★ model when assuming 10 M seconds of visual experience are required for adult ventral stream:
1 image and 1.4 million updated synapses per second

Summary

Relative to the current leading model of the adult ventral stream, we demonstrate a reduction of supervised weight updates using three complementary strategies:
- Only 2% of supervised updates (epochs and images) are needed to achieve~80% of the match to adult ventral stream
- With improving the random distribution of synaptic connectivity 54% of the brain match can already be achieved "at birth" (i.e. notraining at all)
- When training only~5% of model synapses, nearly 80% of the match to the ventral stream can be achieved.

When combining these strategies, new models achieve~80% of a fully trained model's match to the brain, while using two orders of magnitude fewer supervised synaptic updates

Future work

- Replace back-propagation update mechanism with a more biological plausible update mechanism
- Explore unsupervised learning apporaches
- Transfer to other systems (auditory systems)