



A brain-inspired
measure for assessing
the quality of deep net
representations

Blake Richards

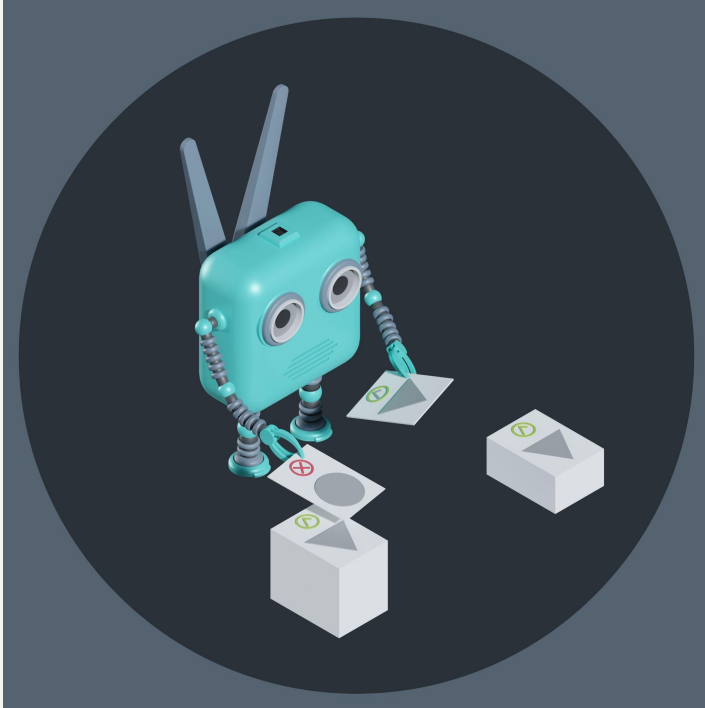
World Summit AI Americas
April 25, 2024



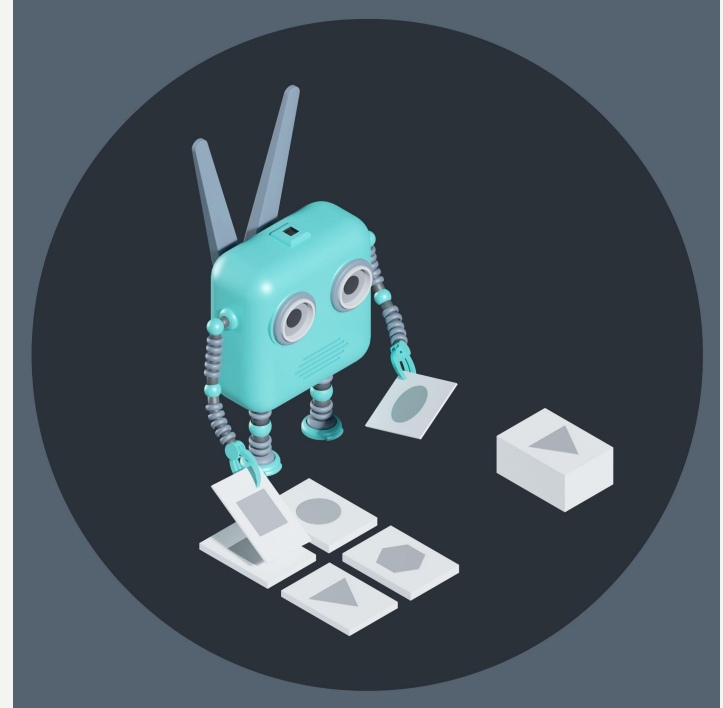
McGill

CIFAR

The most important recent advances in machine learning are in **unsupervised** and **self-supervised** learning



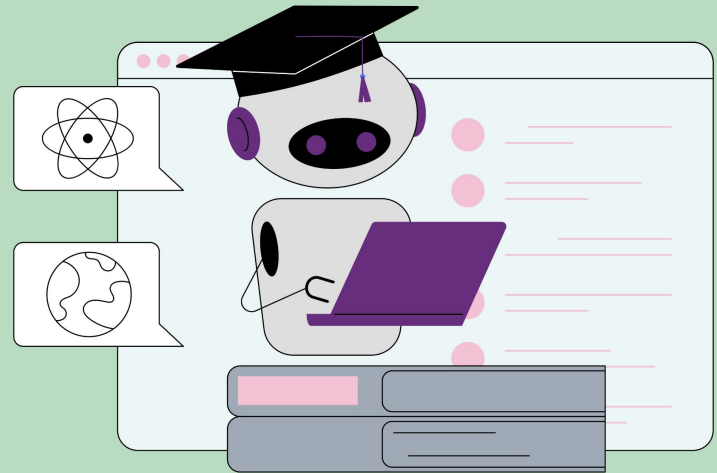
Supervised: requires human annotation



Self-supervised: no annotation required

Problem:

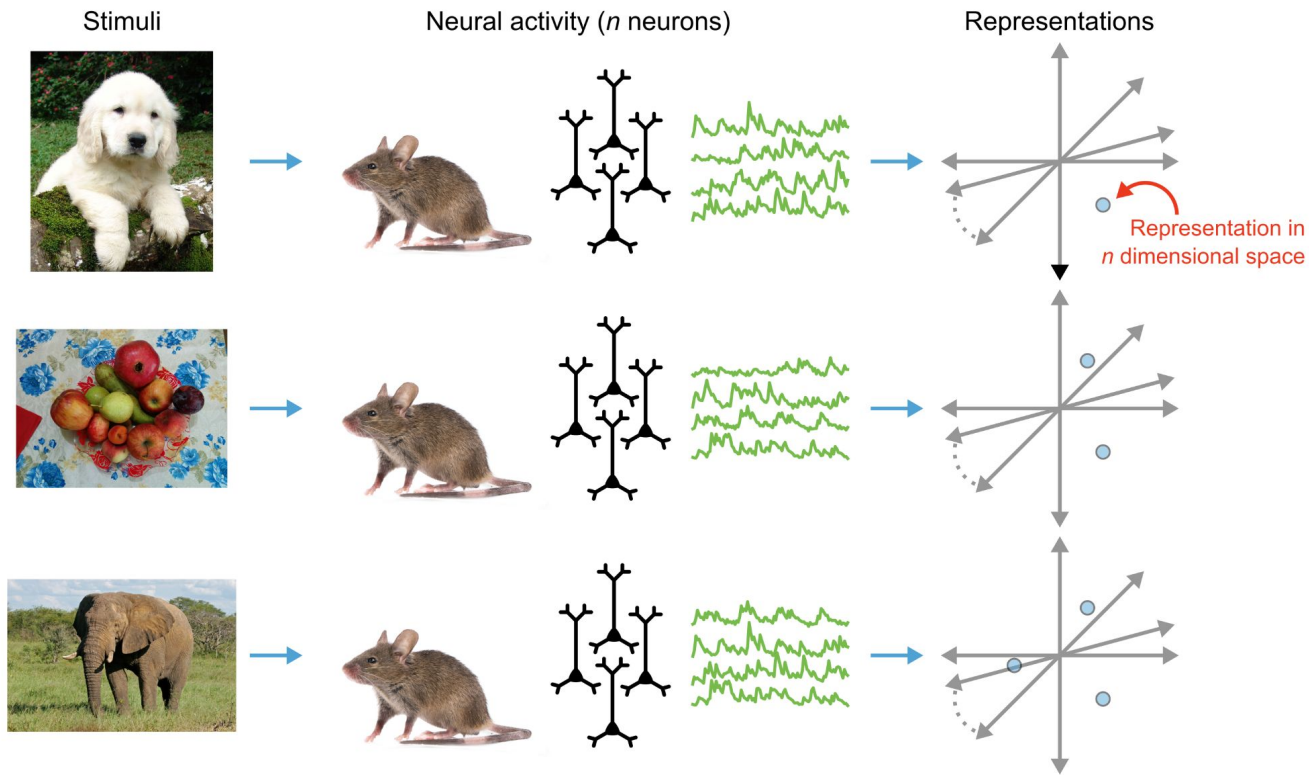
How do you know when self-supervised learning is working well?





We turned to the brain for help!

Determining the geometry of representations in the brain

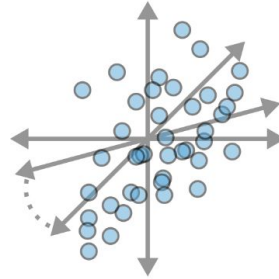


Determining the geometry of representations in the brain

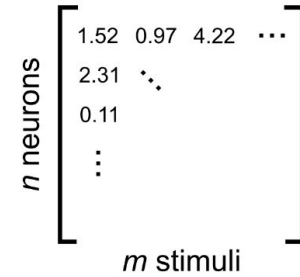
m stimuli



m representations



$m \times n$ matrix of representations

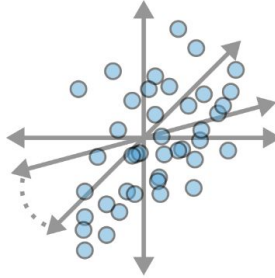


Determining the geometry of representations in the brain

m stimuli



m representations



$m \times n$ matrix of representations

$$\begin{matrix} n \text{ neurons} \\ \left[\begin{array}{cccc} 1.52 & 0.97 & 4.22 & \dots \\ 2.31 & \ddots & & \\ 0.11 & & & \\ \vdots & & & \end{array} \right] \\ m \text{ stimuli} \end{matrix}$$

$$\begin{bmatrix} 1.52 & 0.97 & 4.22 & \dots \\ 2.31 & \ddots & & \\ 0.11 & & & \\ \vdots & & & \end{bmatrix} \cdot \begin{bmatrix} 1.52 & 2.31 & 0.11 & \dots \\ 0.97 & \ddots & & \\ 4.22 & & & \\ \vdots & & & \end{bmatrix}^T = \begin{bmatrix} \text{cross-correlation} \\ \text{matrix} \end{bmatrix}$$

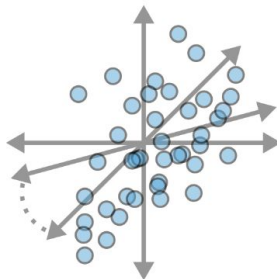
← Calculate the eigenvalues of this matrix

Determining the geometry of representations in the brain

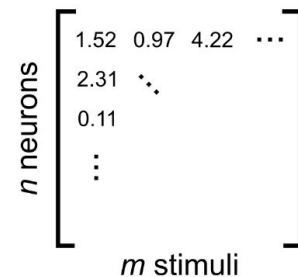
m stimuli



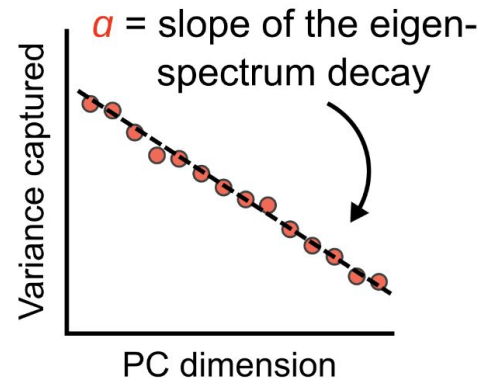
m representations



$m \times n$ matrix of representations



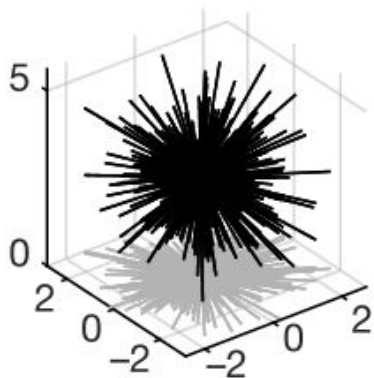
$$\begin{bmatrix} 1.52 & 0.97 & 4.22 & \dots \\ 2.31 & \ddots & & \\ 0.11 & & & \\ \vdots & & & \end{bmatrix} \cdot \begin{bmatrix} 1.52 & 2.31 & 0.11 & \dots \\ 0.97 & \ddots & & \\ 4.22 & & & \\ \vdots & & & \end{bmatrix}^T = \begin{bmatrix} \text{cross-correlation} \\ \text{matrix} \end{bmatrix}$$



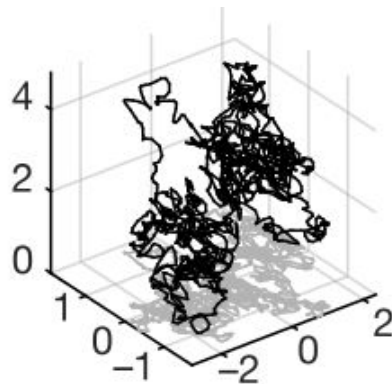
Determining the geometry of representations in the brain

Representations

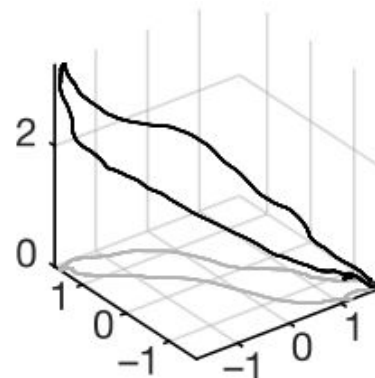
Not smooth, high-D



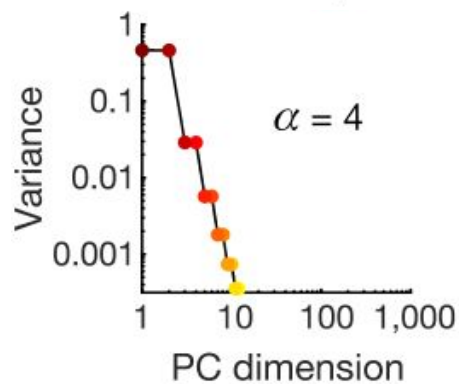
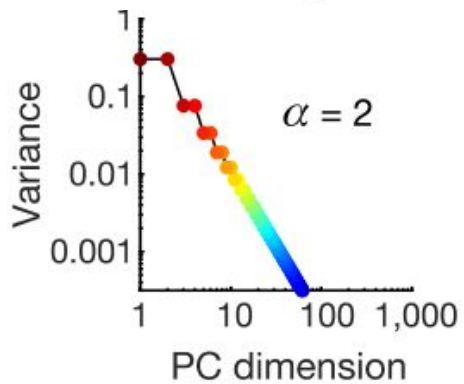
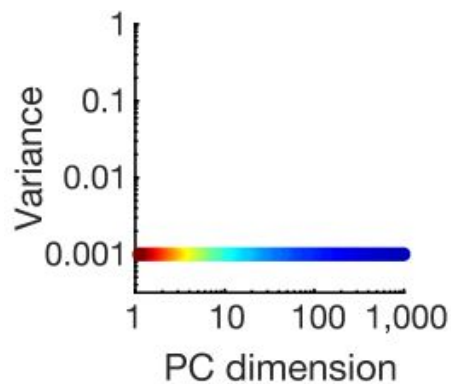
Semi-smooth, med-D



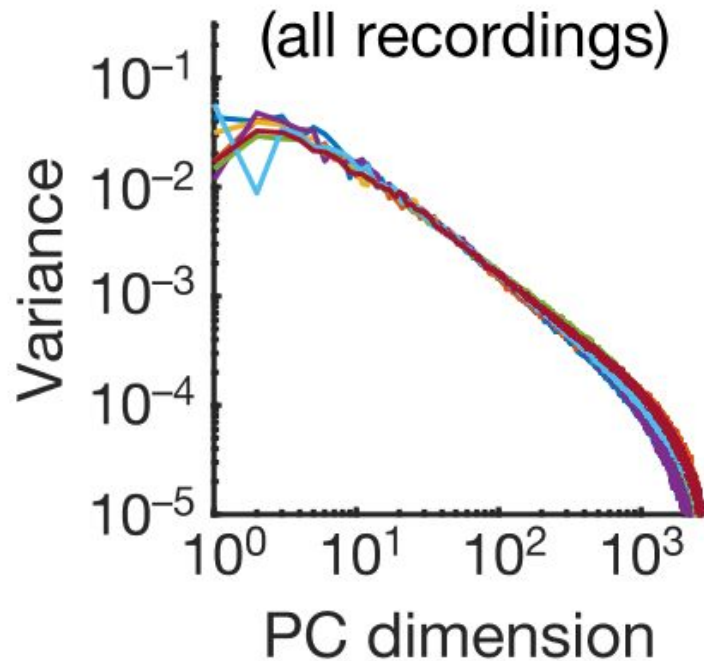
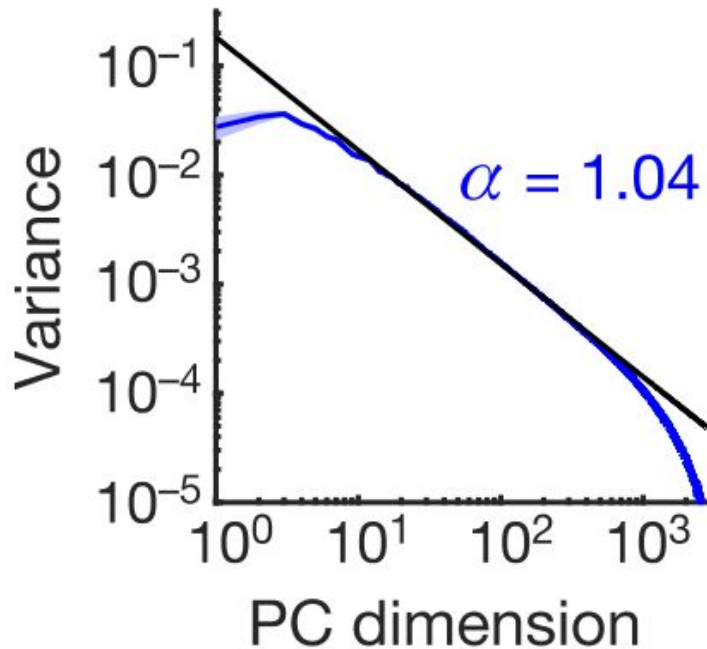
Smooth, low-D



Eigenspectra



Visual cortex has an eigenspectrum decay of roughly 1



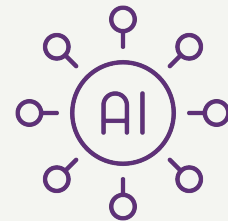
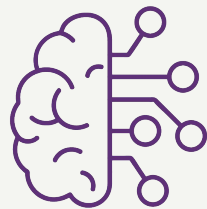
Stringer et al. (2019) *Nature*



Hypothesis:

the geometry of representations observed in the brain is best for general performance on natural data

How to test this hypothesis?



1. Predict performance

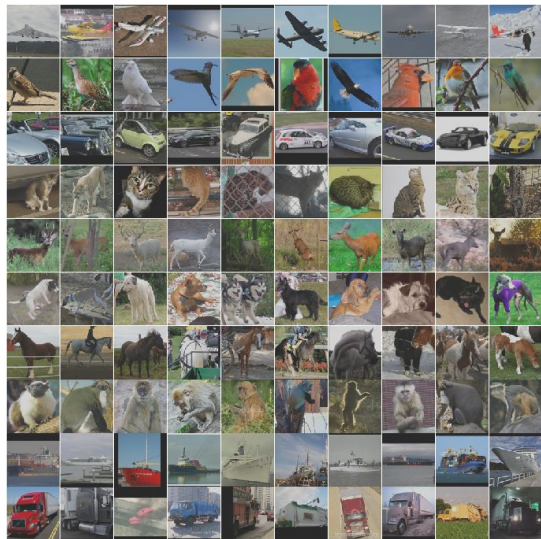
Can we predict how well a deep neural network will perform on new data by comparing to brain's α ?

2. Model selection

Can we select the best AI models for downstream applications using α ?

α close to 1 predicts better performance

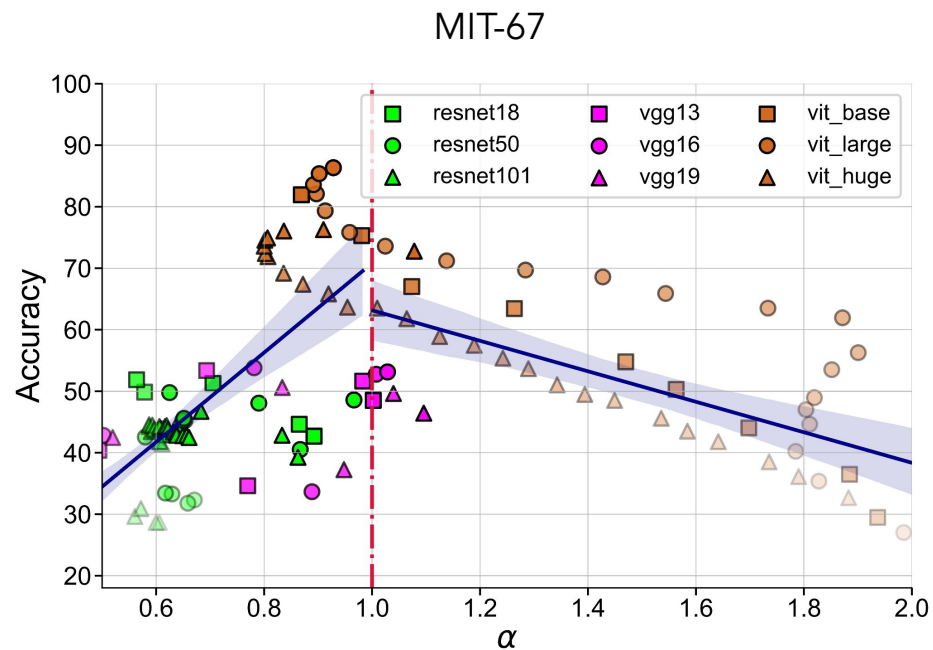
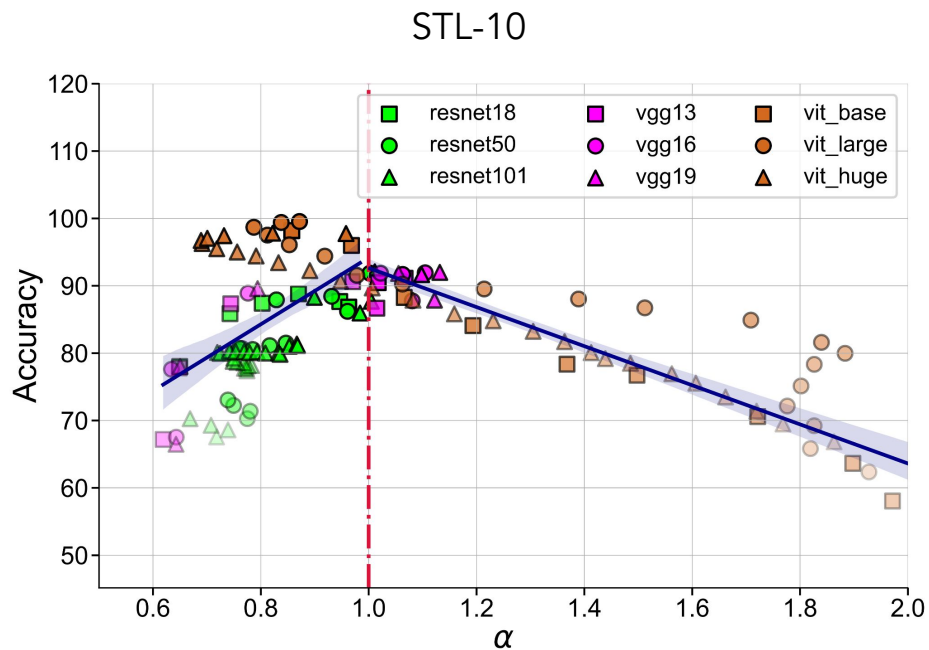
STL-10



MIT-67
















α close to 1 predicts better performance

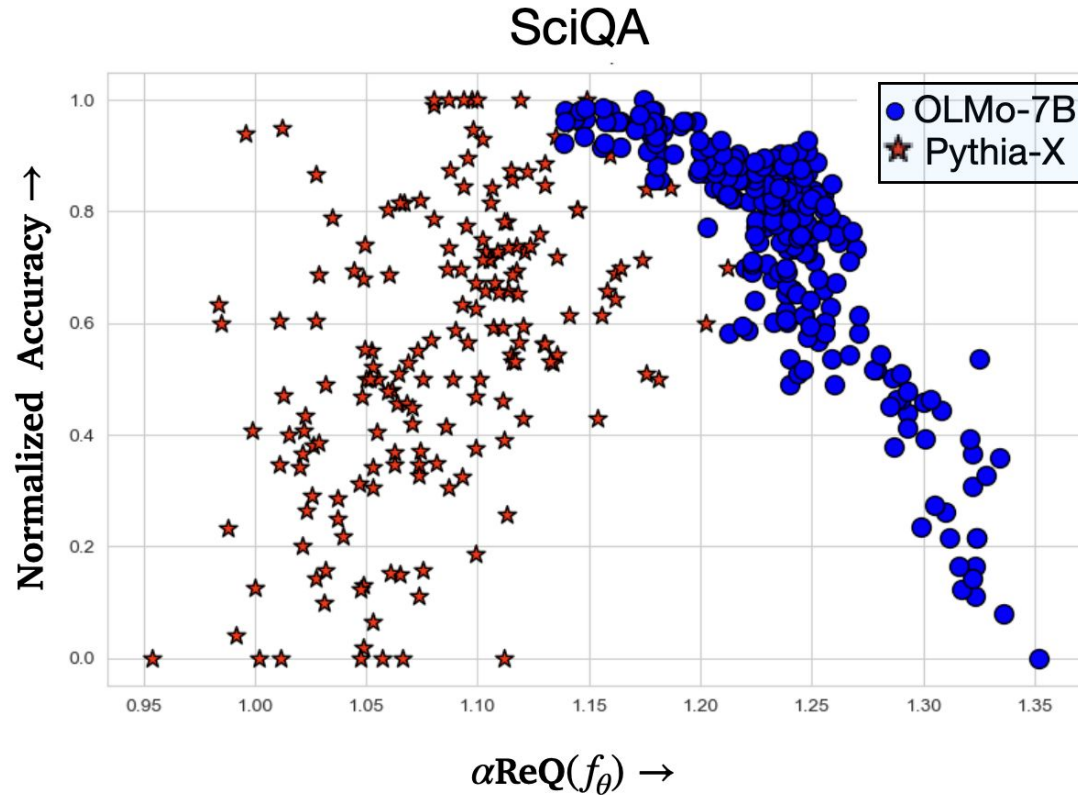


α close to 1 predicts better performance

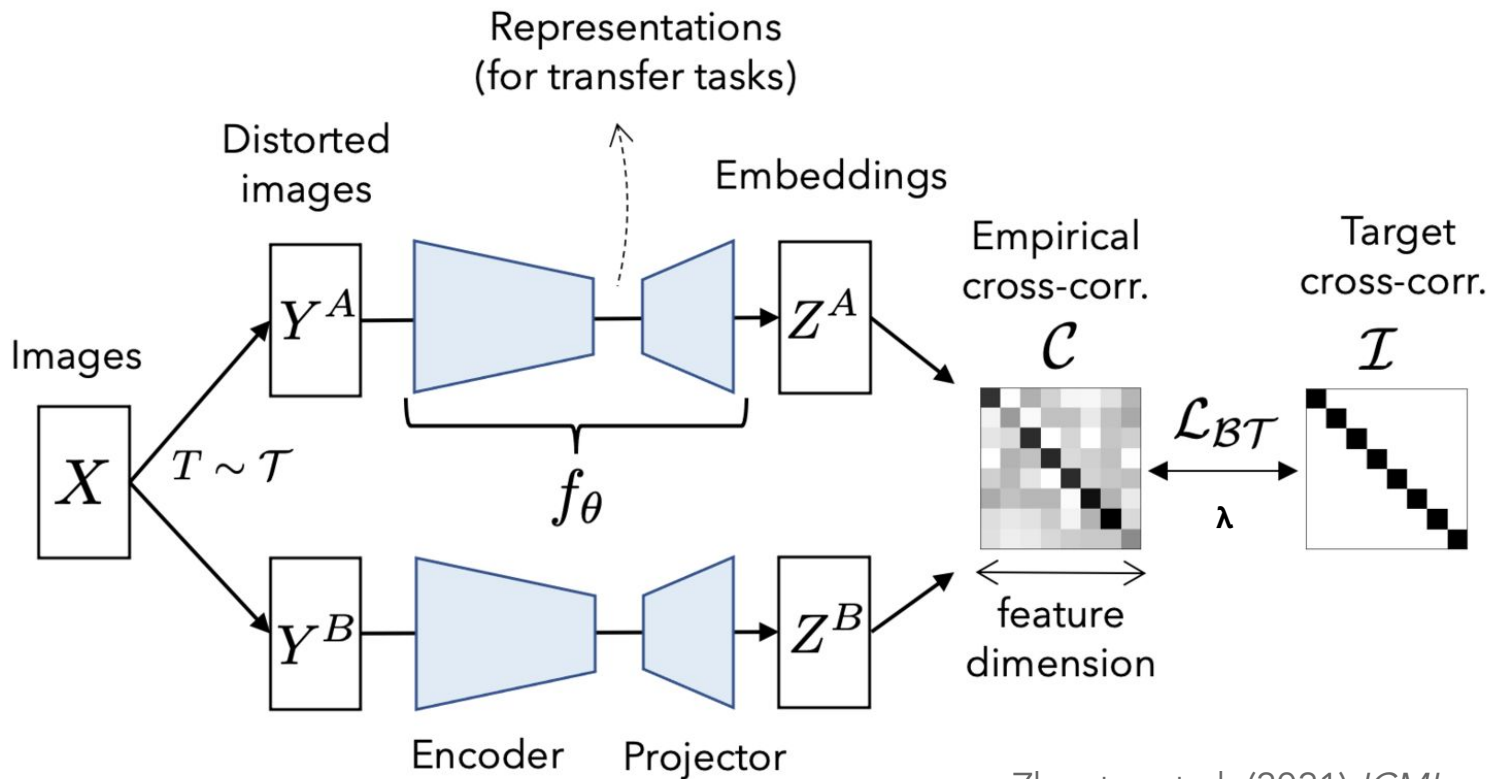
SciQA

| | | | | |
|--|--|--|---|---|
| Biology Genes to traits Classification Adaptations Traits and heredity Ecosystems Classification Scientific names Heredity Ecological interactions Cells Plants Animals Plant reproduction |  Physics Materials Magnets Velocity and forces Force and motion Particle motion and energy Heat and thermal energy States of matter Kinetic and potential energy Mixture |  Geography State capitals Geography Maps Oceania: geography Physical Geography The Americas: geography Oceans and continents Cities States |  History Colonial America English colonies in North America The American Revolution |  Civics Social skills Government The Constitution |
| | | | | |
| Earth Science Weather and climate Rocks and minerals Astronomy Fossils Earth events Plate tectonics |  Chemistry Solutions Physical and chemical change Atoms and molecules Chemical reactions | Writing Strategies Supporting arguments Sentences, fragments, and run-ons Word usage and nuance Creative techniques Audience, purpose, and tone Pronouns and antecedents Persuasive strategies Editing and revising |  Vocabulary Categories Shades of meaning Comprehension strategies Context clues |  Verbs Verb tense |
| |  Engineering Designing experiments Engineering practices | | | |
| |  Units and Measurement Weather and climate |  Visual elements Opinion writing |  Figurative Language Literary devices |  Global Studies Society and environment |
| | | | |  Punctuation Fragments |
| | | | | Phonology Rhyming |
| | | | | Reference Research skills |

α close to 1 predicts better performance

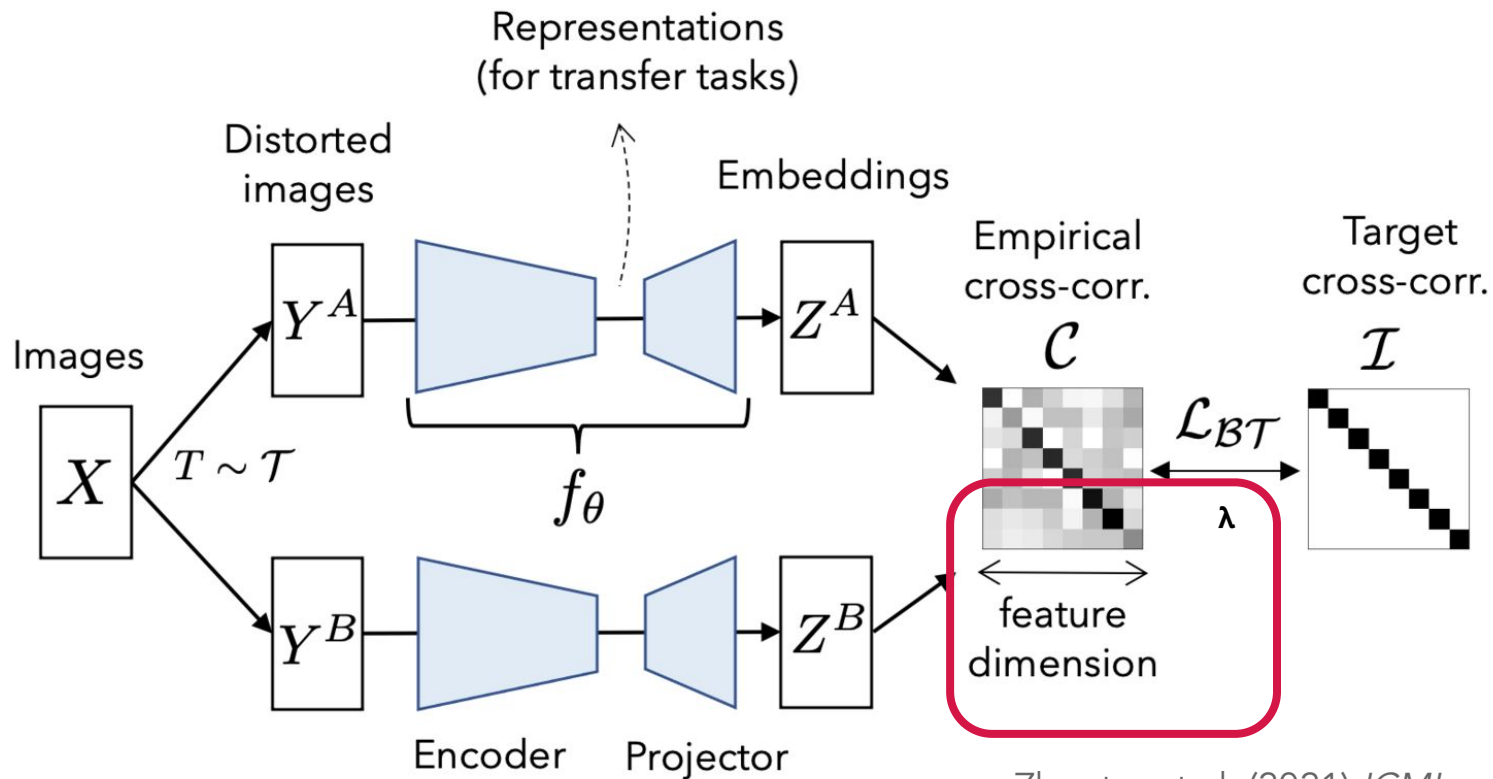


Model selection for Barlow Twins



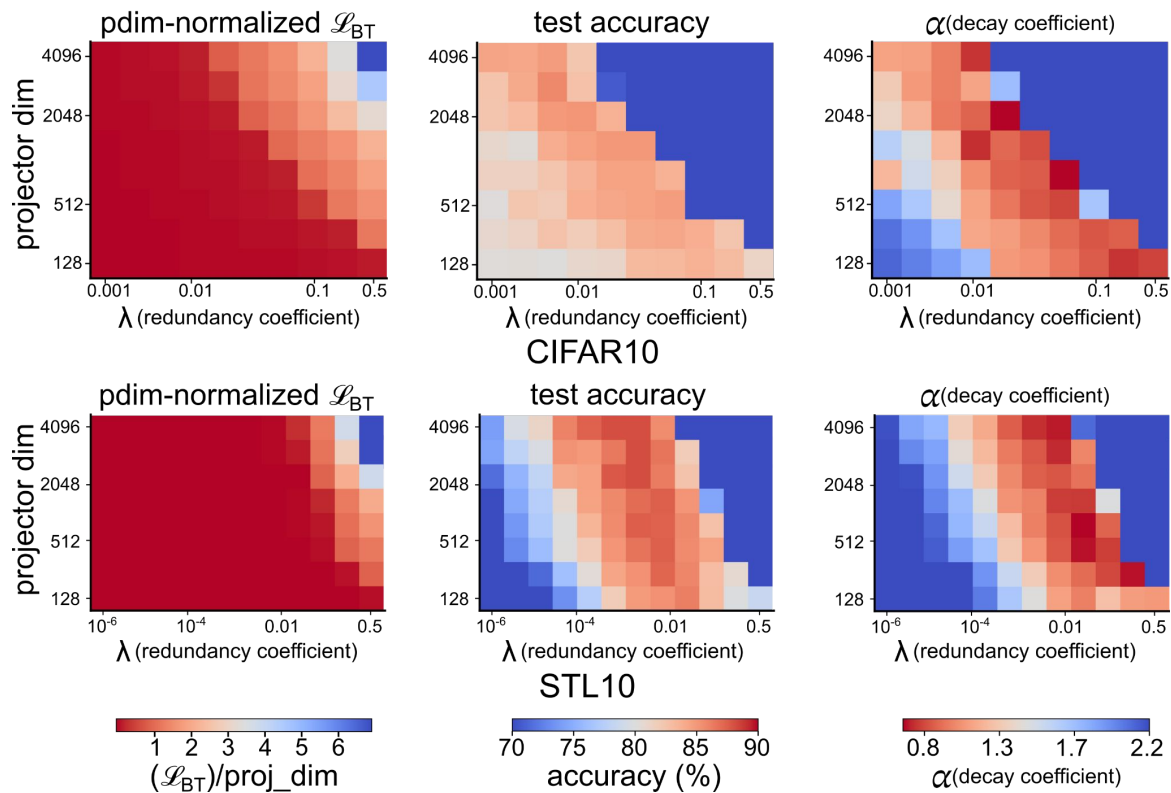
Zbontar et al. (2021) *ICML*

Model selection for Barlow Twins



Zbontar et al. (2021) *ICML*

The values for λ that bring α close to 1 lead to the best accuracy



Conclusion

Self-supervised learning is critical to modern AI, but it is not obvious how best to measure the quality of representations learned by self-supervised learning - *we took inspiration from the brain*

Conclusion

Self-supervised learning is critical to modern AI, but it is not obvious how best to measure the quality of representations learned by self-supervised learning - *we took inspiration from the brain*

Geometry of reps

What is the geometry of the representations in a network (\mathbf{a})?



Conclusion

Self-supervised learning is critical to modern AI, but it is not obvious how best to measure the quality of representations learned by self-supervised learning - *we took inspiration from the brain*

Geometry of reps

What is the geometry of the representations in a network (α)?



Brain-inspiration

Brains have “goldilocks” geometry (α close to 1)



Conclusion

Self-supervised learning is critical to modern AI, but it is not obvious how best to measure the quality of representations learned by self-supervised learning - *we took inspiration from the brain*

Geometry of reps

What is the geometry of the representations in a network (α)?



Brain-inspiration

Brains have “goldilocks” geometry (α close to 1)



Prediction

We can predict model performance using α close to 1



Conclusion

Self-supervised learning is critical to modern AI, but it is not obvious how best to measure the quality of representations learned by self-supervised learning - *we took inspiration from the brain*

Geometry of reps

What is the geometry of the representations in a network (α)?



Brain-inspiration

Brains have “goldilocks” geometry (α close to 1)



Prediction

We can predict model performance using α close to 1



Selection

We can select models, i.e. tune hyperparameters, using α close to 1



Thanks for listening!



Arna Ghosh

Acknowledgements

Arna Ghosh

Arnab Kumar Mondal

Kumar Krishna Agrawal

Melody Li



McGill



NSERC
CRSNG



HEALTHY BRAINS
FOR HEALTHY LIVES