Mila

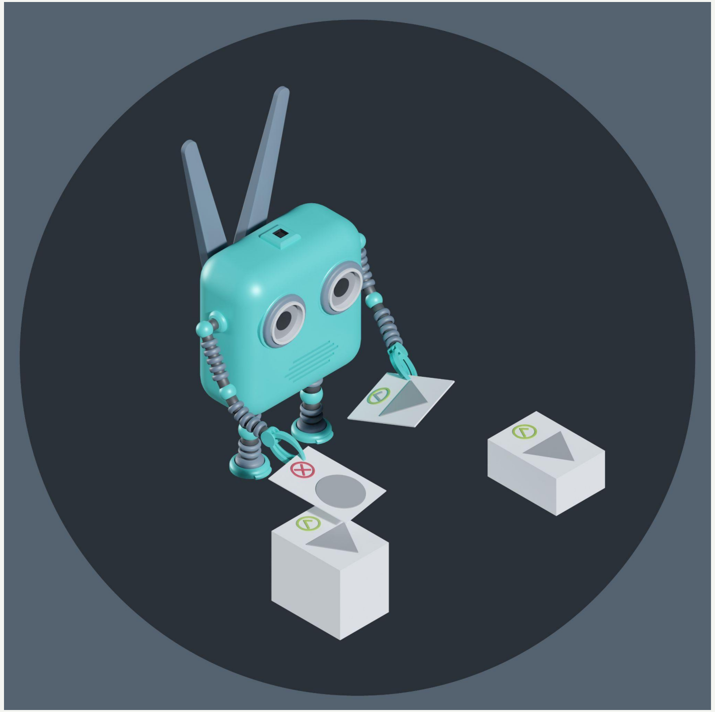
# A brain-inspired measure for assessing the quality of deep net representations

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



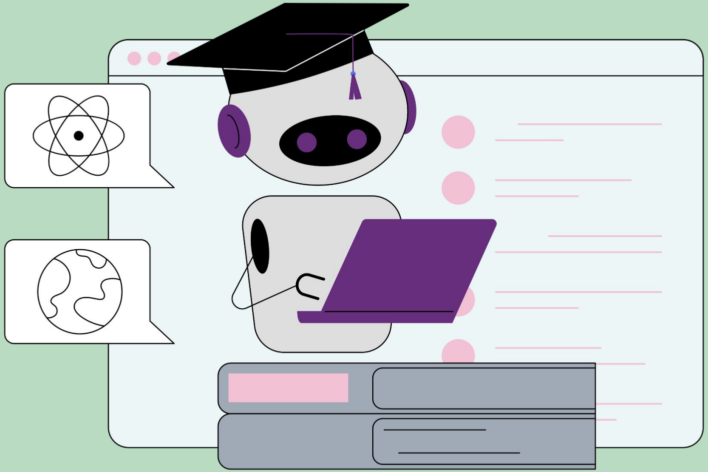
**Supervised:** requires human annotation

**Self-supervised:** no annotation required

Mila

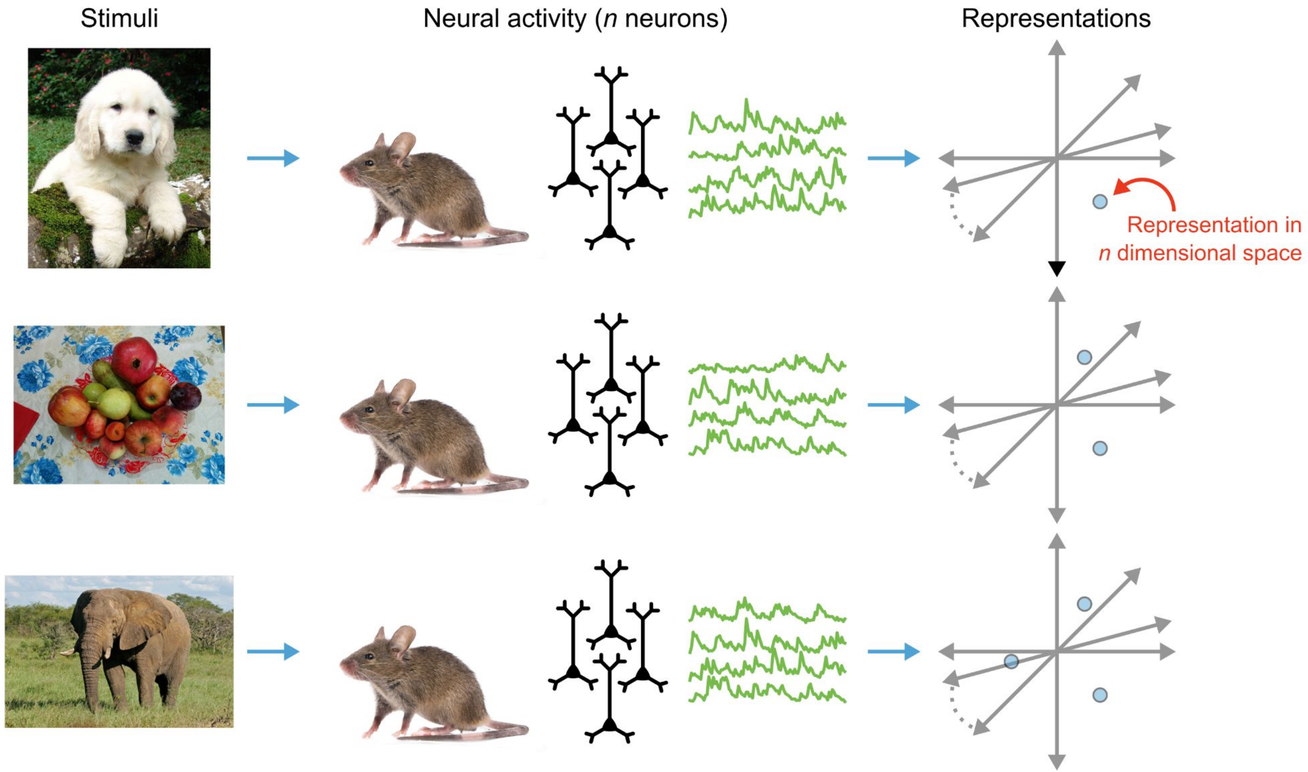
# Problem:

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



We turned to the brain for help!

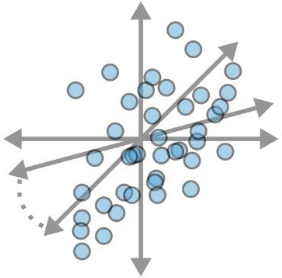
# Determining the geometry of representations in the brain



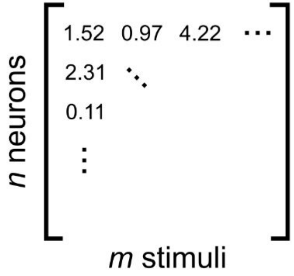
# Determining the geometry of representations in the brain

*m* stimuli

*m* representations



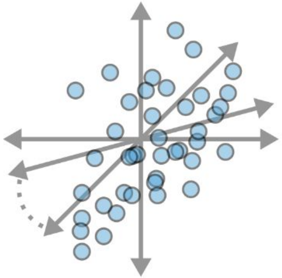
*m* x n matrix of representations



# Determining the geometry of representations in the brain

*m* stimuli

*m* representations



*m* x n matrix of representations

*n* neurons

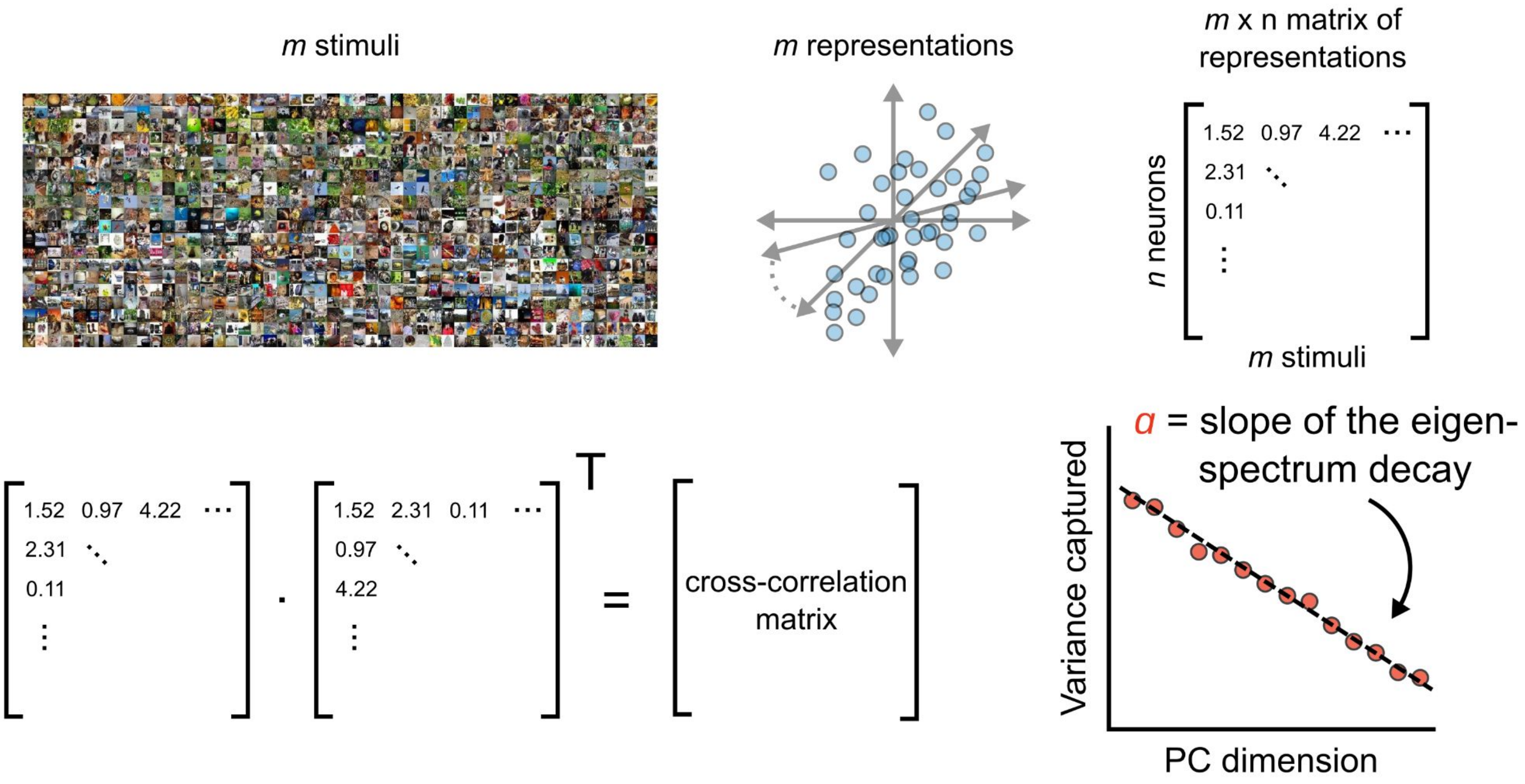
$$\begin{bmatrix} 1.52 & 0.97 & 4.22 & \cdots \\ 2.31 & \ddots & & \\ 0.11 & & & \\ \vdots & & & \end{bmatrix}$$

*m* stimuli

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

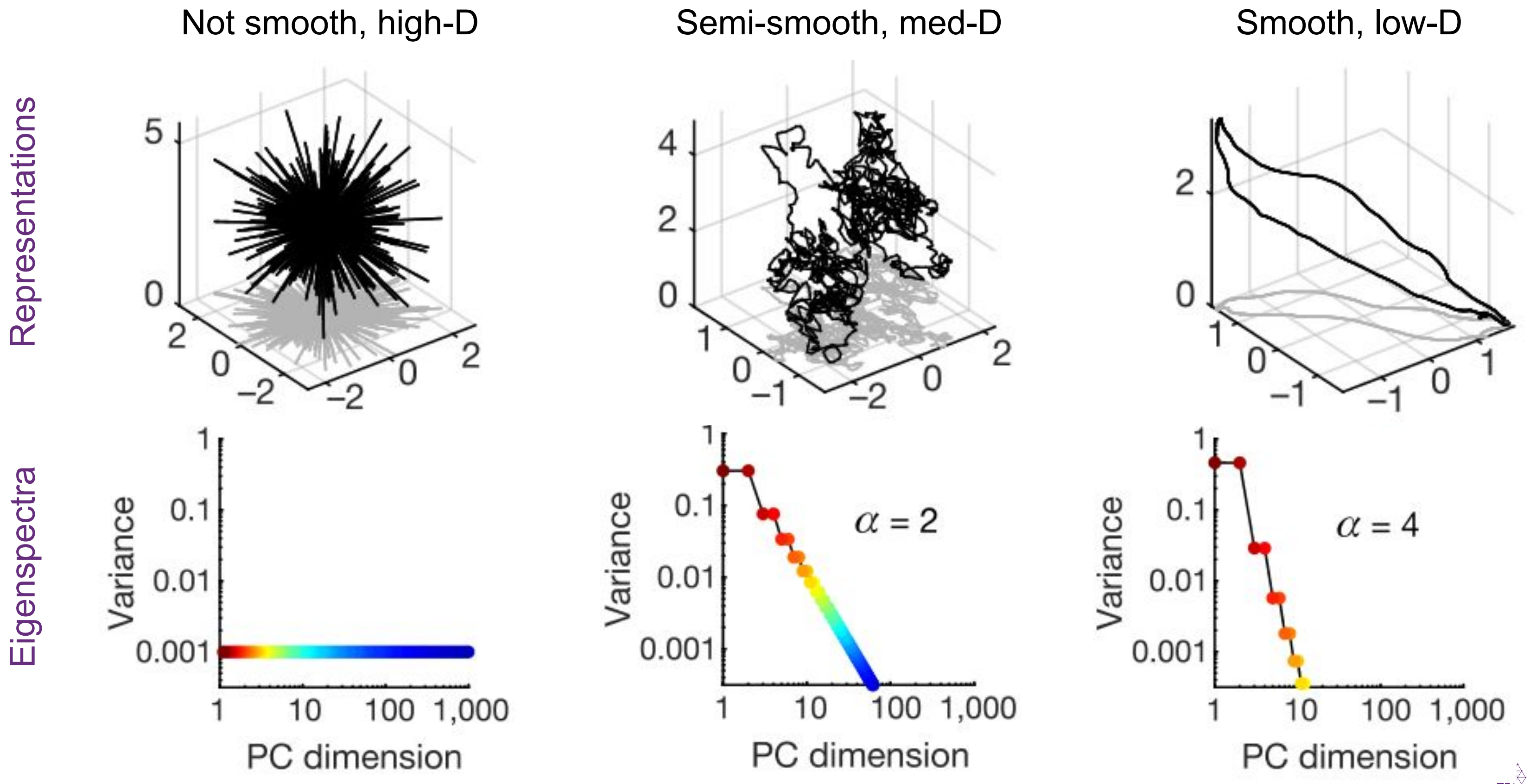
Calculate the eigenvalues of this matrix

# Determining the geometry of representations in the brain



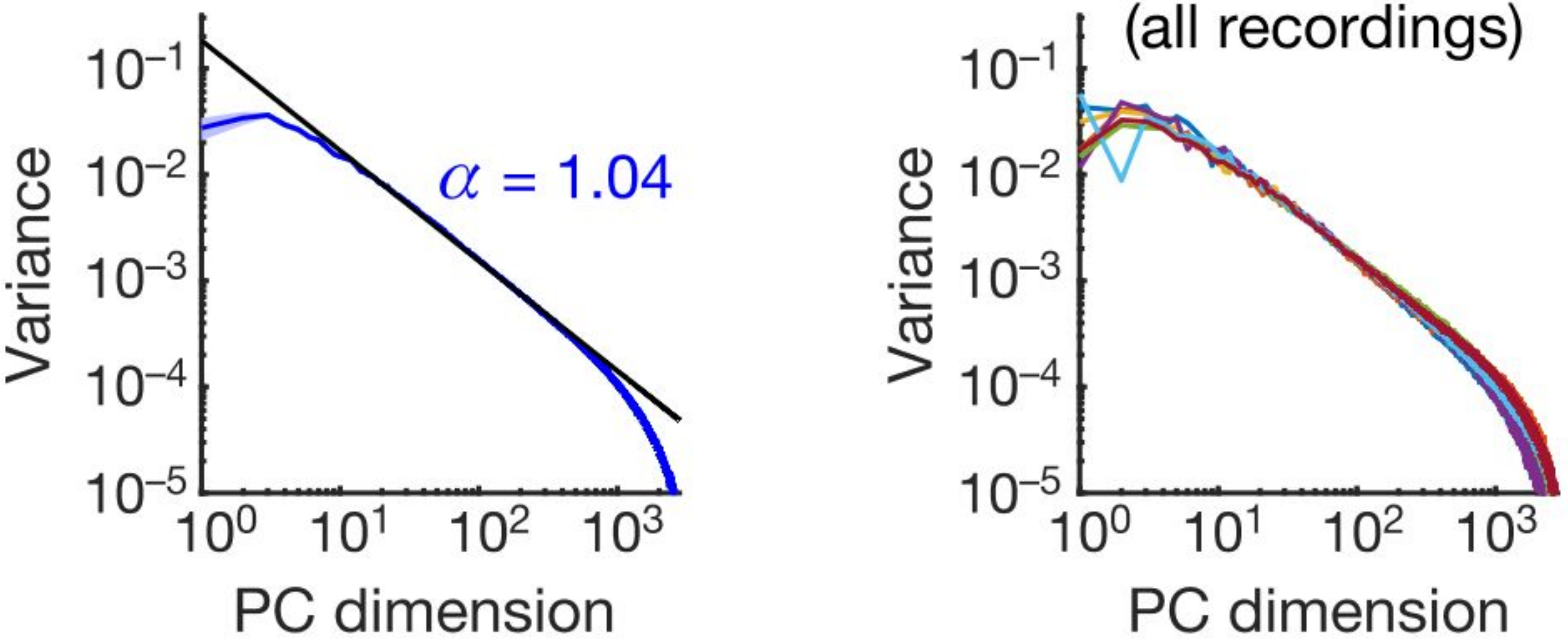

Mila

# Determining the geometry of representations in the brain




Mila

# 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

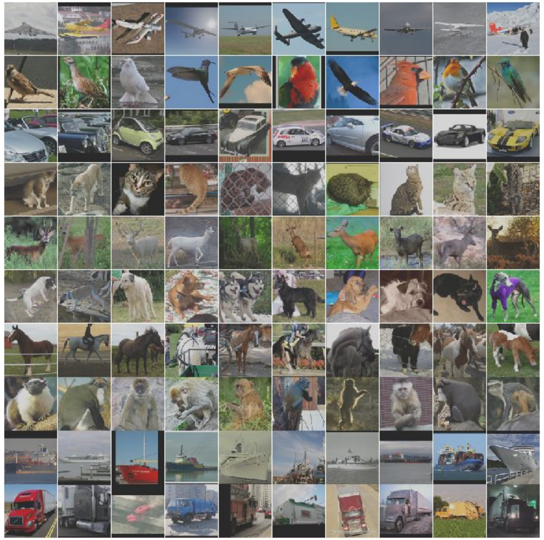
Can we predict how well a deep neural network will perform on new data by comparing to brain’s **α**?

## 2. Model selection

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

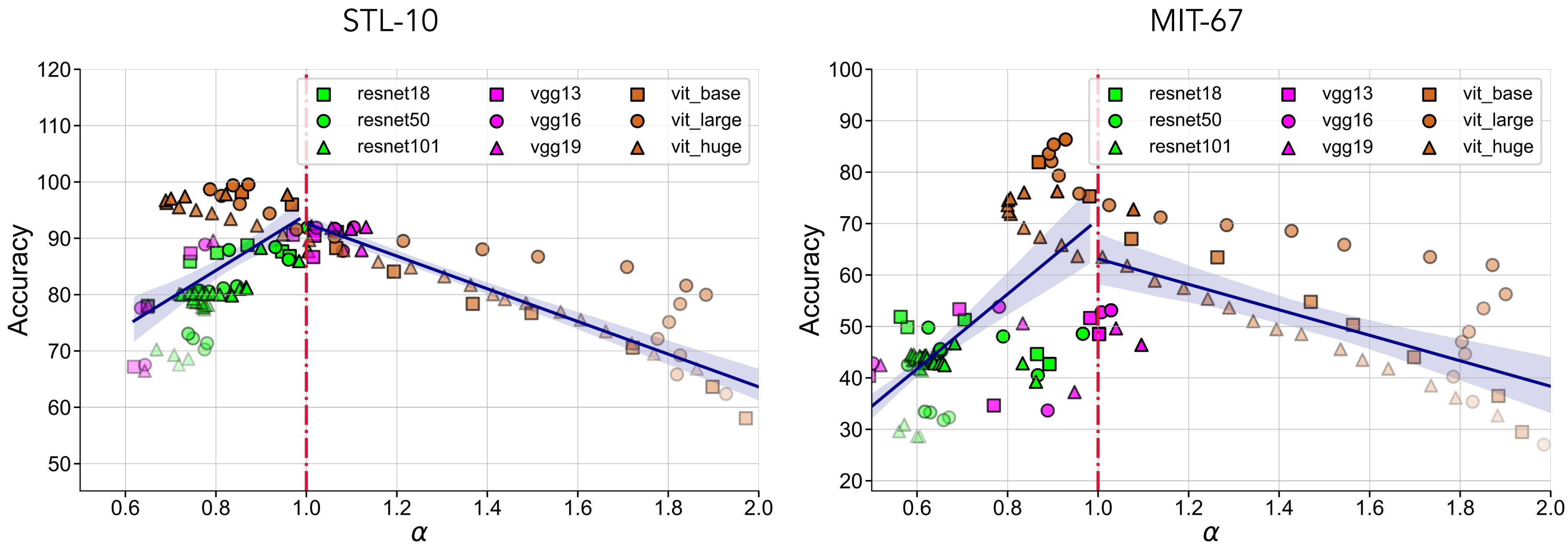
# α close to 1 predicts better performance

STL-10



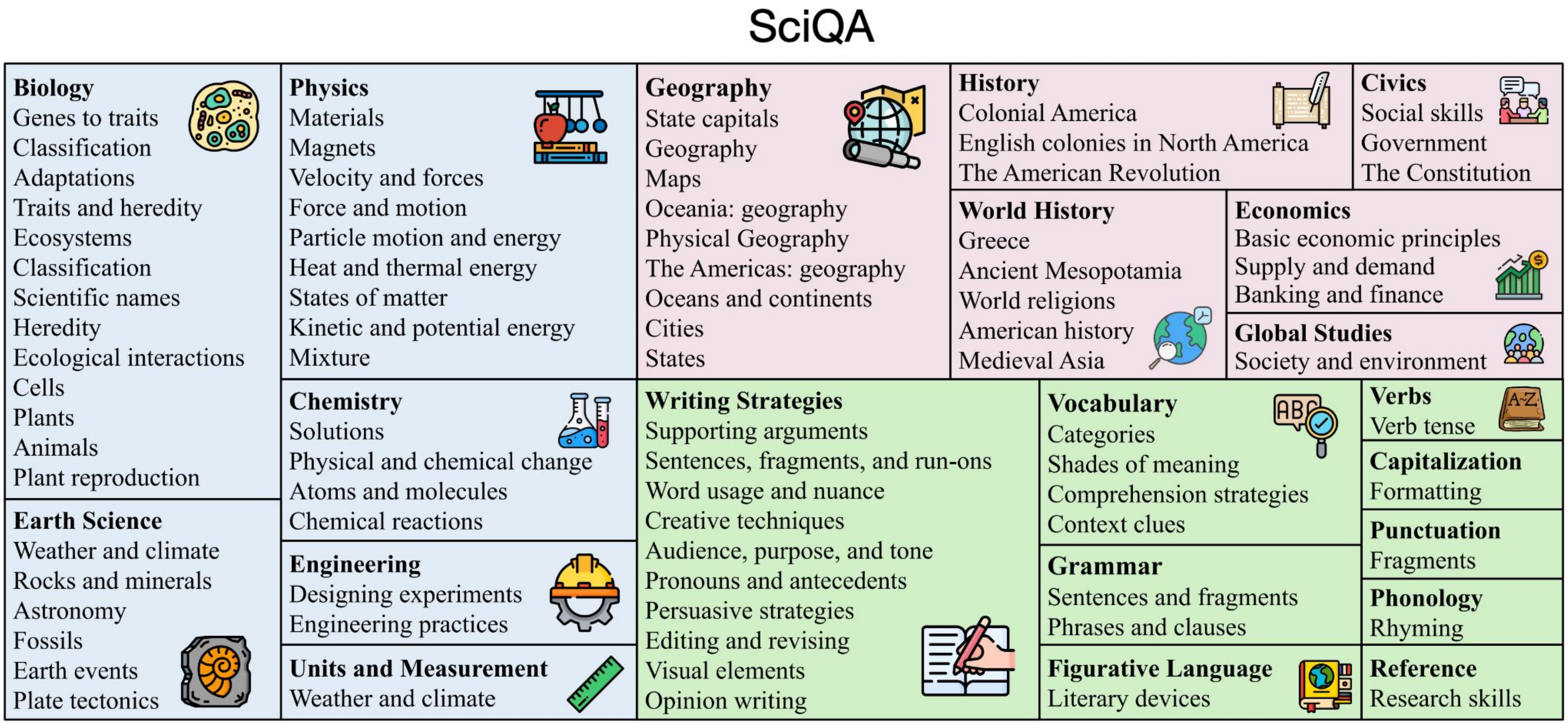
MIT-67

# α close to 1 predicts better performance

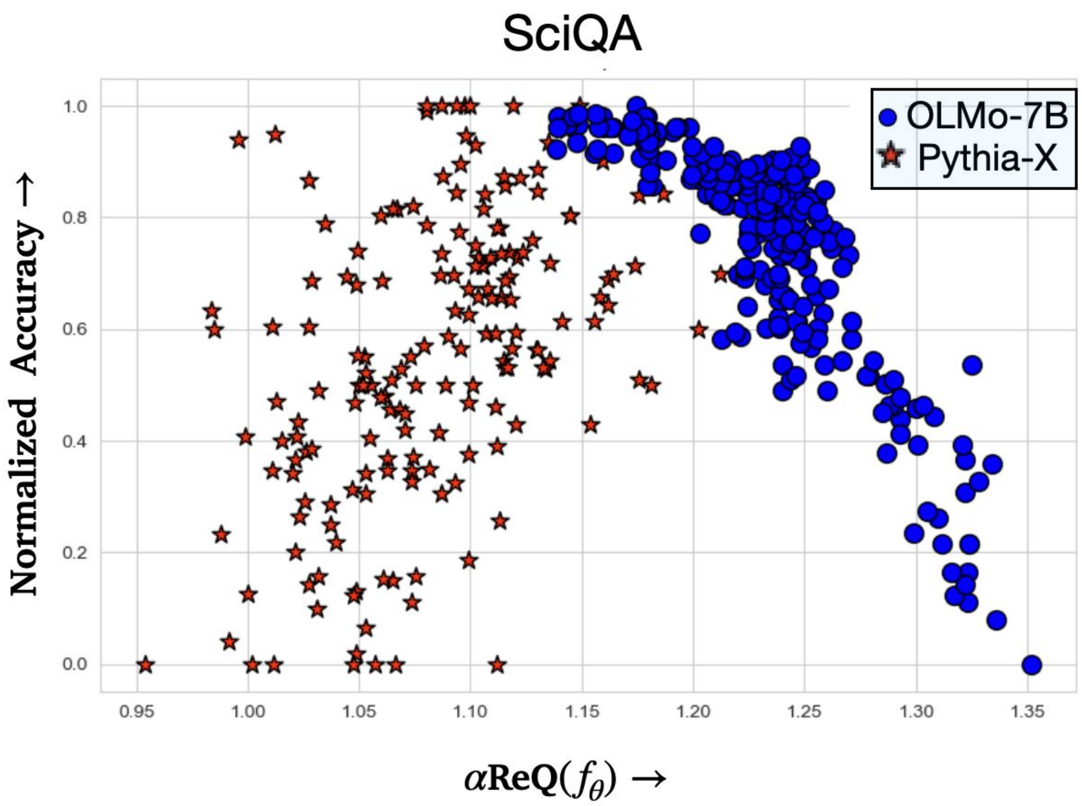



Mila

# α close to 1 predicts better performance

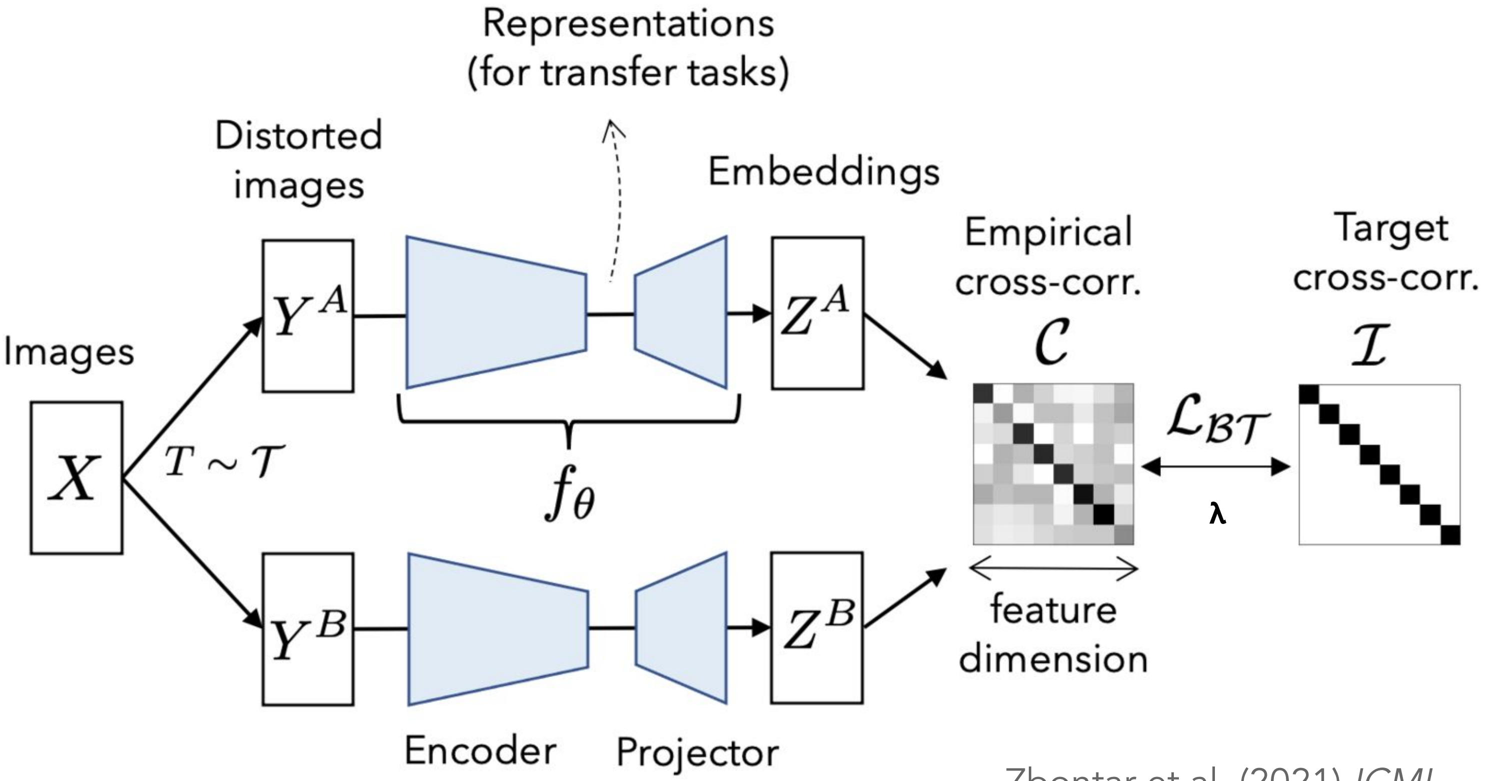


# α close to 1 predicts better performance



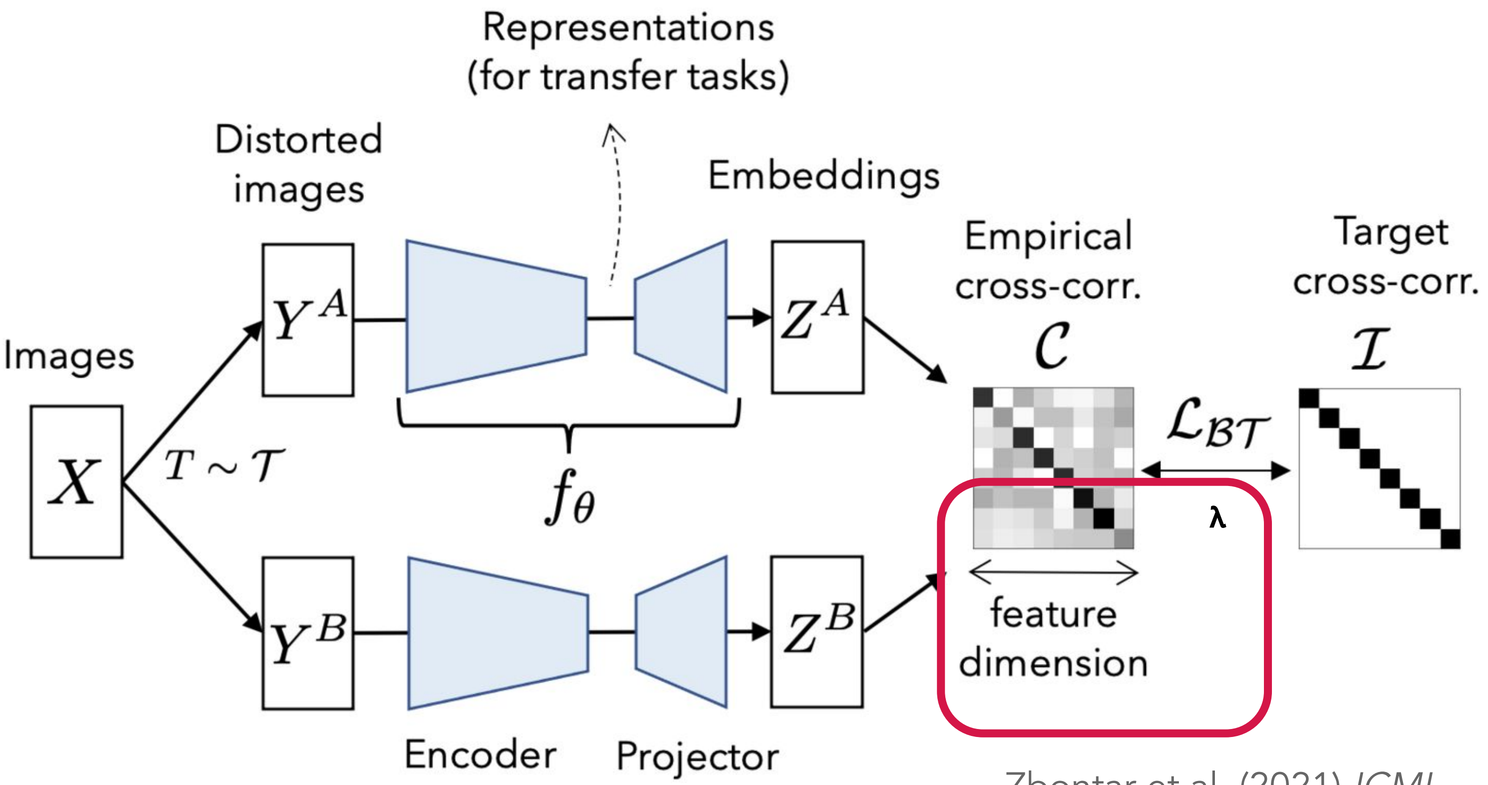

Mila

# Model selection for Barlow Twins



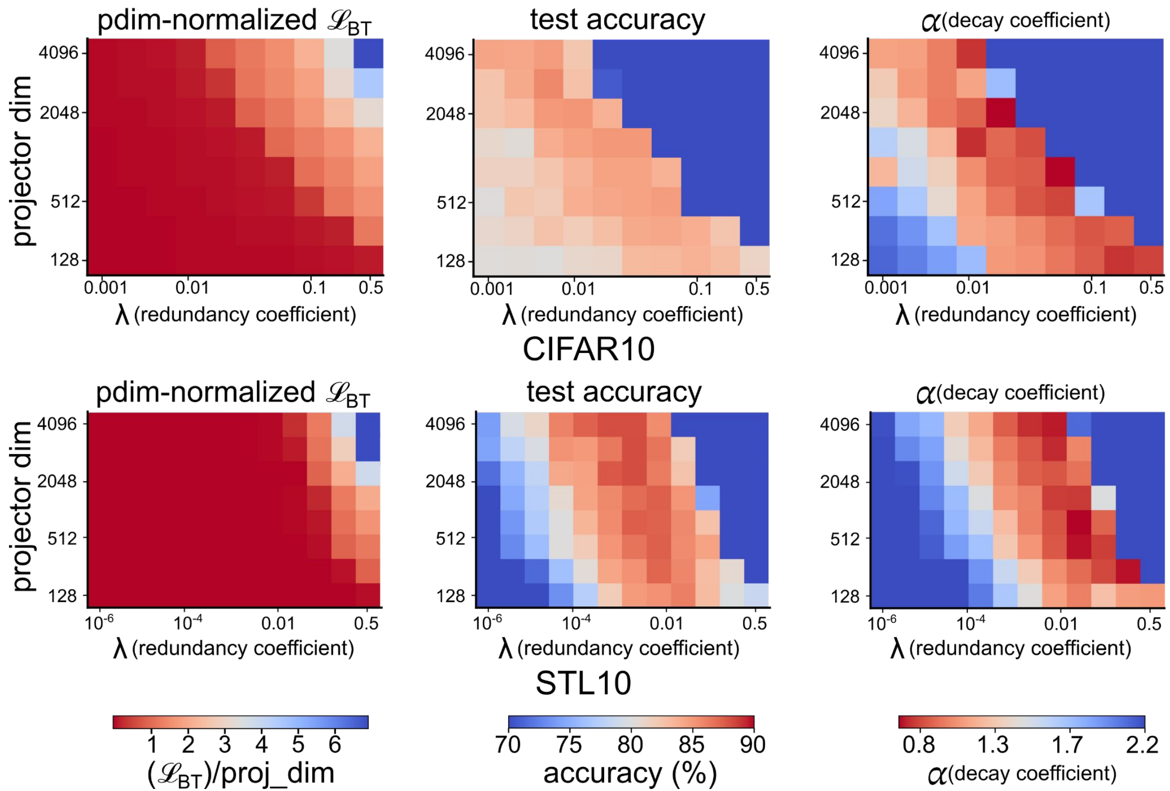

Zbontar et al. (2021) *ICML*

# 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 (α)?

# 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



Mila

# 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

Mila

# Thanks for listening!

Arna Ghosh


## Acknowledgements

Arna Ghosh
Arnab Kumar Mondal
Kumar Krishna Agrawal
Melody Li

neuro
Institut-Hôpital neurologique de Montréal
Montreal Neurological Institute-Hospital

CIFAR

McGill

NSERC CRSNG

HEALTHY BRAINS FOR HEALTHY LIVES