Do Self-Supervised and Supervised Methods Learn Similar Visual Representations?

Tom George Grigg, Dan Busbridge, Jason Ramapuram, Russ Webb
NeurIPS 2021 Workshop on Self-Supervised Learning · Apple Inc.

Abstract

Self-Supervised Learning (SSL) has been closing the performance gap with Supervised Learning (SL). How is this achieved?

Questions:

- What do SSL and SL representations have in common?
- How do SSL and SL representations differ?

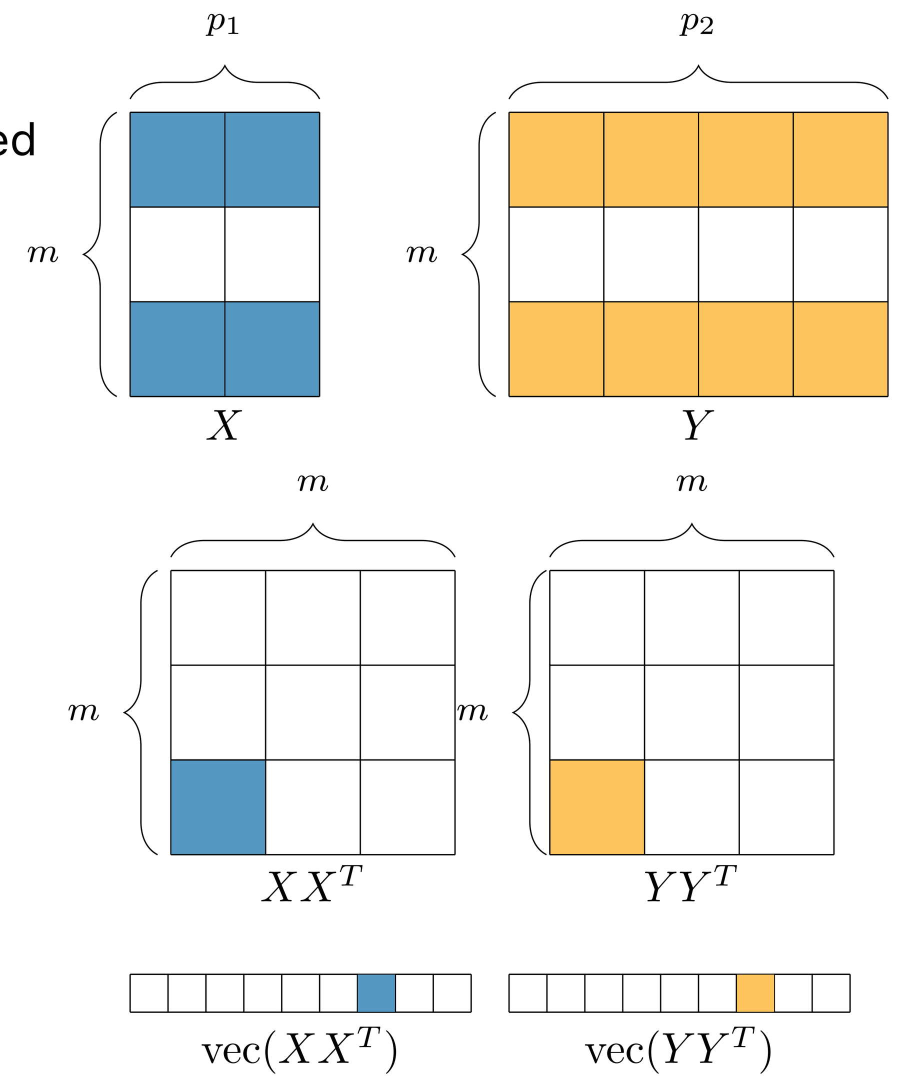
Approach:

- Produce models whose only difference is the training objective (SL or SSL).
- Study similarity of these models using a representational similarity index.

Representational Similarity with CKA

We use linear Centered Kernel Alignment (CKA) to compute the similarity of two representations:

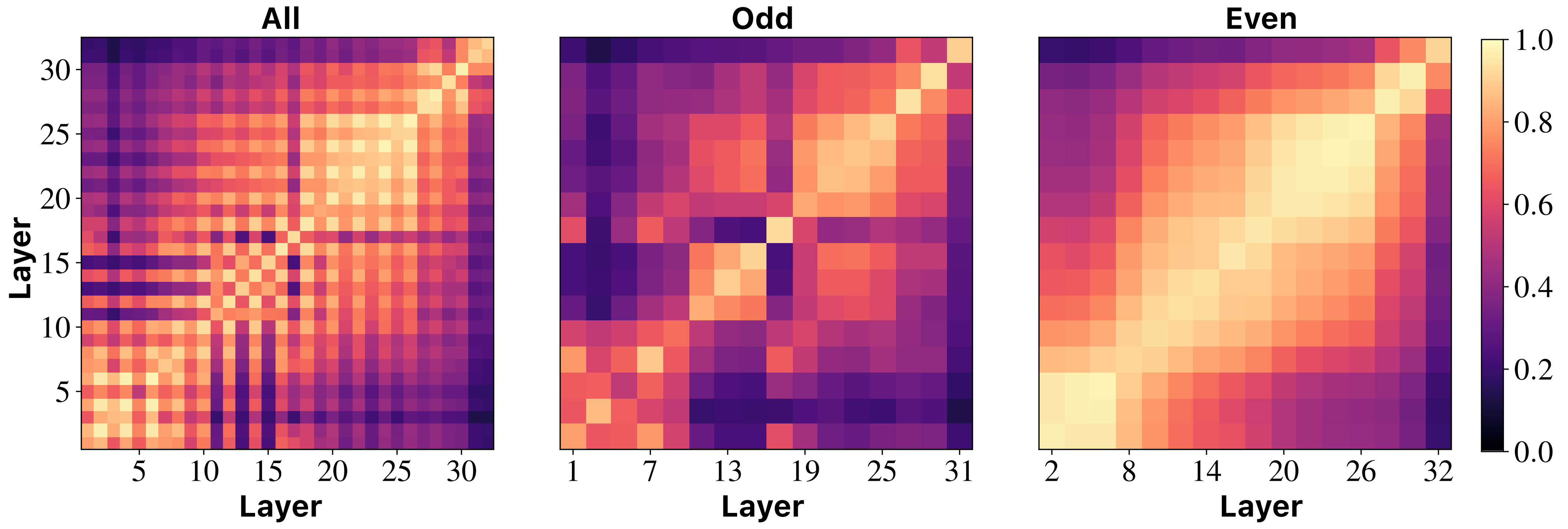
1. Compute centered features X and Y
2. Compute Gram matrices
3. Flatten
4. Compute inner product



$$\mathrm{CKA}(X,Y) \sim \mathrm{vec}(XX^T)^T \mathrm{vec}(YY^T)$$

Results: SimCLR vs SimCLR

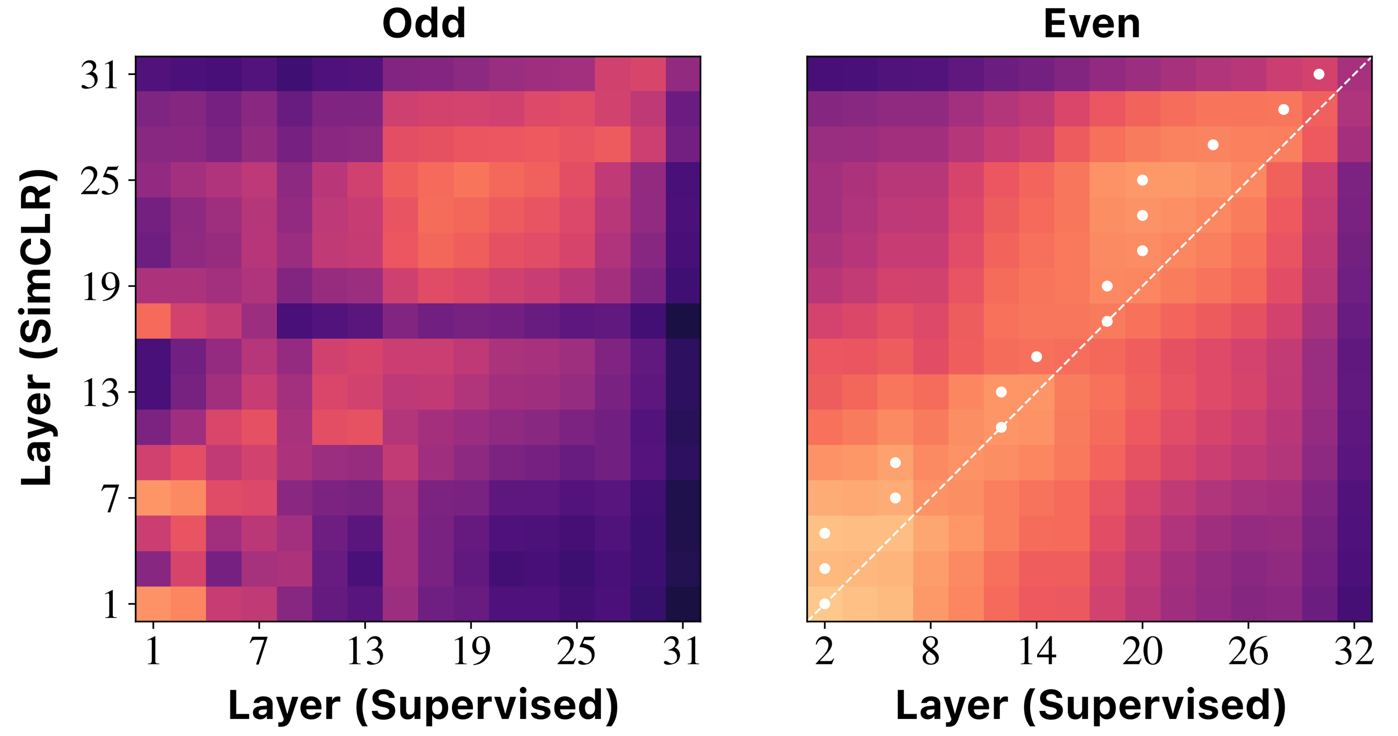
Layer-wise heatmaps of CKA values



- *Checkerboard All*: ResNet inductive bias in SSL consistent with SL.
- *Diagonal Odd + Even*: Representational similarity is stable across initializations.

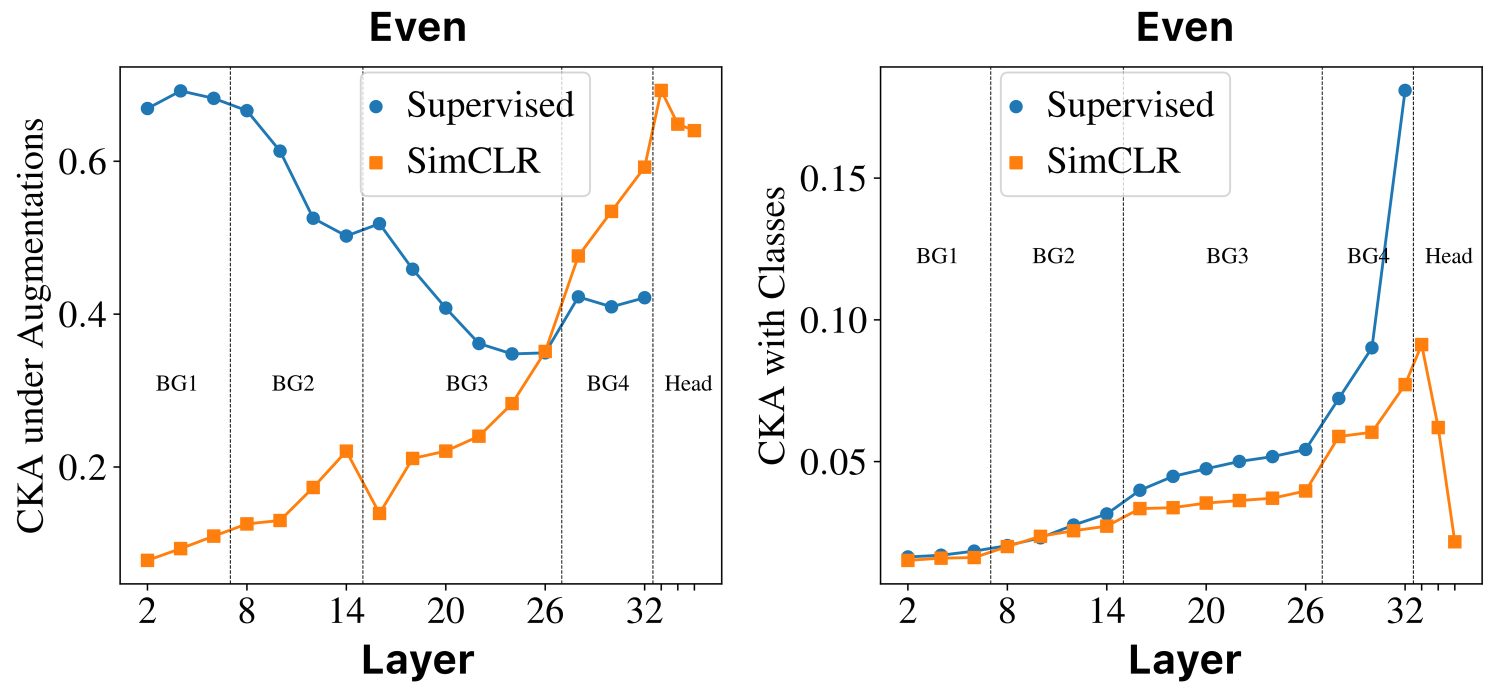
Results: SimCLR vs Supervised

Layer-wise heatmaps of CKA values



Investigating similarities and differences

- *Low Odd*: Odd representations (residuals) not shared.
- *Diagonal Even*: Dissimilar residuals gives rise to similar post-residual representations.
- *High early evens*: SSL and SL have common primitives.
- *White dots indicate row-wise maxima*: SSL "lags" behind SL.



Investigating the impact of objectives

- *Left*: SSL learns augmentation invariance, SL does not.
- *Right*: SL maps to the class simplex, SSL does not.
- BG = Bottleneck Group.

Experimental Details

General setting

- ResNet50s used for SL + SSL backbone.
- SimCLR used as representative SSL method.
- 3 seeds per model.
- All models trained from scratch on CIFAR10 train (50K samples).
- Representation evaluated on CIFAR10 test (10K samples).

Odd and Even representations

Analyzed representations are extracted from the ResNet50 bottleneck layers or after final max pool.

Odd representations are residual and even representations are post-residual.

$$\mathrm{even}(x) = x + \mathrm{odd}(x)$$

Conclusion

- CKA can be used across learning objectives.
- CKA revealed similarities as well as differences between the SSL and supervised representations.
- Representational differences can be understood in terms of the distinct learning objectives.
- Going forward: identifying aspects of supervised learning to infuse into SSL can aid the design of new learning objectives.