Little Label Learners

Winston Li, Marcel Mateos Salles, John Ryan Byers

May 2024

BROWN
Computer Science

Outline

- Group
- Introduction
- Methodology
- Results
- Discussion

Our Group

The Little Label Learners Group

Winston Li

- wli115
- winston_y_li@brown.edu

Marcel Mateos Salles

- mmateoss
- marcel_mateos_salles@brown.edu

John Ryan Byers

- jbyers3
- john_byers@brown.edu

Introduction

Definitions

Fully-Supervised

- All data inputs have labels

Self-Supervision

- No labels associated with inputs

Semi-Supervised

- ***Some*** data inputs are labeled

Inspiration

- Children must learn to identify all types of objects
    - They tend to learn differences between animals, cars, and other everyday things even if they don’t learn the actual name
- Standard image classification models like ResNet can classify images accurately, but use an enormous amount of data
- Could we design a model using both unlabeled and labeled data to create a classifier more similar to a biological learning process?

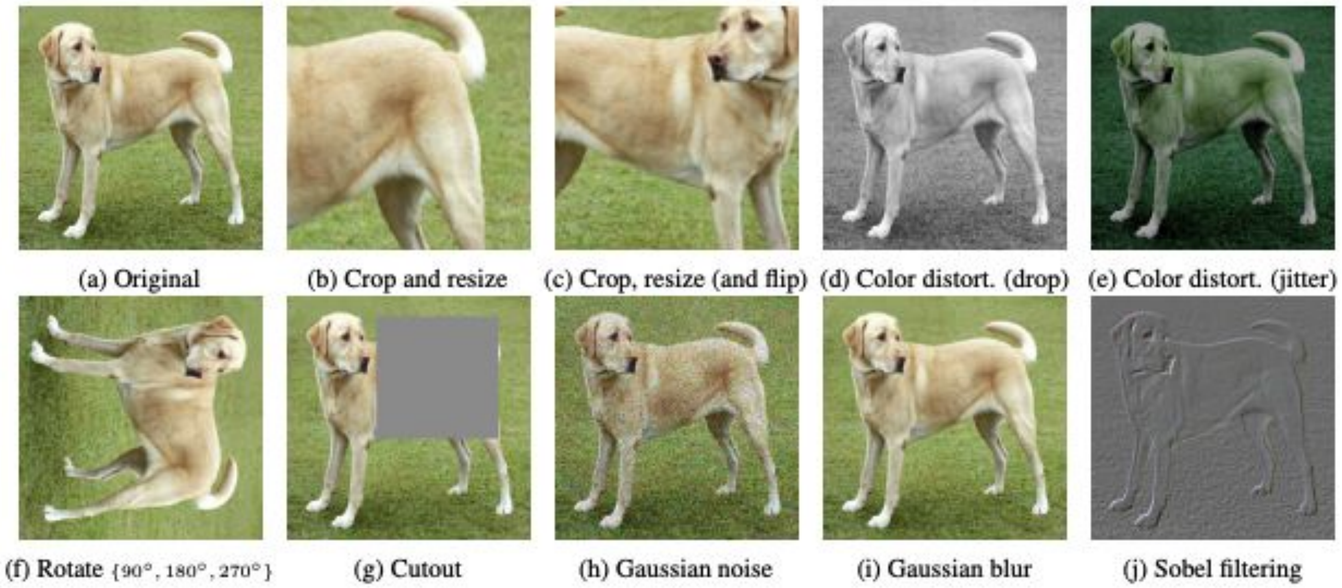
https://www.pnmag.com/gear/three-steps-to-a-cleaner-nursery

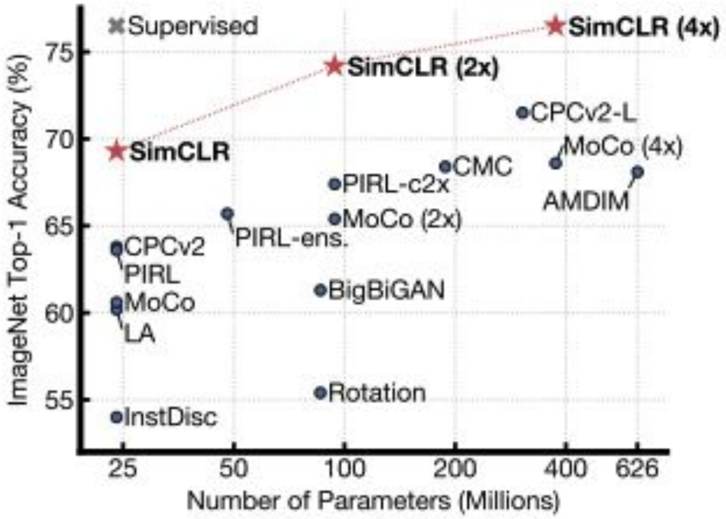
BROWN

Related Work

A Simple Framework for Contrastive Learning of Visual Representations (SimCLR)

- SimCLR uses self-supervision to achieve higher accuracy than traditional fully-supervised models
- Contrastive Learning comes from augmentations to data
    - Augmented images are passed in pairs to generate a latent space embedding



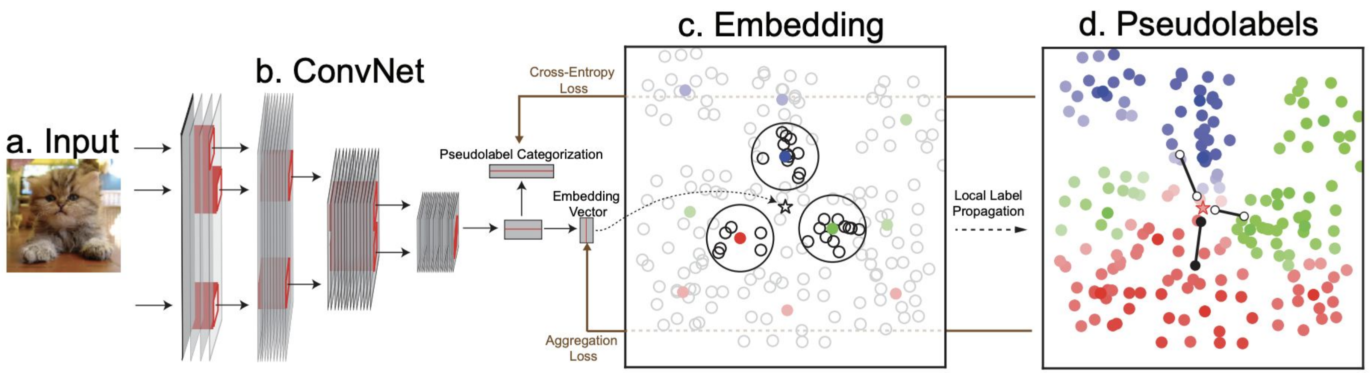


https://arxiv.org/pdf/2002.05709.pdf

BROWN

Local Label Propagation for Large-Scale Semi-Supervised Learning

- Zhuang et al. (2019)
- Representation Learning: update encoder by minimizing CCE between predicted and propagated pseudolabels and maximizing global aggregation
- Label Propagation: weighted KNN from top K nearest *labeled* examples for each *unlabeled* sample



https://arxiv.org/pdf/1905.11581.pdf

BROWN

Methodology

Dataset

CIFAR10

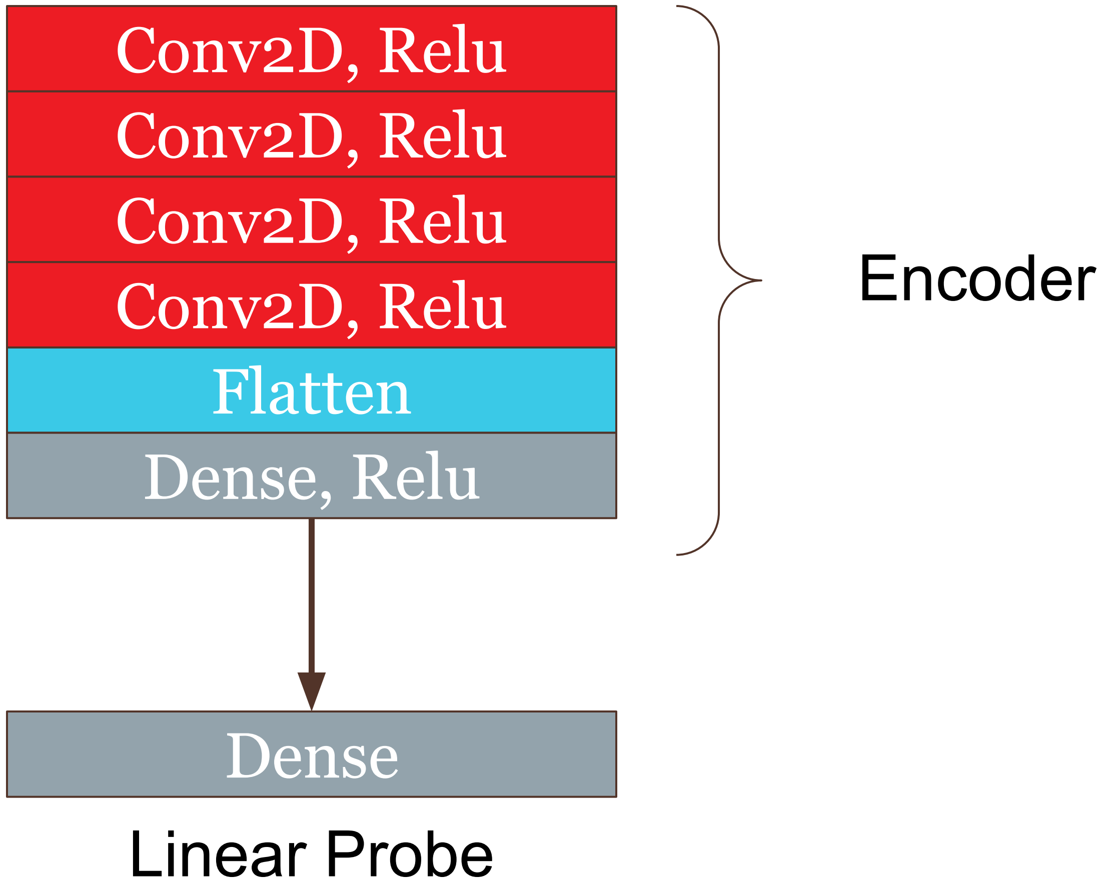
- 50,000 training images and 10,000 testing images
- Labeled into 10 Categories
- 32x32 and Colored
    - Allows quick preprocessing and training
- Easily accessible through keras load function

https://keras.io/api/datasets/cifar10/
https://www.cs.toronto.edu/~kriz/cifar.html

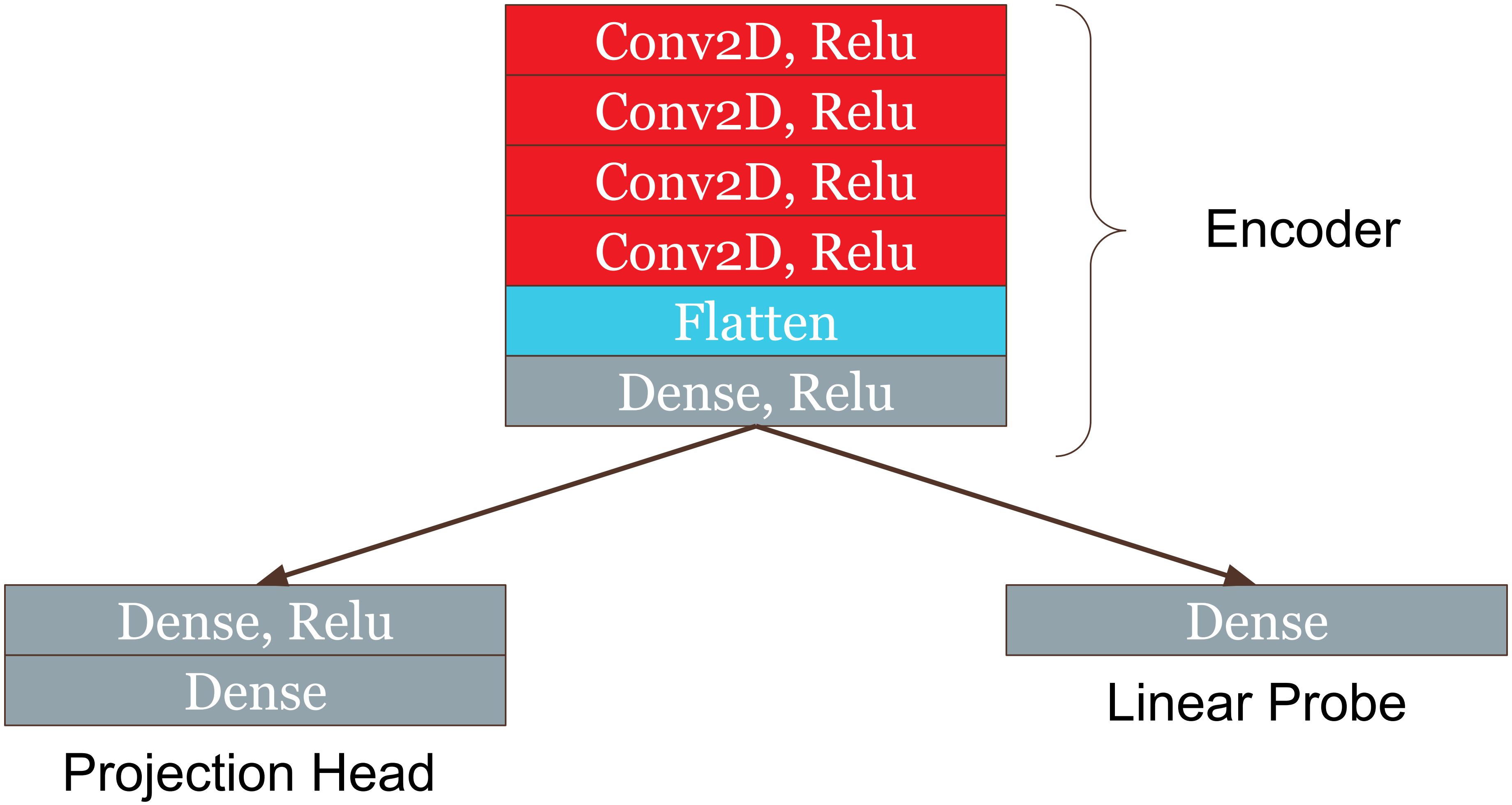
Dataloader (and Preprocessing)

- We needed to modify the CIFAR10 dataset to have both labeled and unlabeled data for our training pipeline
- We batched labeled and unlabeled data together using a custom DataLoader class using Python OOP
- Allows us to decide percentage of labels in the data and also the number of labels present in the data being used

Baseline Model



Self- & Gradual-Supervised Model



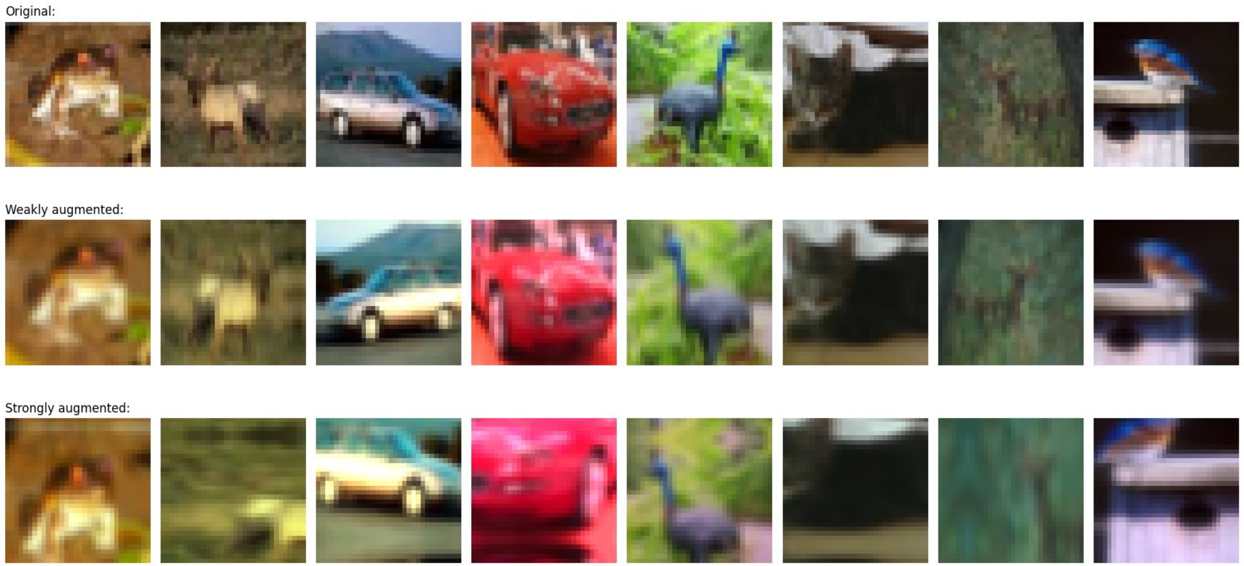
Contrastive Loss

- Takes two augmentations of the same image
- Loss function rewards when these two embedded, same images are close together in the embedding space

Original

Weakly Augmented

Strongly Augmented



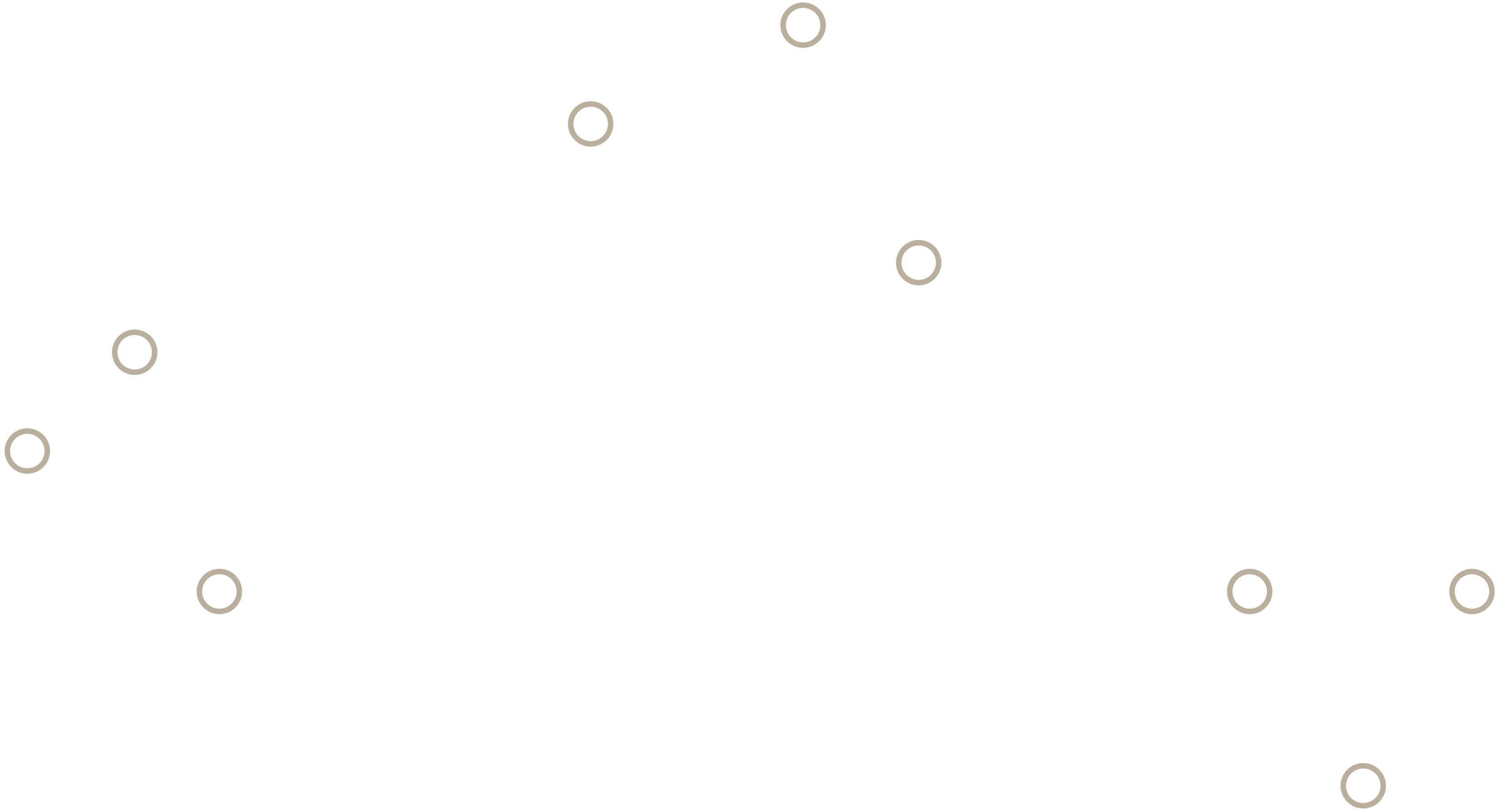
BROWN

Contrastive Loss - NT-Xent/N-pairs

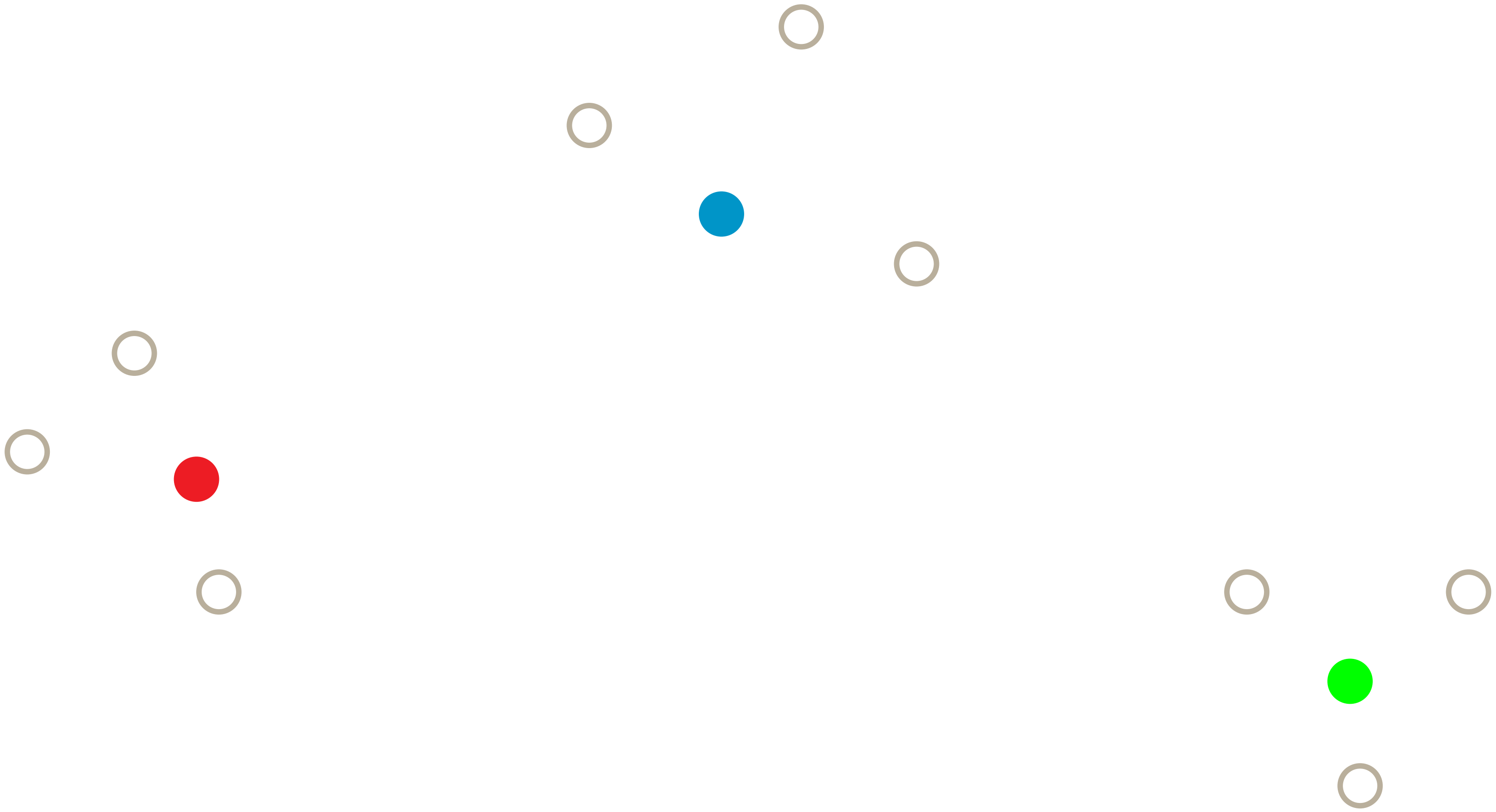
$$\mathcal{L}_{NT-X_{ent}} = -\frac{1}{N}\sum_{i=1}^{N} \frac{\exp(\mathrm{sim}(z_i, z_j)/\tau)}{\sum_{k=1,(i\neq k)}^{2N} \exp(\mathrm{sim}(z_i, z_k)/\tau)}$$

$$\mathrm{sim}(z_i, z_j) = \frac{z_i \cdot z_j}{||z_i|| \cdot ||z_j||}$$

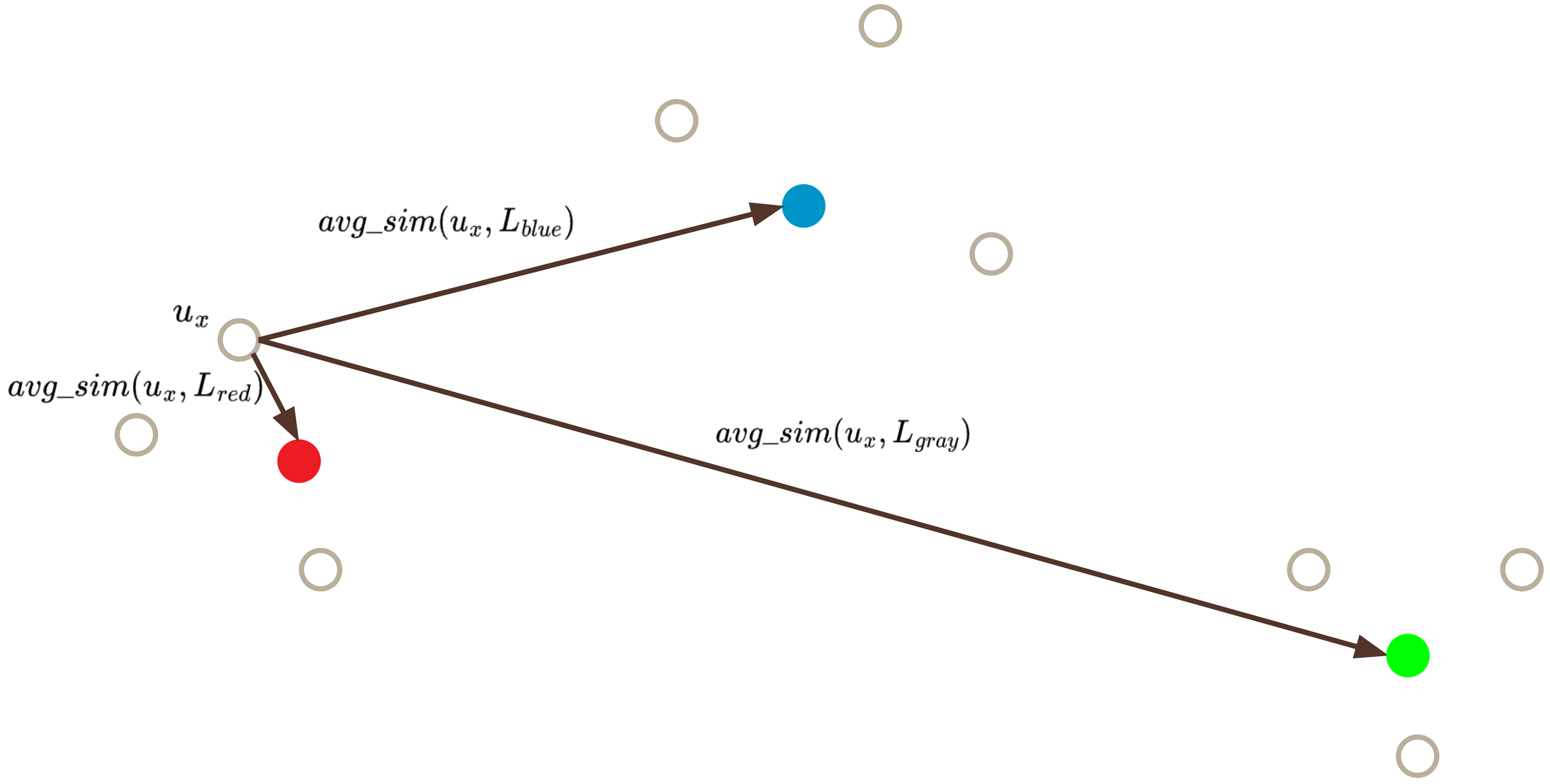
Pseudo-Classifier



Pseudo-Classifier

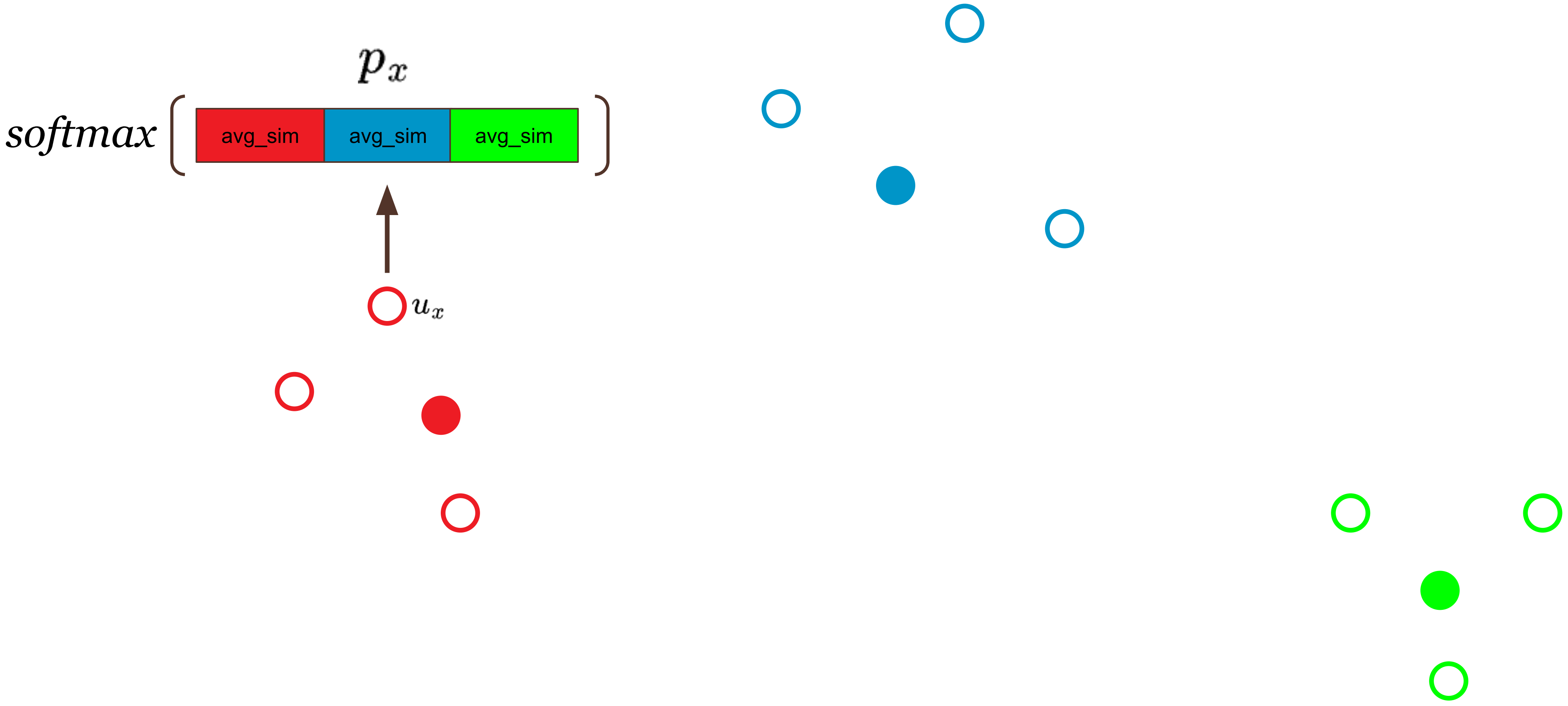


Pseudo-Classifier



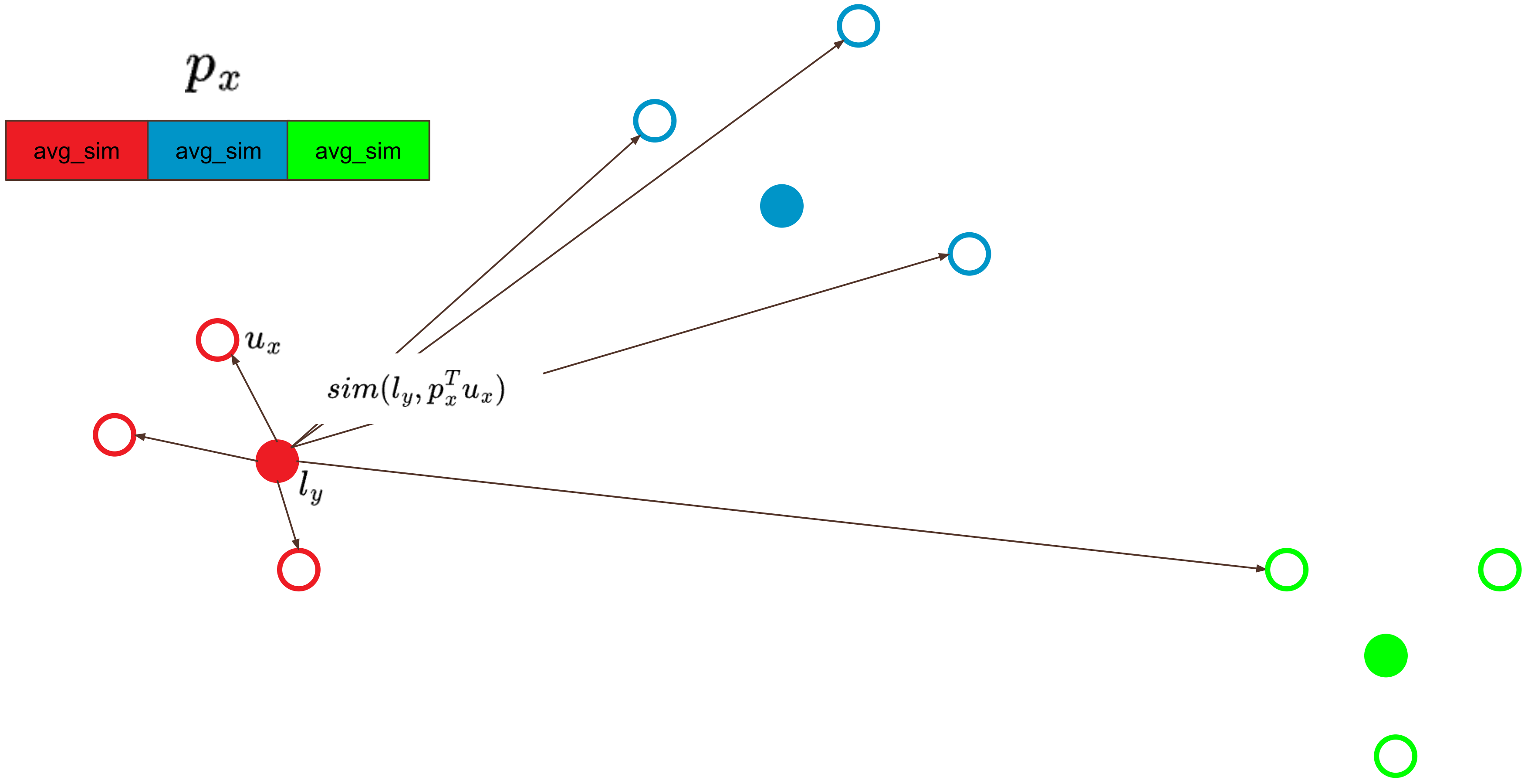
BROWN

Pseudo-Classifier

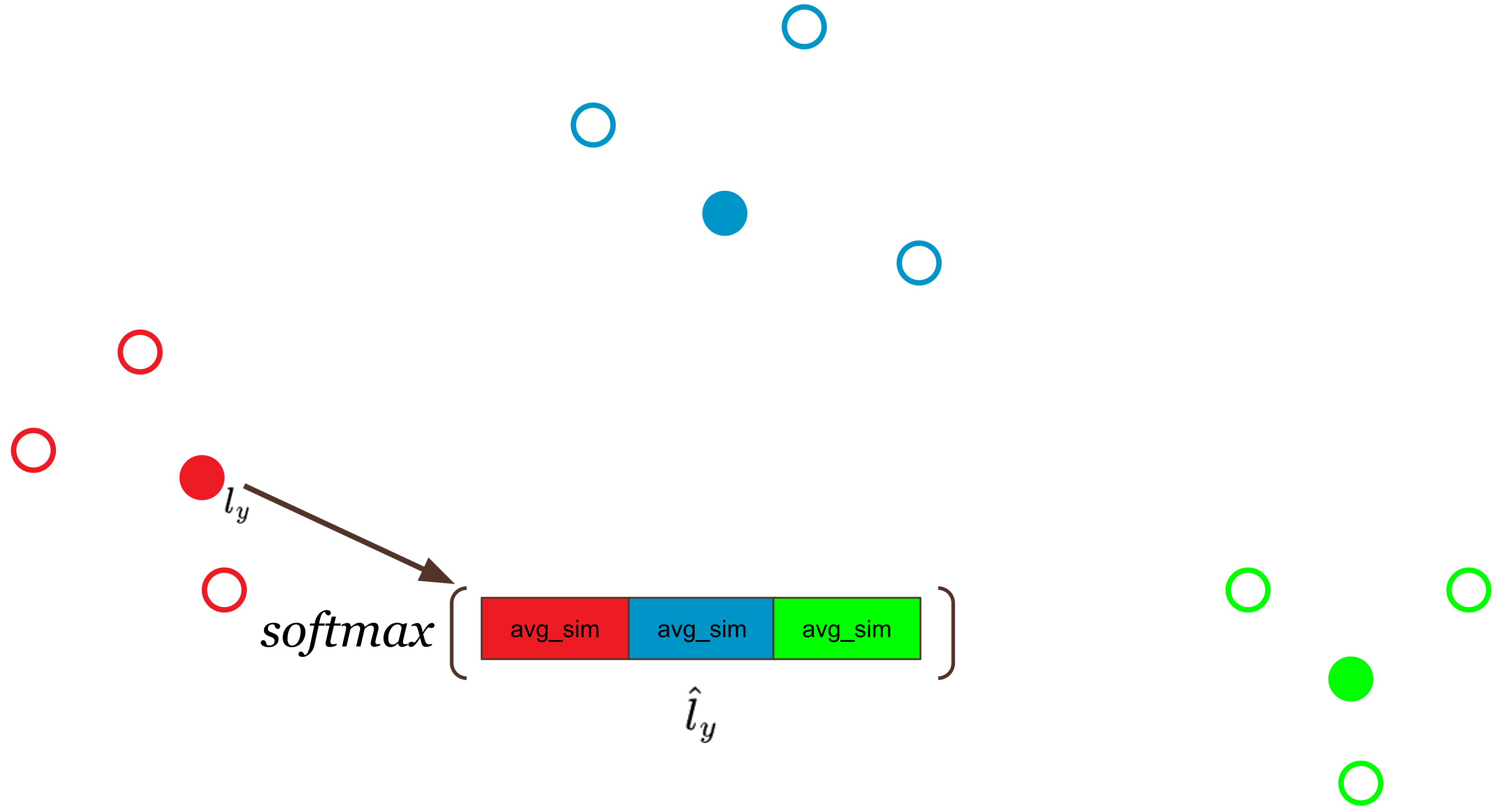


BROWN

Pseudo-Classifier

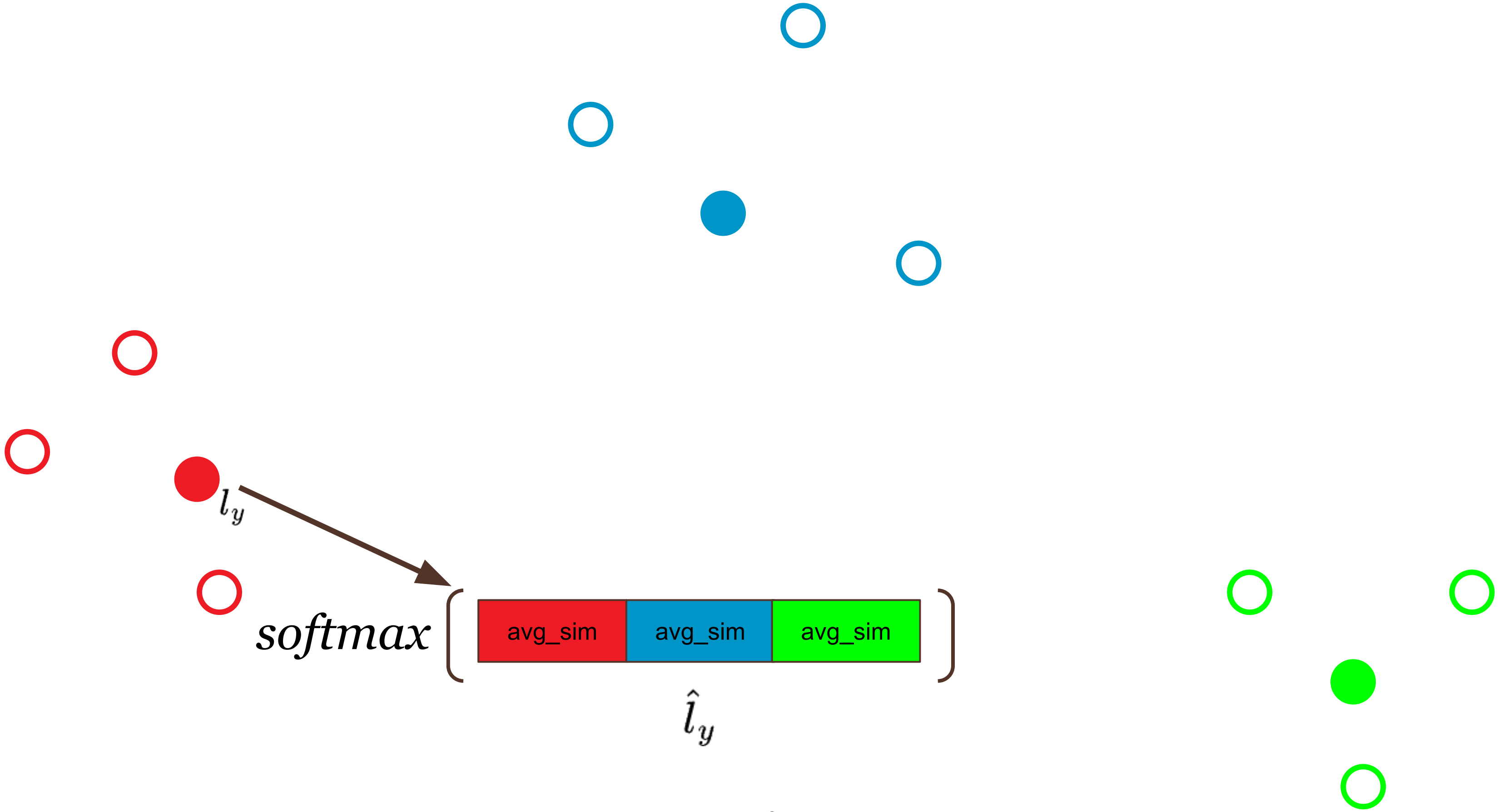


Pseudo-Classifier



BROWN

Pseudo-Classifier



$$CCE(l_y, \hat{l}_y)$$

Epoch-Scheduled Callback

- Epoch-based callback
- Class Representation
  - Determines how many classes are represented in datasets. Gradually work with more labels over epochs to represent continual learning

$$\alpha = epoch$$

- Labeled/Unlabeled Split Rate
  - Determines how much data to use as labeled vs. unlabeled samples. Gradually use more labeled samples over epochs

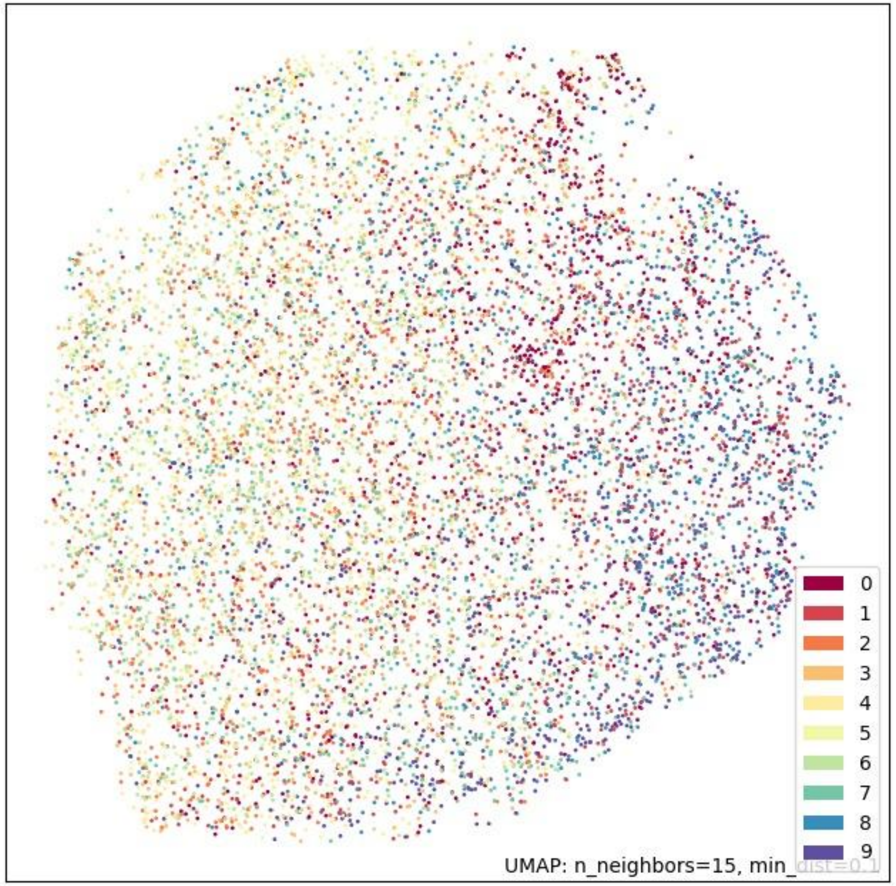
$$\beta = \frac{epoch}{total\ \ epochs}$$

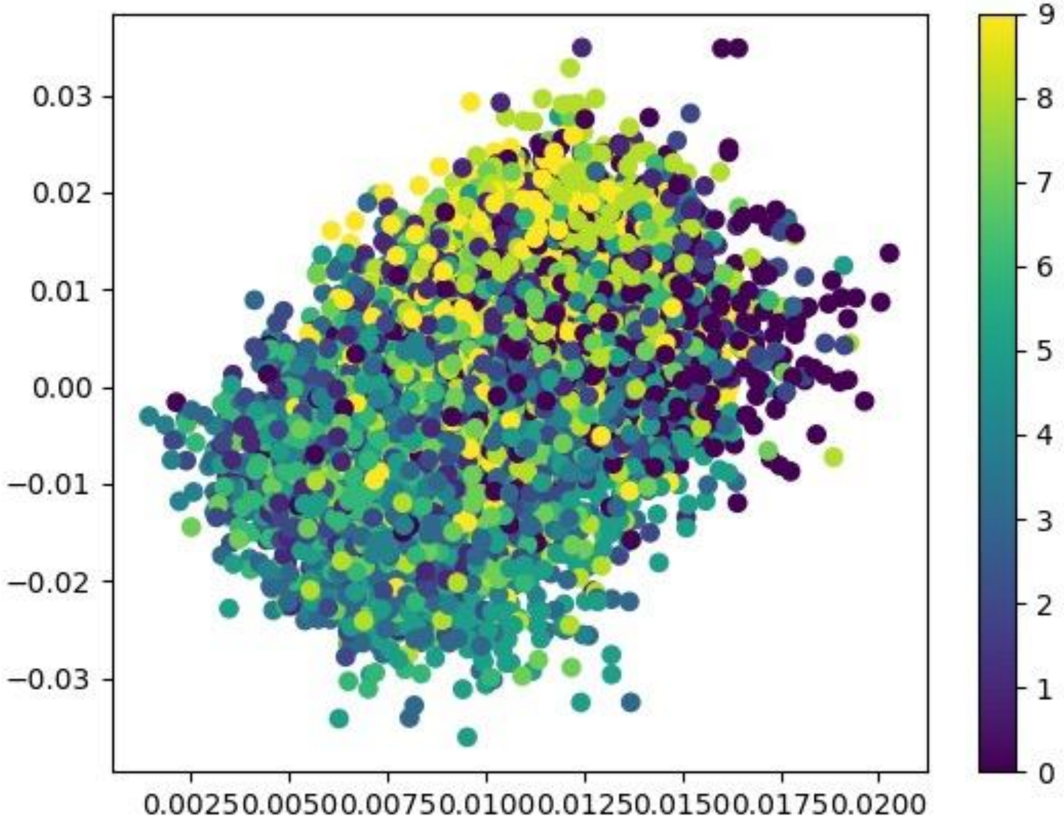
- Contrastive/Probe Learning Rate
  - Encoder is more “fundamental”: hard to re-train at late epochs

$$\eta = \eta_{ceil} \cdot 0.1^{epoch/500}$$

Results

Embedding Spaces





BROWN

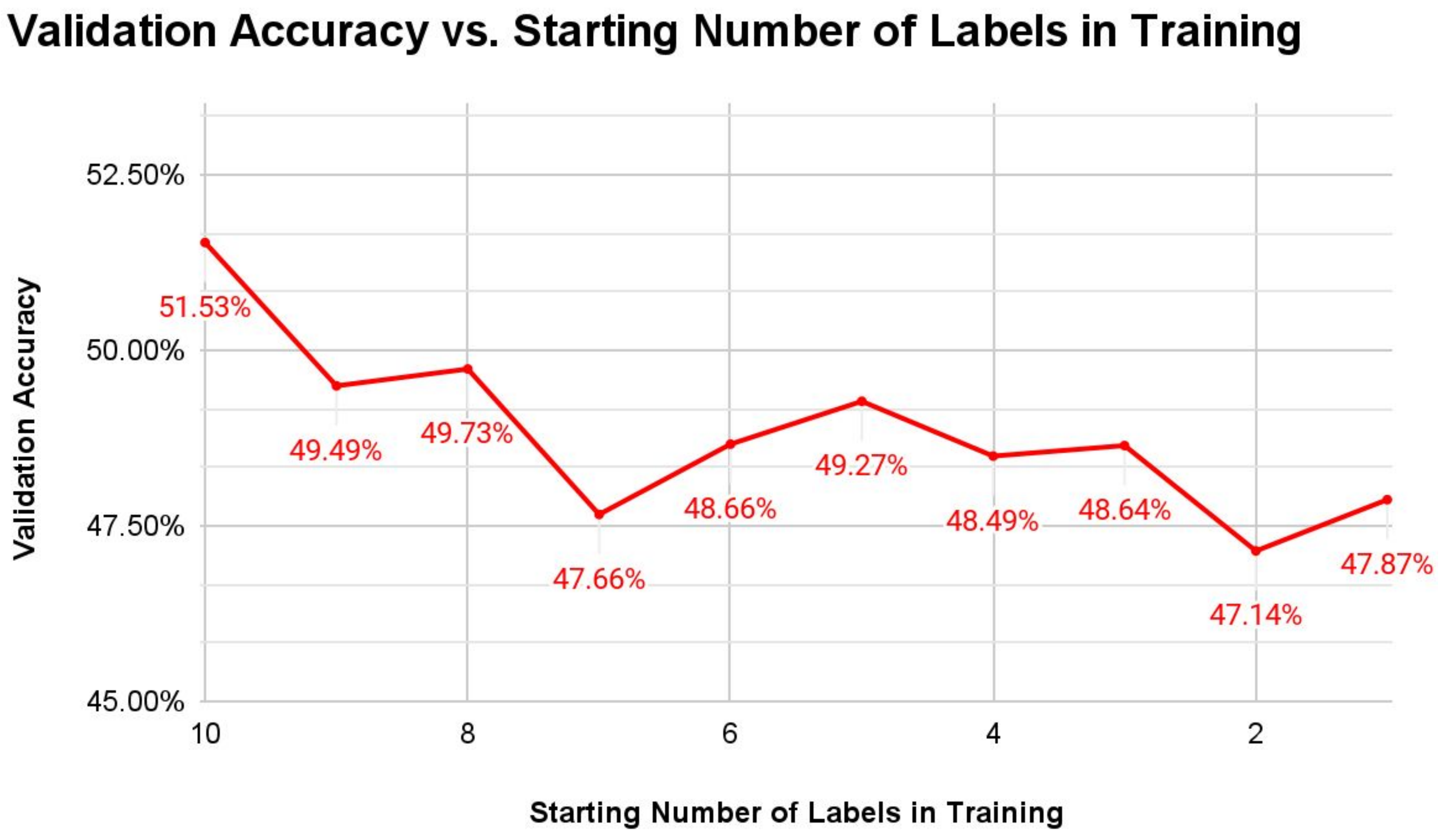
Quantitative (20 epochs)

Model	Learning Type	Dataset	Train/Test Data	Max Validation (%)
Baseline	Fully Supervised	CIFAR10	Full train, full test	59.18
Baseline	Fully Supervised	CIFAR10	train 7 classes; full test	43.11
Baseline	Fully Supervised	CIFAR10	train 7 classes; test 7 classes	59.59
Baseline	Fully Supervised	CIFAR10	Train full; test 7	54.97
SimCLR	Self Supervised	CIFAR10	Full train, full test	31.27
SimCLR	Self Supervised	CIFAR10	train 7 classes; full test	25.30
SimCLR	Self Supervised	CIFAR10	train 7 classes; test 7 classes	37.61
SimCLR	Self Supervised	CIFAR10	Train full; test 7	32.00

Quantitative (20 epochs)

Model	Learning Type	Dataset	Train/Test Data	Max Validation (%)
Baseline	Fully Supervised	CIFAR10	Full train, full test	59.18
SimCLR	Self Supervised	CIFAR10	Full train, full test	31.27
Novel Model	Gradual Supervision	CIFAR10	Start training **with 10**, finish training with 10	51.53
Novel Model	Gradual Supervision	CIFAR10	Start training with **9 to 10,** finish training with 10	49.49
Novel Model	Gradual Supervision	CIFAR10	Start training with **8 to 10**, finish training with 10	49.73
Novel Model	Gradual Supervision	CIFAR10	Start training with **7 to 10**, finish training with 10	47.66
Novel Model	Gradual Supervision	CIFAR10	Start training with **6 to 10**, finish training with 10	48.66
Novel Model	Gradual Supervision	CIFAR10	Start training with **5 to 10**, finish training with 10	49.27
Novel Model	Gradual Supervision	CIFAR10	Start training with **4 to 10**, finish training with 10	48.49
Novel Model	Gradual Supervision	CIFAR10	Start training with **3 to 10**, finish training with 10	48.64
Novel Model	Gradual Supervision	CIFAR10	Start training with **2 to 10**, finish training with 10	47.14
Novel Model	Gradual Supervision	CIFAR10	Start training with **1 to 10**, finish training with 10	47.87

Quantitative



- Gradually Supervised Model Training Regime:
  - Gradually Introducing more labels for the model to train with as epochs progress
  - Gradually increase the percentage of labeled data used in training (1% to 90%)
  - Gradually change the training rates for contrastive learning and linear probe.

BROWN

Major Findings

- Compared to purely self-supervised SimCLR, our model had better accuracies; we also learned with less labeled samples
- Our novel gradual learning approach allowed the model to learn new classes across epochs
- Low learning rate for the encoder was especially helpful; led to ~5% better probe classification
- Pseudo-classification led to better contrastive performance, but did not significantly impact classification

Discussion

Lessons Learned

- When you begin with less categories and gradually add them in, there is a lower accuracy then starting with all categories
- Contrastive learning approaches in self and semi supervised learning
- How to use TensorFlow on the Brown Supercomputer OSCAR

Lingering Problems/Limitations

- We would have liked to construct additional loss functions
- Our DataLoader code is not that easily compatible for future datasets such as MiniImageNet

Future Work

- We would have liked to run our model on an additional dataset titled MiniImageNet
    - These higher resolution images might lead to better embeddings and outperform the baseline
- Better label-informed encoder updates

https://www.kaggle.com/datasets/ctrnngtrung/miniimagenet

BROWN

Thank You

BROWN
Computer Science

Sources

Chen, Ting, et al. "A Simple Framework for Contrastive Learning of Visual Representations." *arXiv.Org*, 2020, https://doi.org/10.48550/arxiv.2002.05709.

Chollet, Francois et al. Keras. https://keras.io, 2015.

Krizhevsky, Alex, et al. "CIFAR-10 (Canadian Institute for Advanced Research)." 2009, http://www.cs.toronto.edu/~kriz/cifar.html.

Zhuang, Chengxu, et al. "Local Aggregation for Unsupervised Learning of Visual Embeddings." *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019, pp. 6001–11, https://doi.org/10.1109/ICCV.2019.00610.

Zhuang, Chengxu, et al. "Local Label Propagation for Large-Scale Semi-Supervised Learning." *arXiv.Org*, 2019, https://doi.org/10.48550/arxiv.1905.11581.

BROWN

Bonus Slides

Ethics

Why is Deep Learning a good approach to this problem?

- We are attempting to replicate a biological pipeline
- This biological pipeline is best represented with neurons which we can represent with our architecture

What broader societal issues are relevant to your chosen problem space?

- Our architecture which requires less data than a traditional classifier is more accessible to the average developer
- It does not require advanced GPUs

BROWN