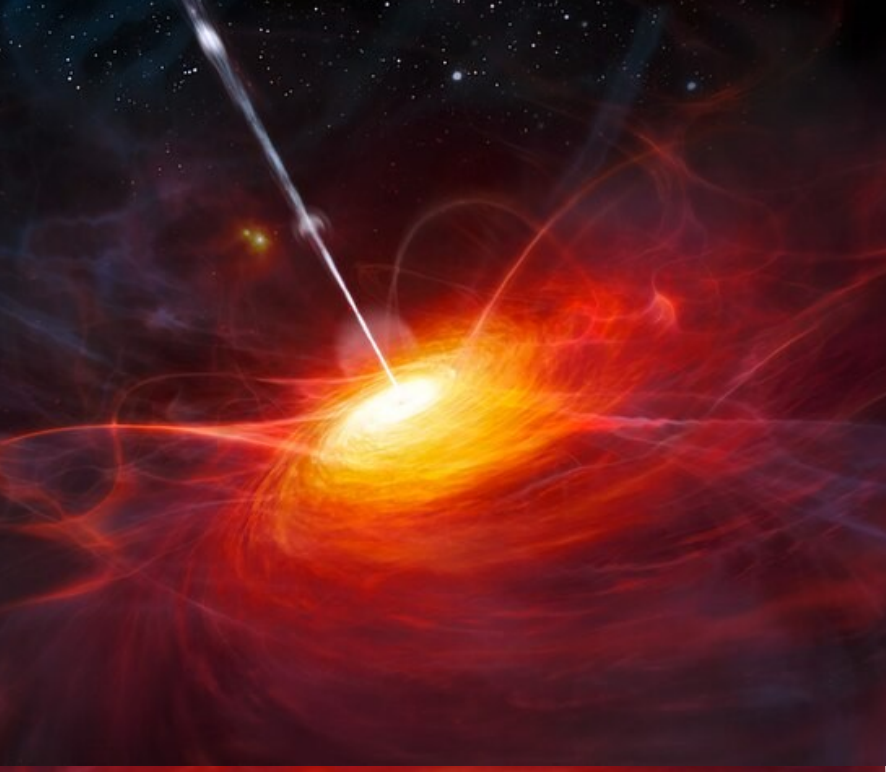


Contrastive Learning
and High-Redshift Quasars

ajnb3@cam --- Xander Byrne --- xbyrne.github.io

1st year PhD Student --- Institute of Astronomy

Outline

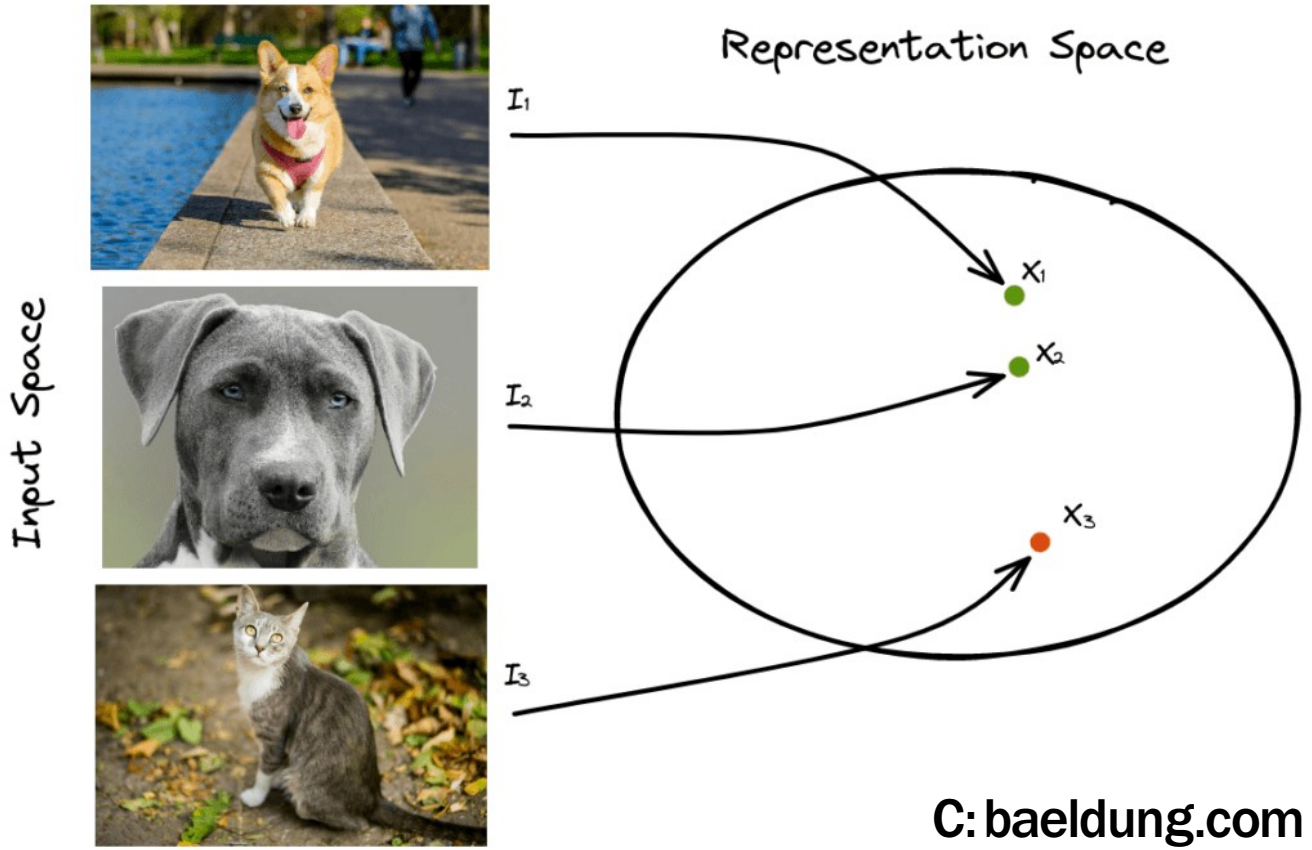
- What is contrastive learning?
- A brief example: high-redshift quasars
- How does contrastive learning work?
- Deep dive into the contrastive framework
- How good is contrastive learning?

What is Contrastive Learning?

- An unsupervised ML method for clustering (Chen+20)
- Datasets (e.g. of images) --> clustered latent vectors

A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen [1] Simon Kornblith [1] Mohammad Norouzi [1] Geoffrey Hinton [1]



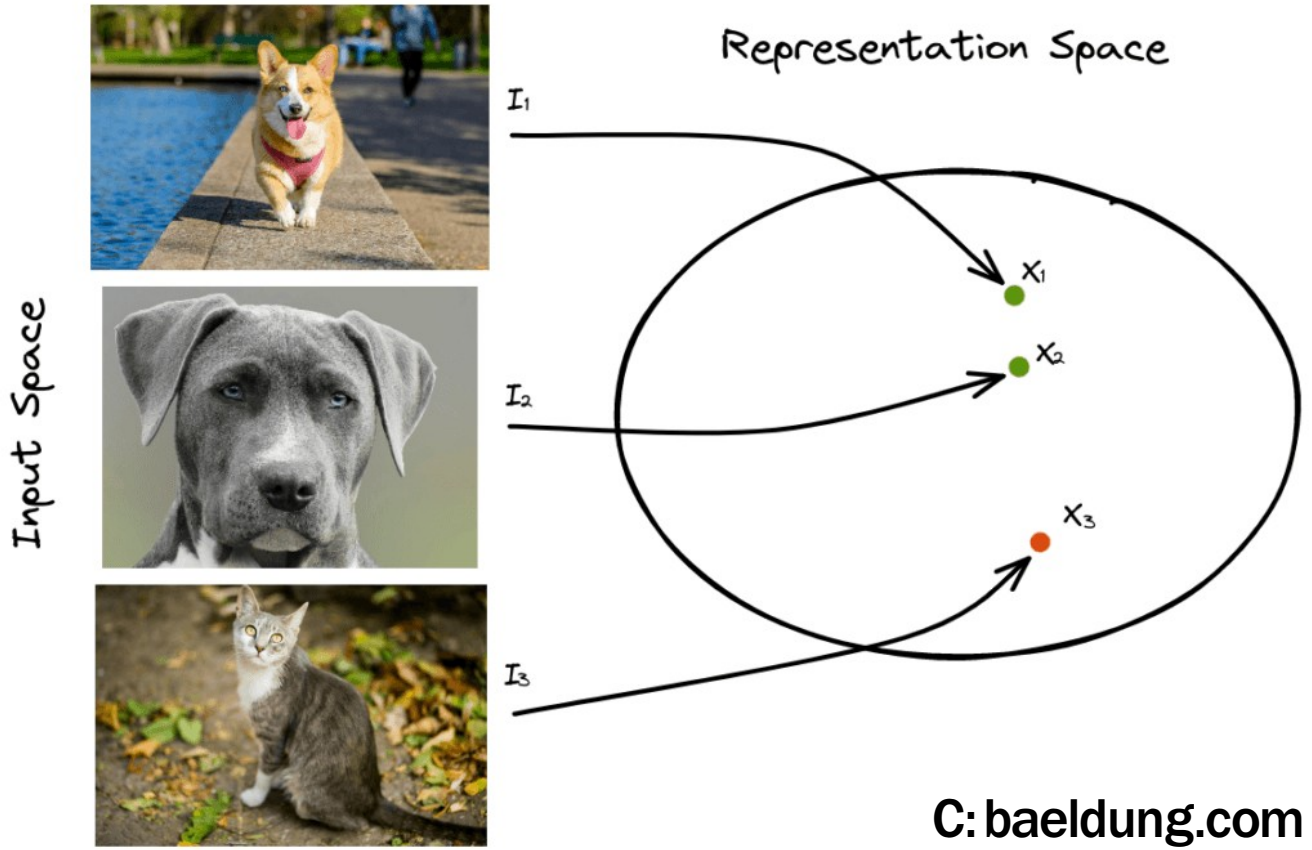
C: baeldung.com

What is Contrastive Learning?

- An unsupervised ML method for clustering (Chen+20)
- Datasets (e.g. of images) --> clustered latent vectors
- No labels, no problem: two transformations of the same image must have the same label

A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen [1] Simon Kornblith [1] Mohammad Norouzi [1] Geoffrey Hinton [1]



C: baeldung.com

A brief example: high-z quasars (HZQ)

A brief example: high-z quasars (HZQ)

- HZQs identified from large-sky surveys (e.g. DES, PanSTARRS) by their photometry

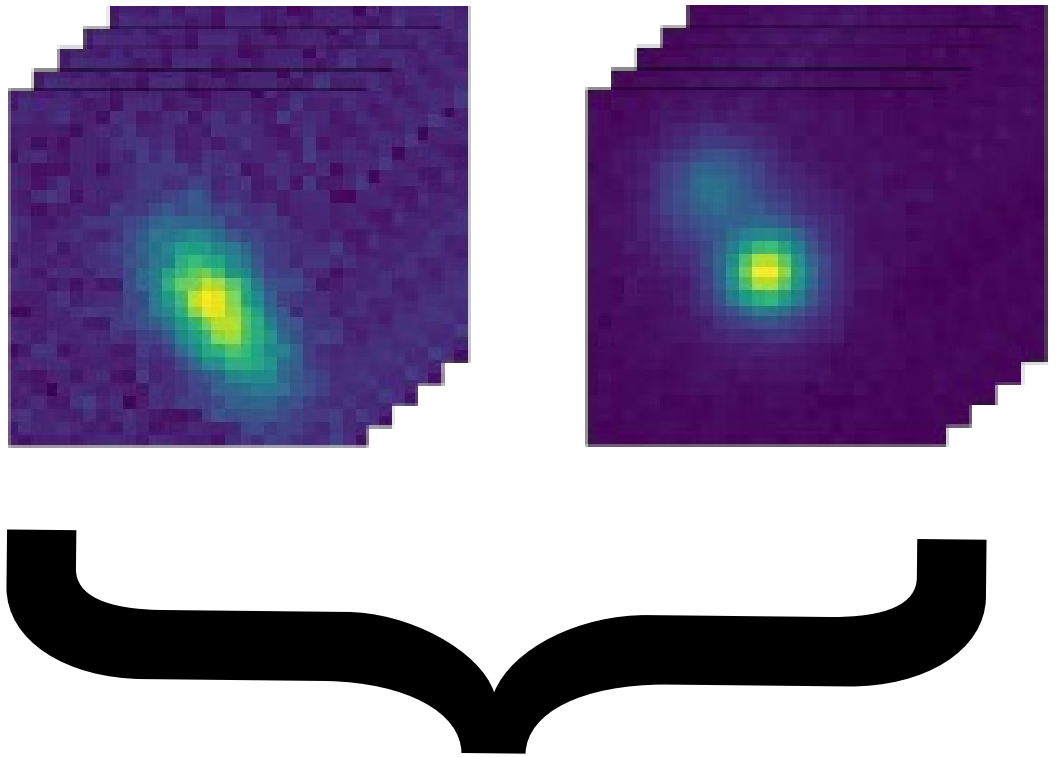
A brief example: high-z quasars (HZQ)

- HZQs identified from large-sky surveys (e.g. DES, PanSTARRS) by their photometry

- But some other sources have HZQ-like colours (BDs, z˜1 gal.s) => Hard to distinguish

A brief example: high-z quasars (HZQ)

- HZQs identified from large-sky surveys (e.g. DES, PanSTARRS) by their photometry
- But some other sources have HZQ-like colours (BDs, z~1 gal.s) => Hard to distinguish
- Could imaging help?



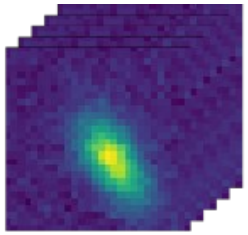
×7,000

Byrne+24, in prep

Byrne+24, in prep

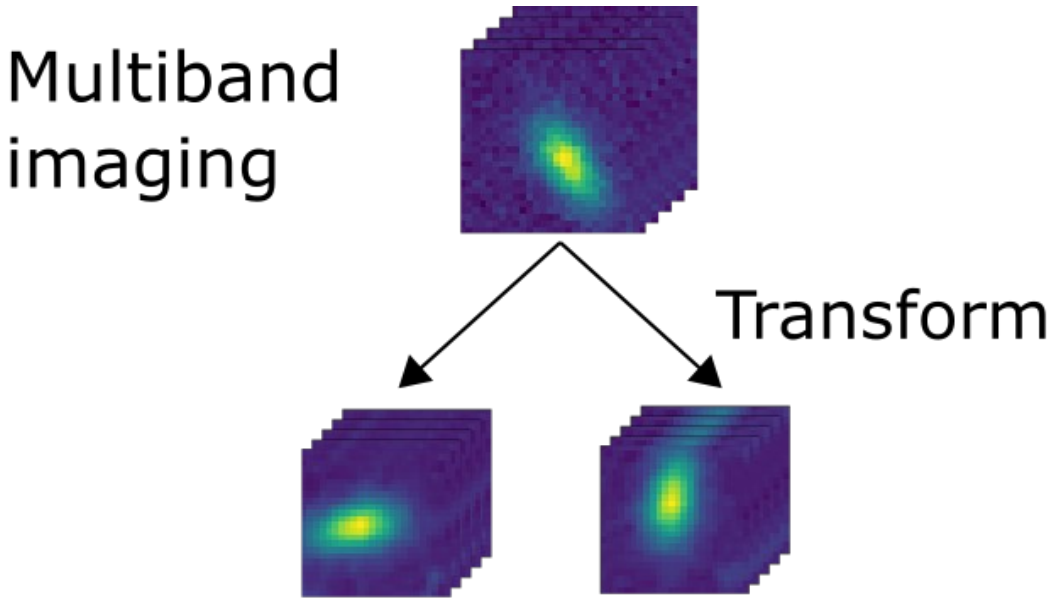
How does Contrastive Learning work?

Multiband imaging



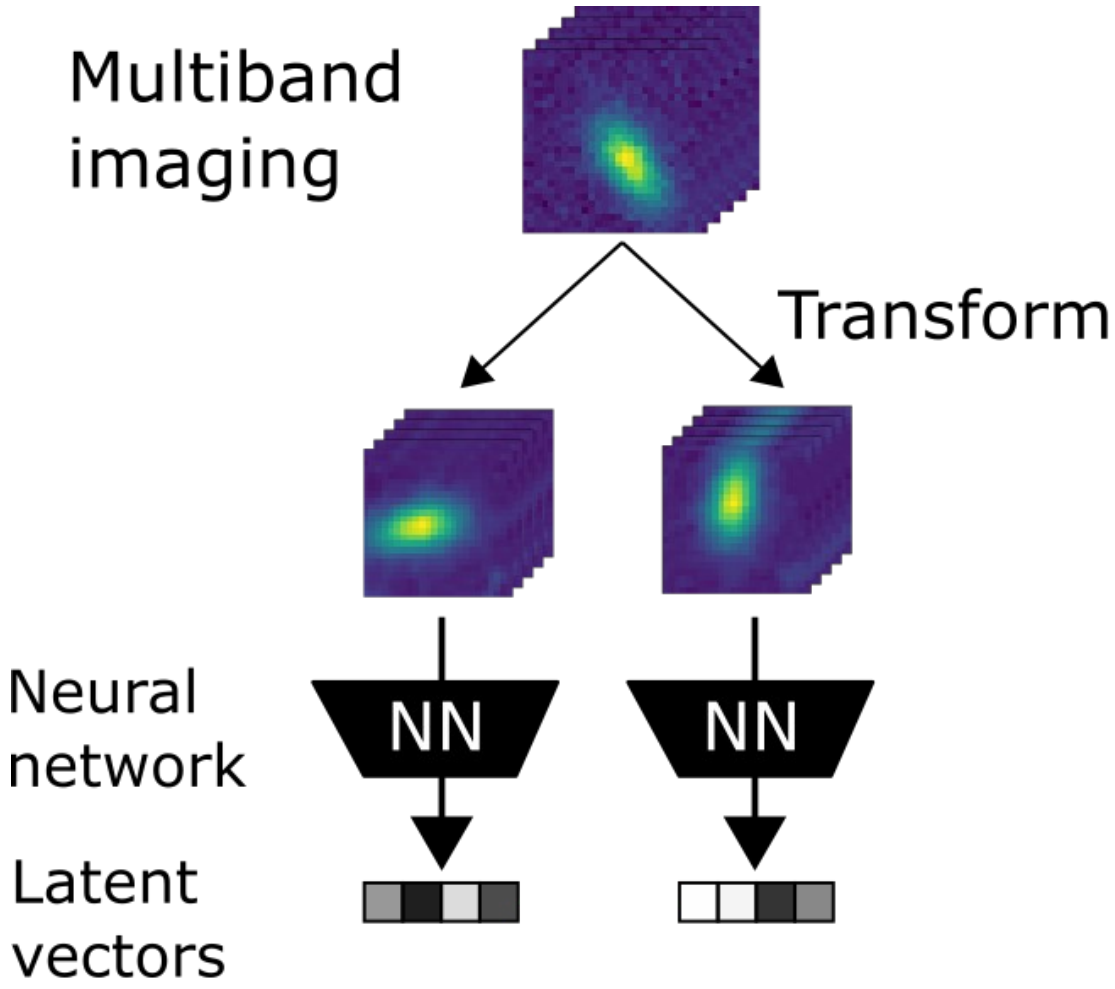
Byrne+24, in prep

How does Contrastive Learning work?



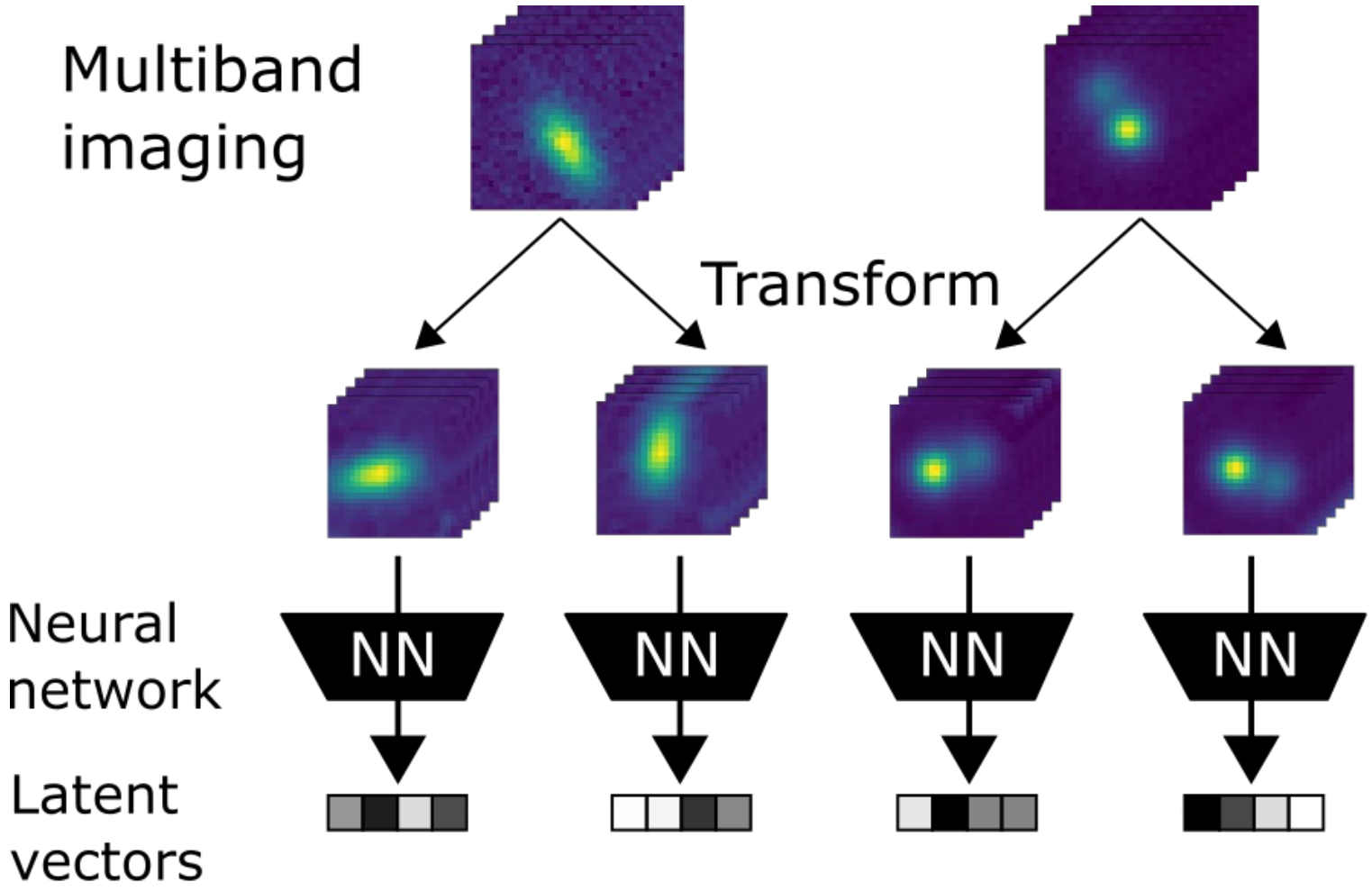
Byrne+24, in prep

How does Contrastive Learning work?



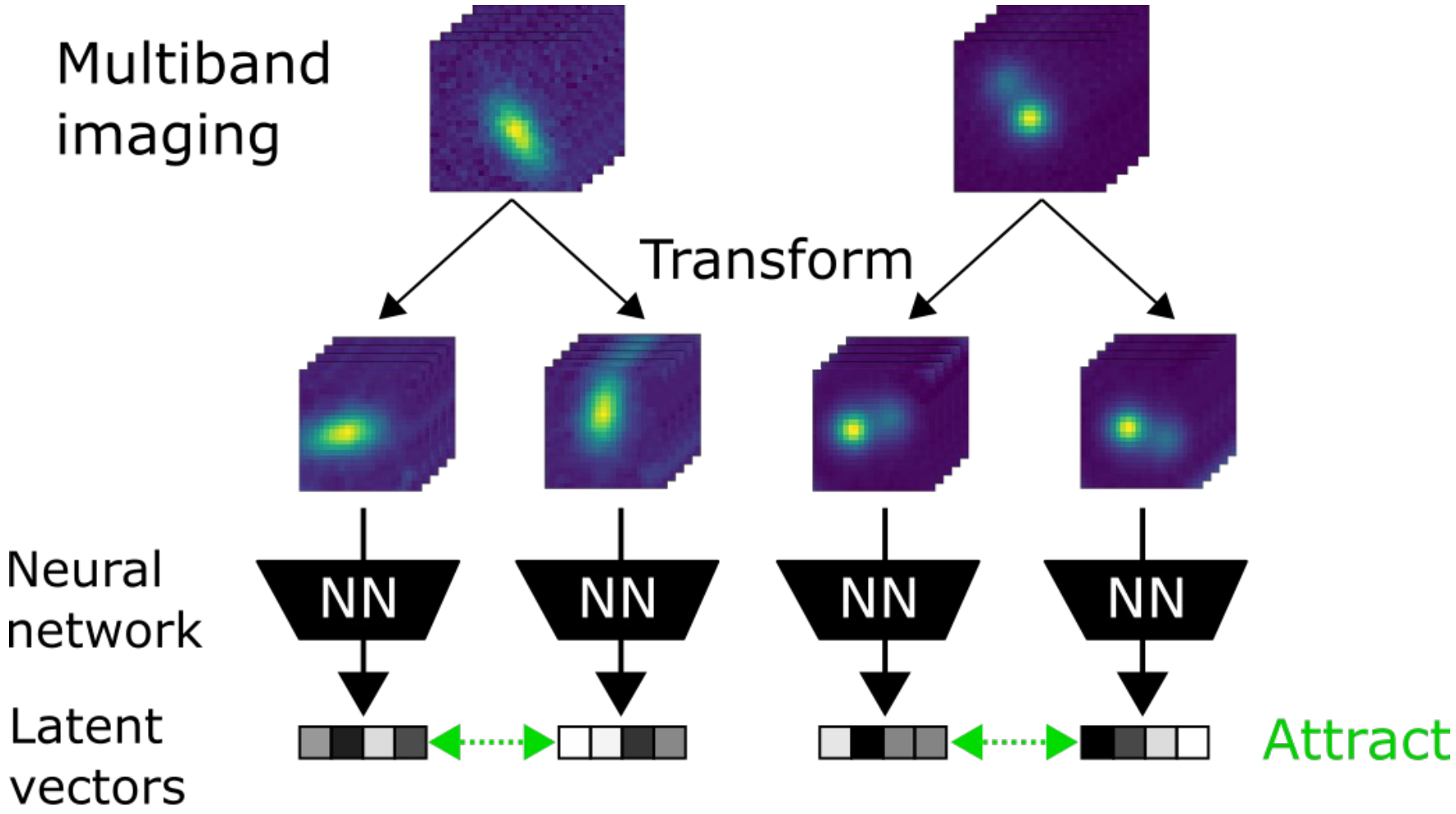
Byrne+24, in prep

How does Contrastive Learning work?



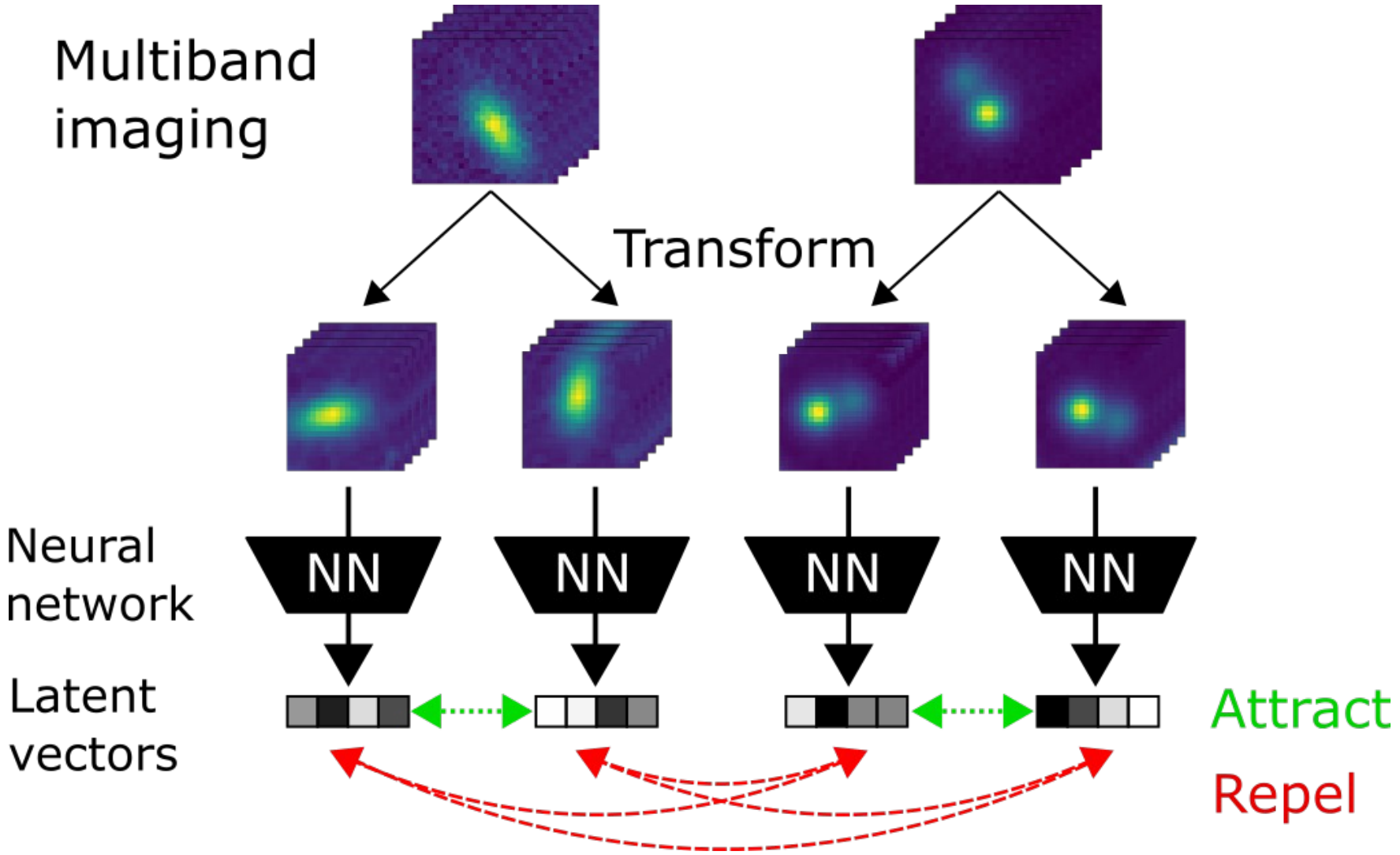
Byrne+24, in prep

How does Contrastive Learning work?



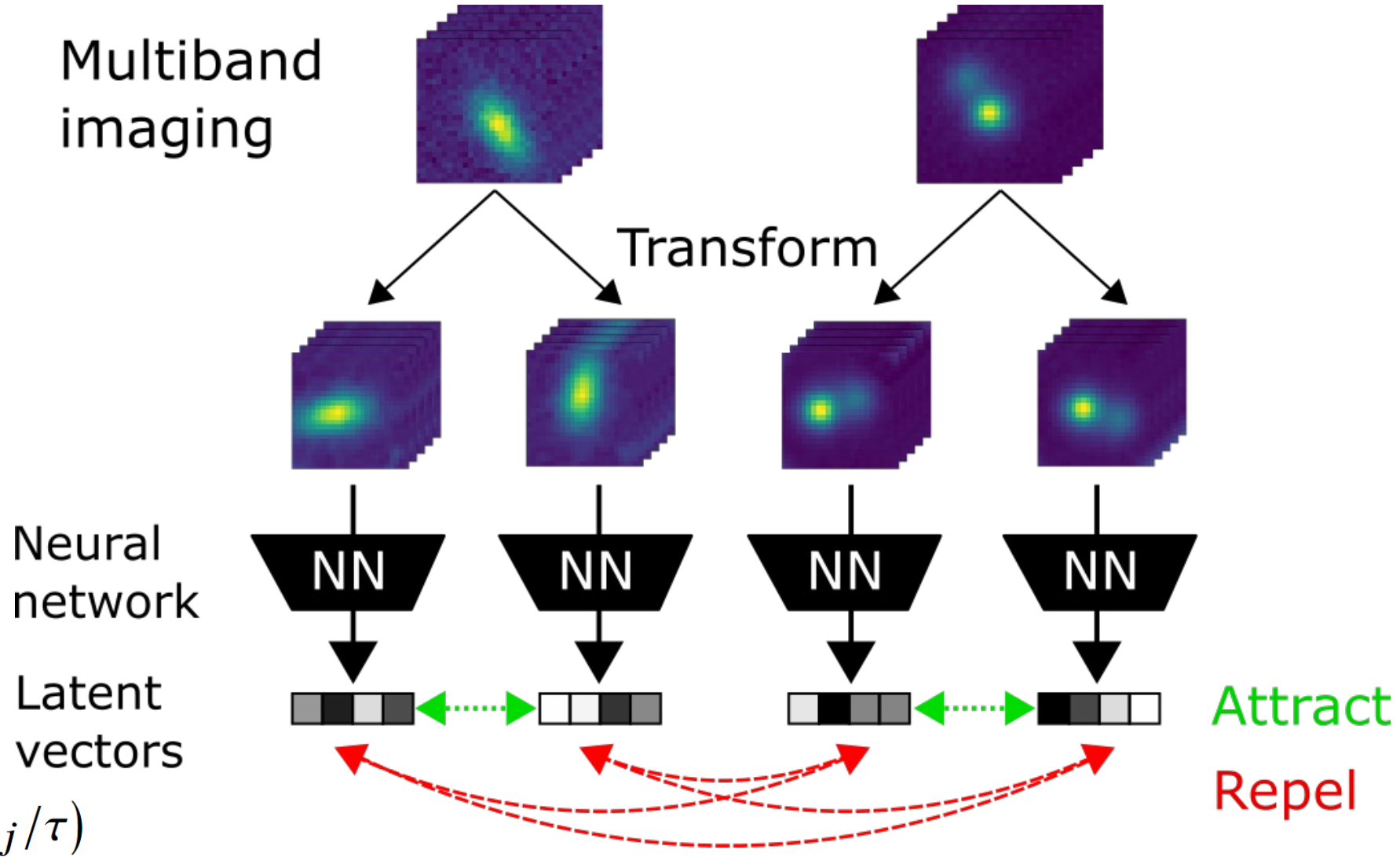
Byrne+24, in prep

How does Contrastive Learning work?



Byrne+24, in prep

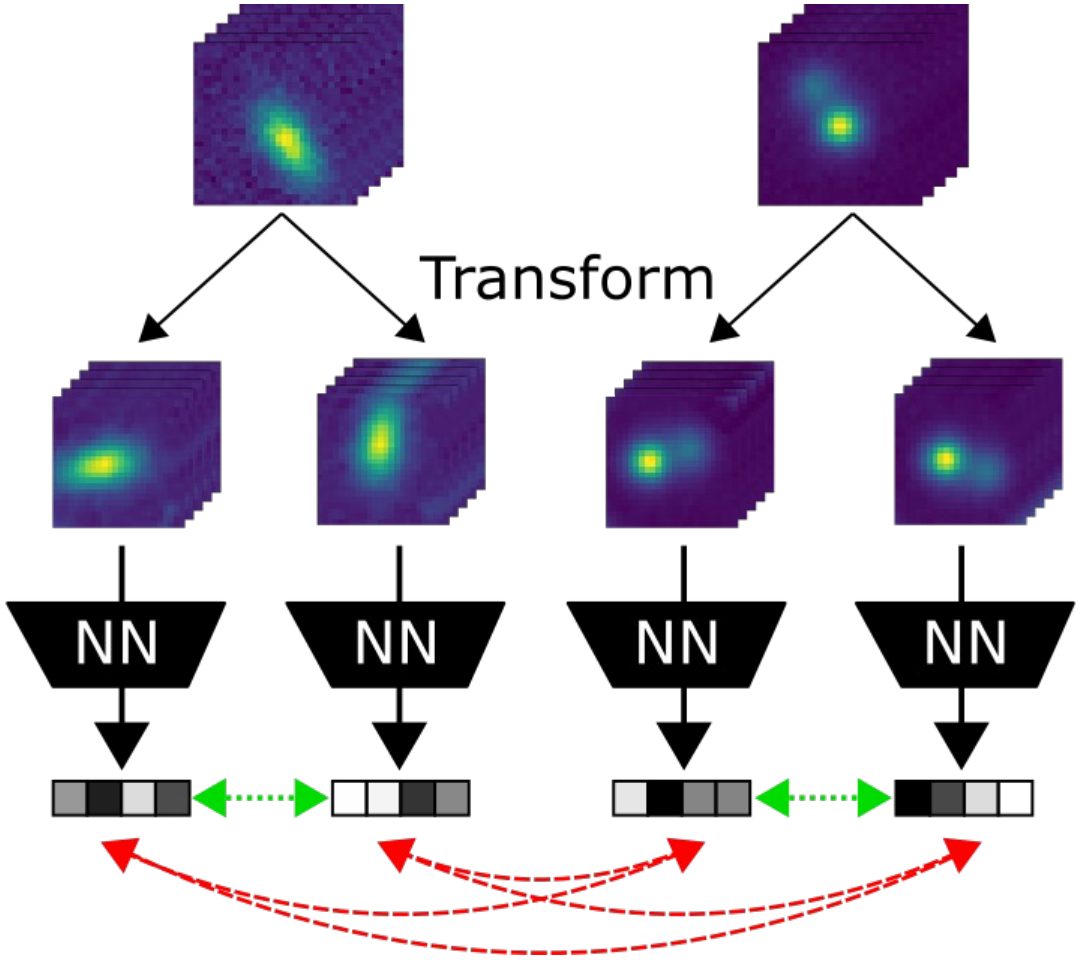
How does Contrastive Learning work?



$$L_{i;j} = -\log \frac{\exp(\hat{z}_i \cdot \hat{z}_j / \tau)}{\sum_{k=1,k\neq i}^{2N} \exp(\hat{z}_i \cdot \hat{z}_k / \tau)}$$

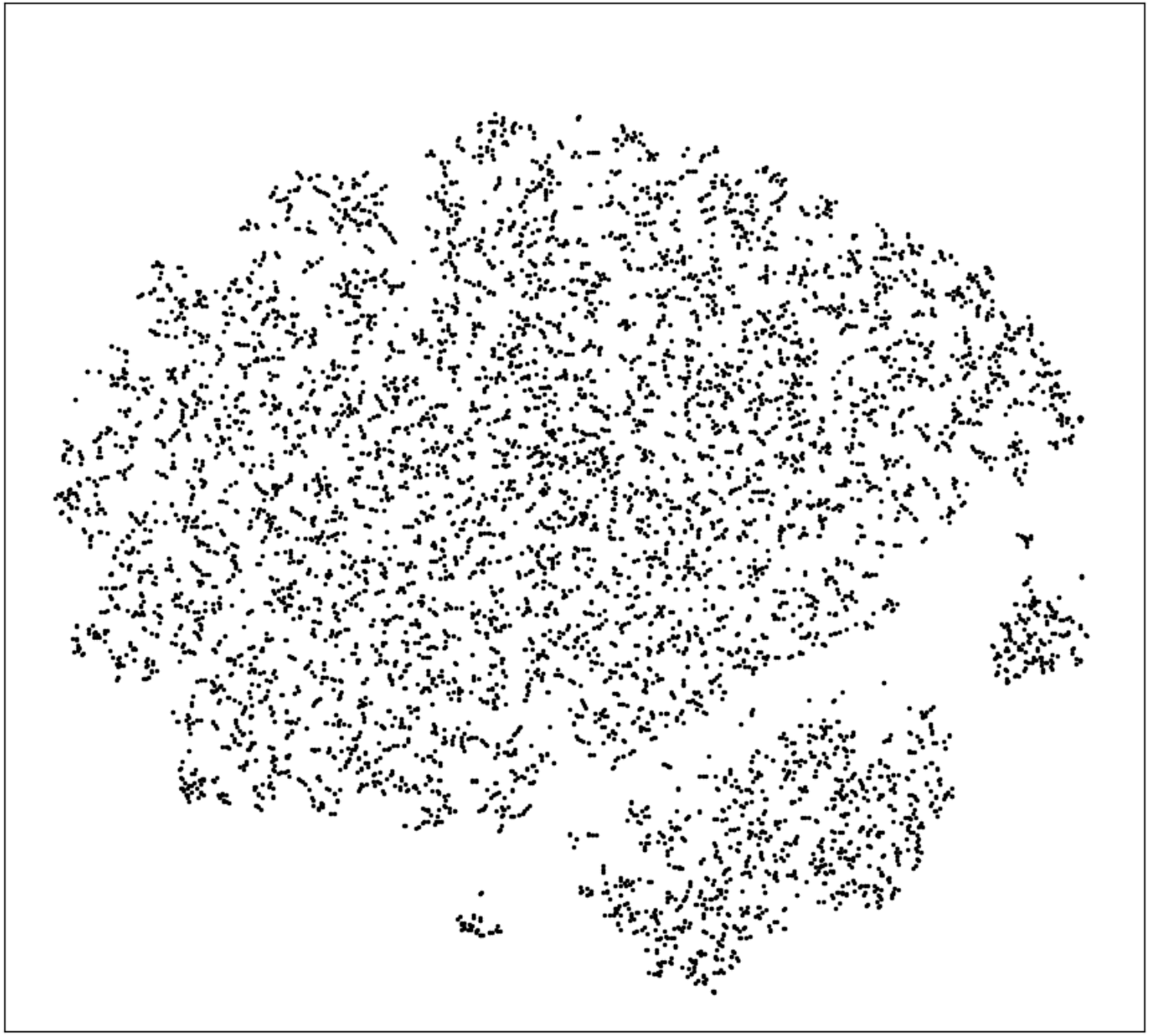
Algorithm summary

- Make two transformations of each training instance
- Project them into a latent space with a NN
- Train NN by...
  - ...rewarding proximity of projections of the same image
  - ...penalising proximity of projections of different images



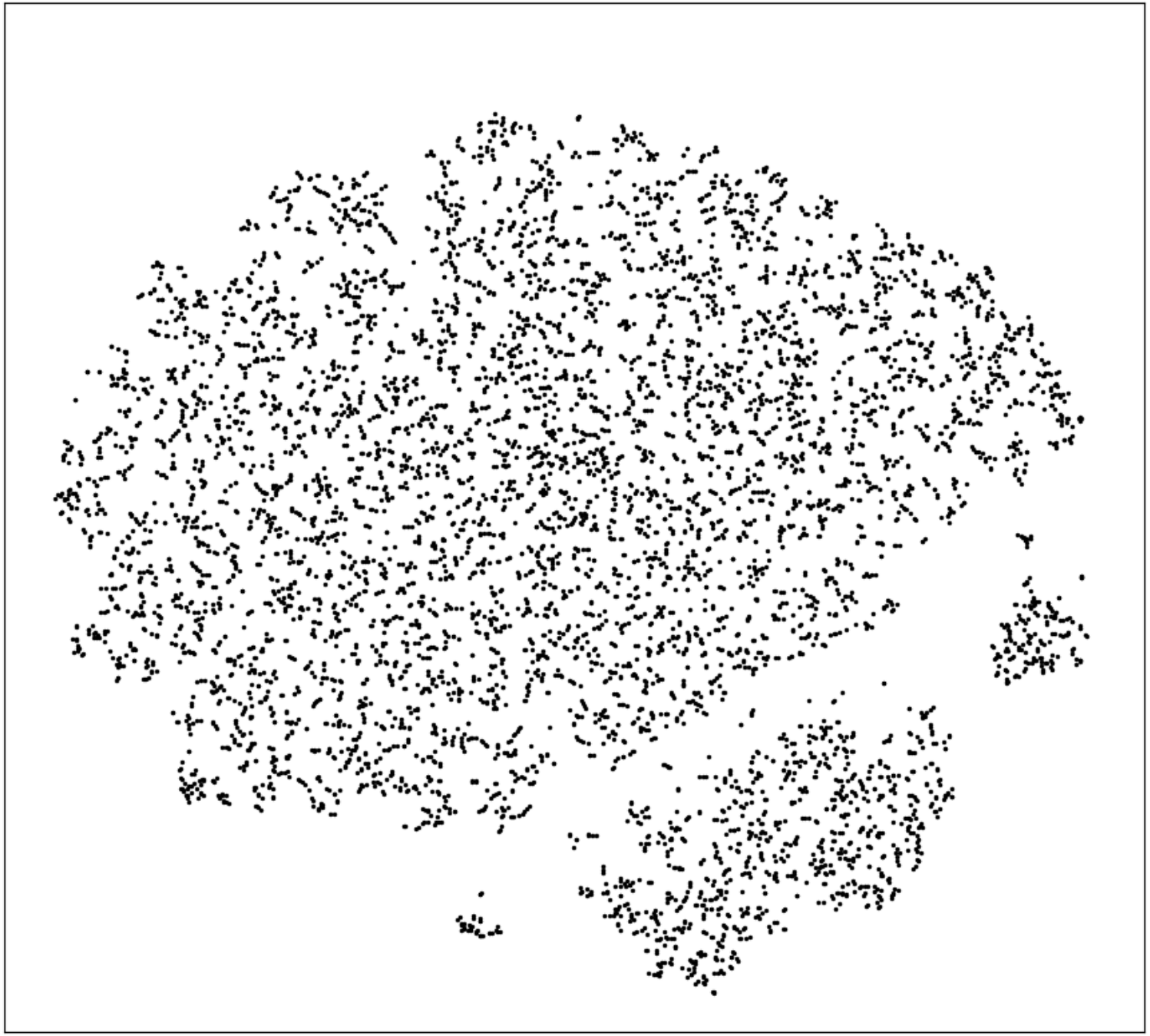
After Training

- Training data then fed into trained NN



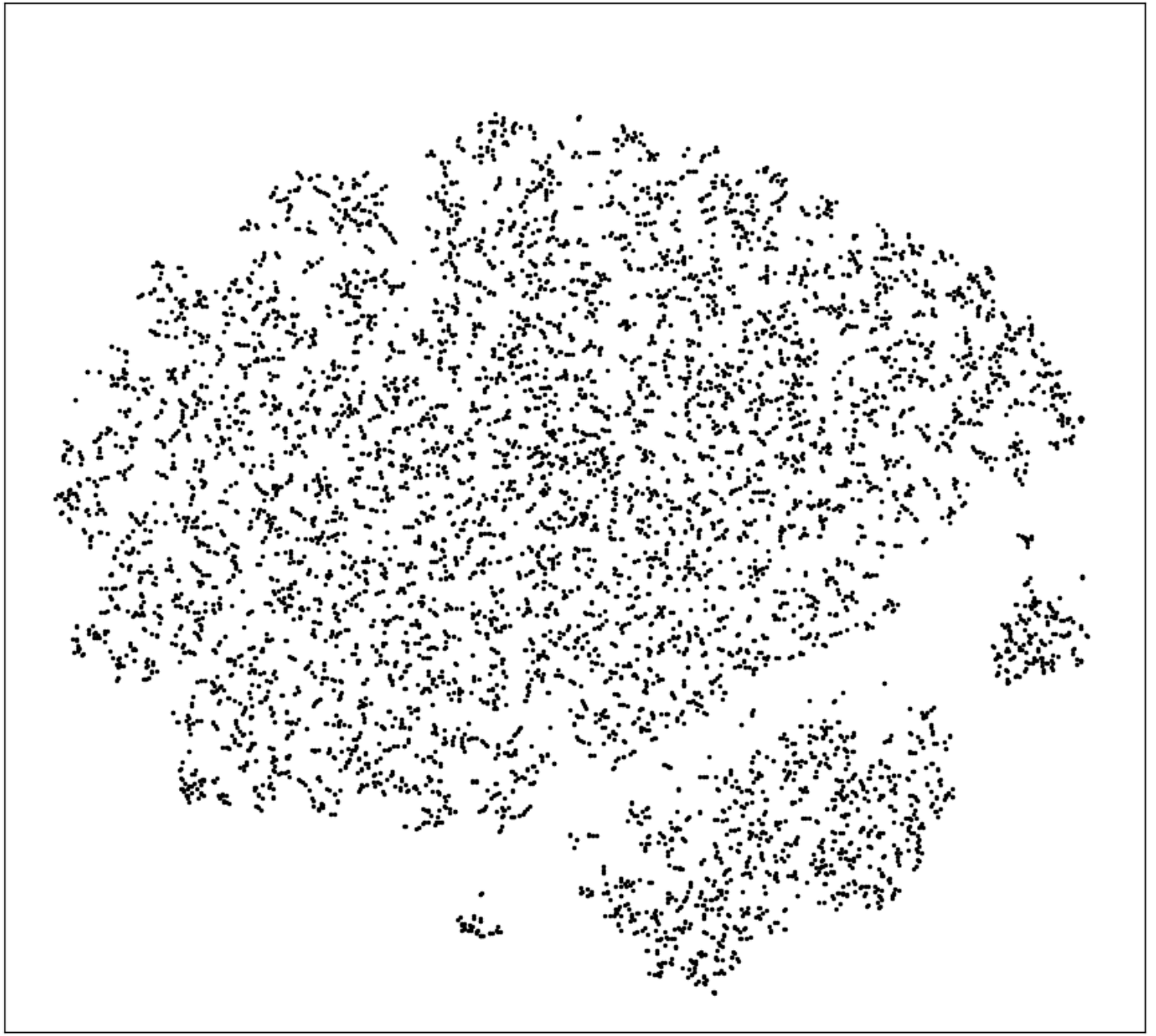
After Training

- Training data then fed into trained NN
- Data clustered in (high-D) latent space



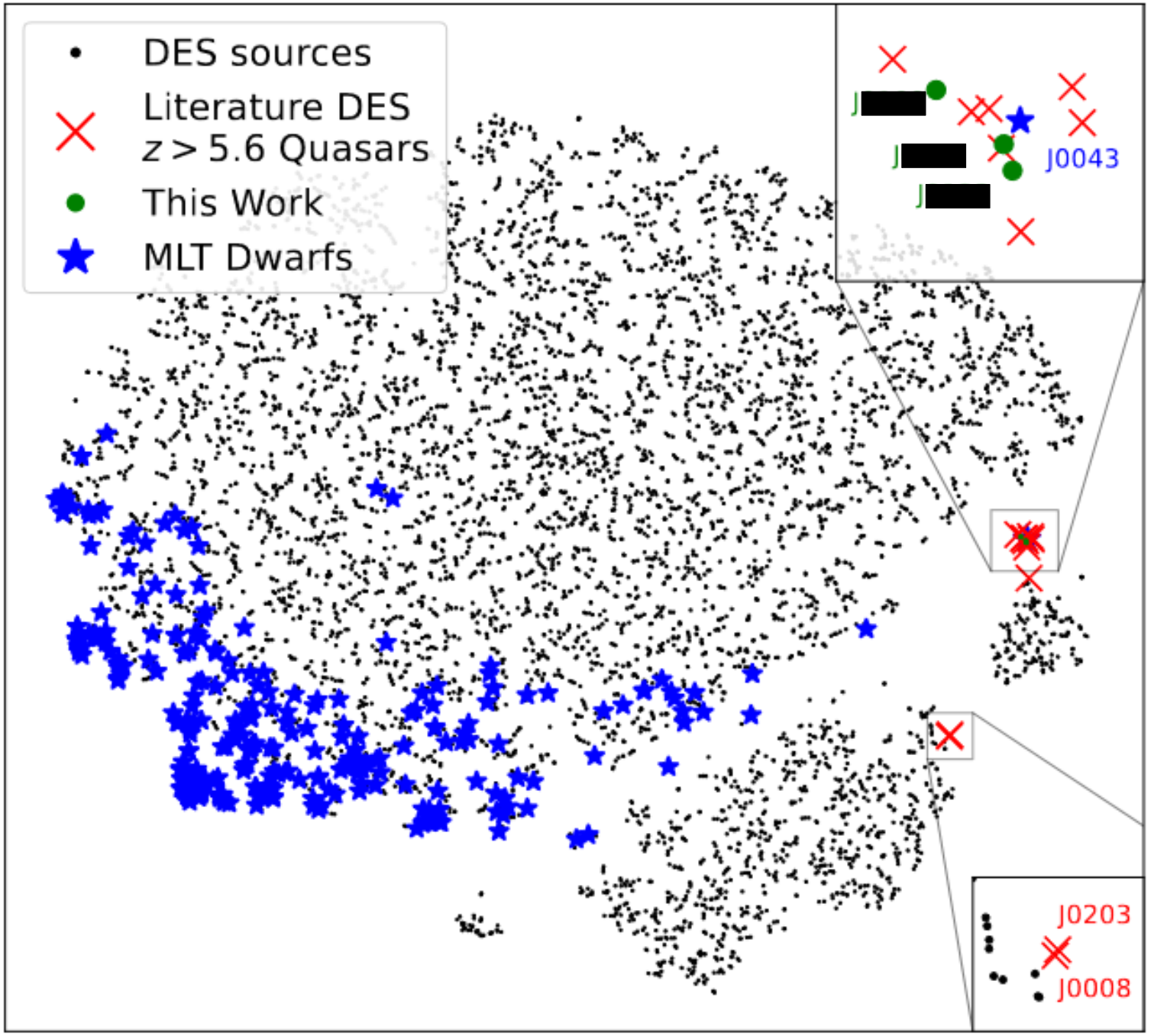
After Training

- Training data then fed into trained NN
- Data clustered in (high-D) latent space
- Dimensionality reduction then employed (PCA, *t*-SNE, etc.)



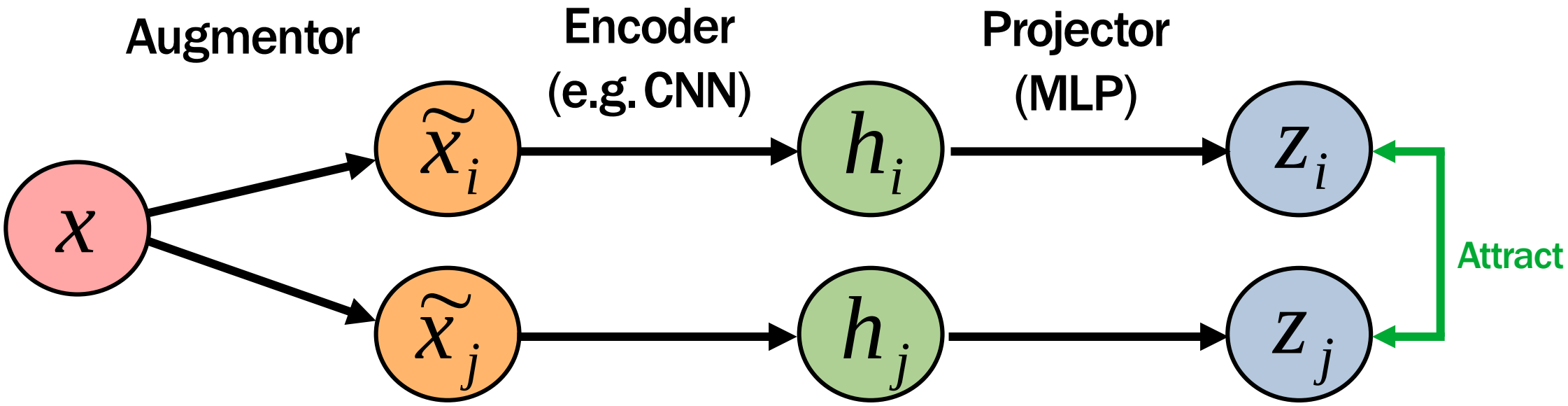
After Training

- Training data then fed into trained NN
- Data clustered in (high-D) latent space
- Dimensionality reduction then employed (PCA, *t*-SNE, etc.)

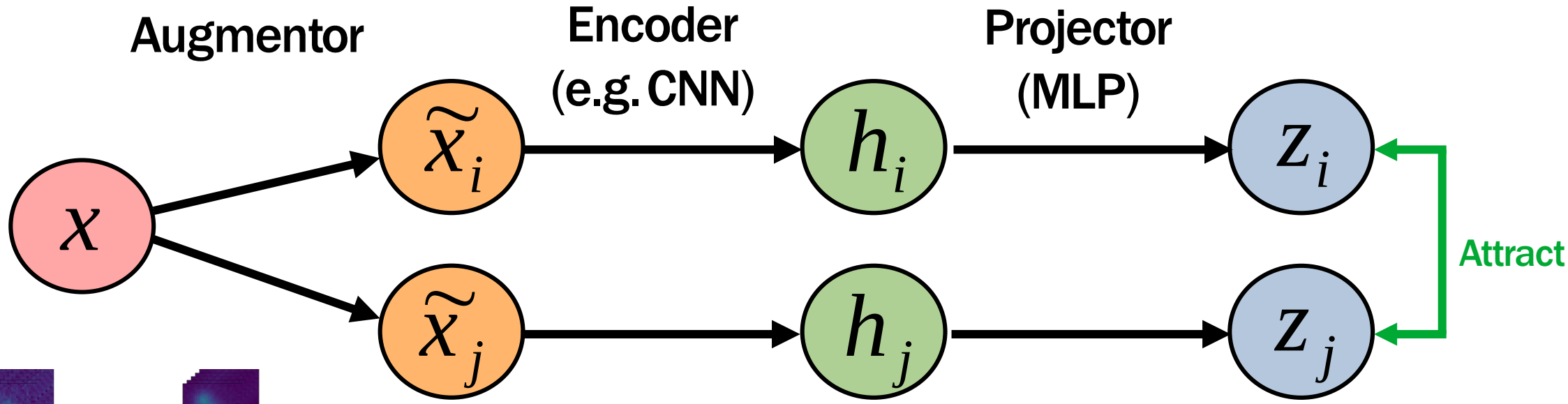


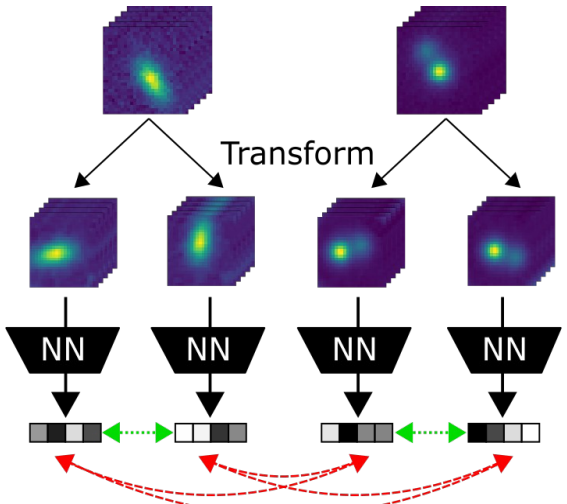
Byrne+24, in prep

Deep Dive into Contrastive Framework



Deep Dive into Contrastive Framework





Deep Dive: Augmentor

- Choice of transformations is...
- ...data-type-dependent

Keras

Preprocessing layers

Image augmentation layers

- RandomCrop layer
- RandomFlip layer
- RandomTranslation layer
- RandomRotation layer
- RandomZoom layer
- RandomContrast layer
- RandomBrightness layer

Deep Dive: Augmentor

- Choice of transformations is...
- ...data-type-dependent

Keras

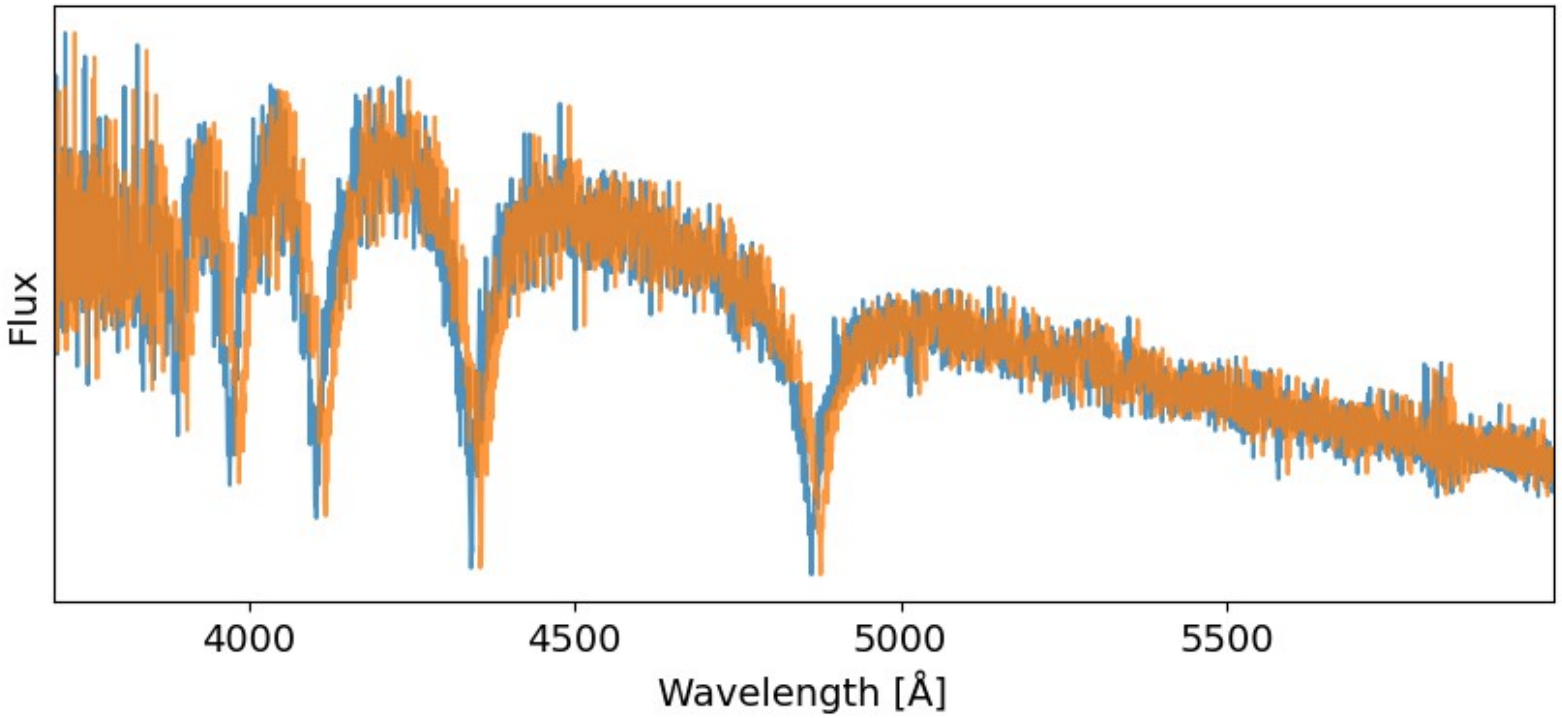
Preprocessing layers

Image augmentation layers

- RandomCrop layer
- RandomFlip layer
- RandomTranslation layer
- RandomRotation layer
- RandomZoom layer
- RandomContrast layer
- RandomBrightness layer

Deep Dive: Augmentor

- Choice of transformations is...
- ...data-type-dependent



Keras

Preprocessing layers

Image augmentation layers

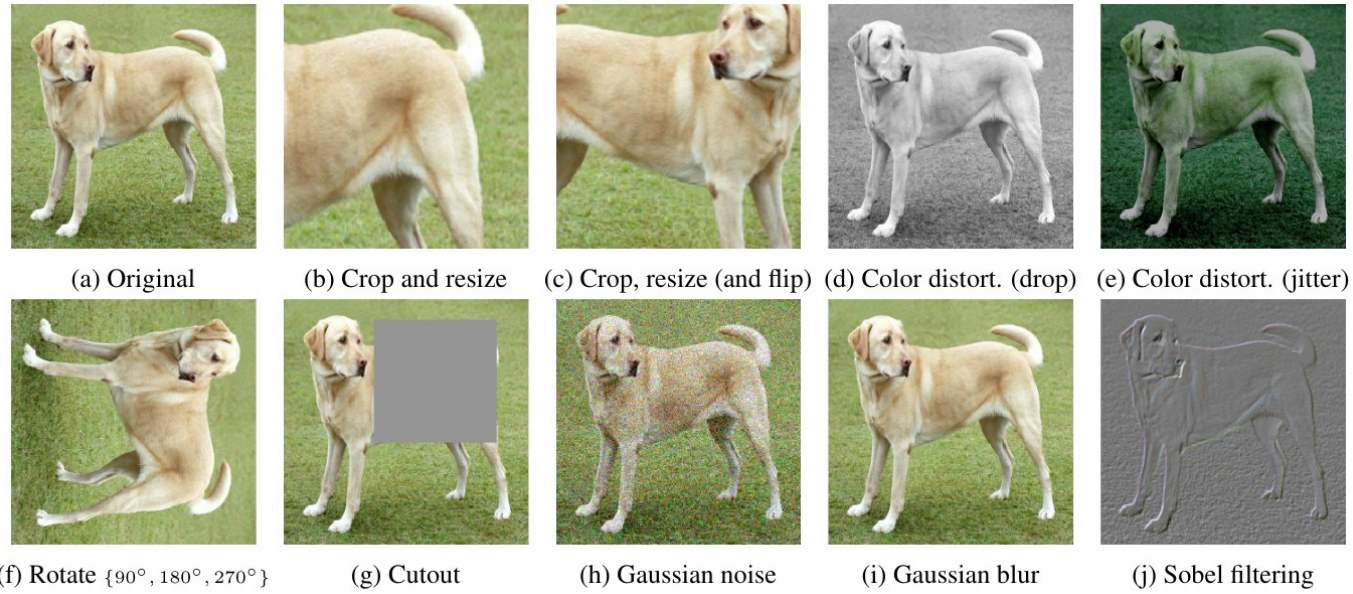
- RandomCrop layer
- RandomFlip layer
- RandomTranslation layer
- RandomRotation layer
- RandomZoom layer
- RandomContrast layer
- RandomBrightness layer

Deep Dive: Augmentor

- Choice of transformations is...
- ...data-type-dependent
- ...important

Deep Dive: Augmentor

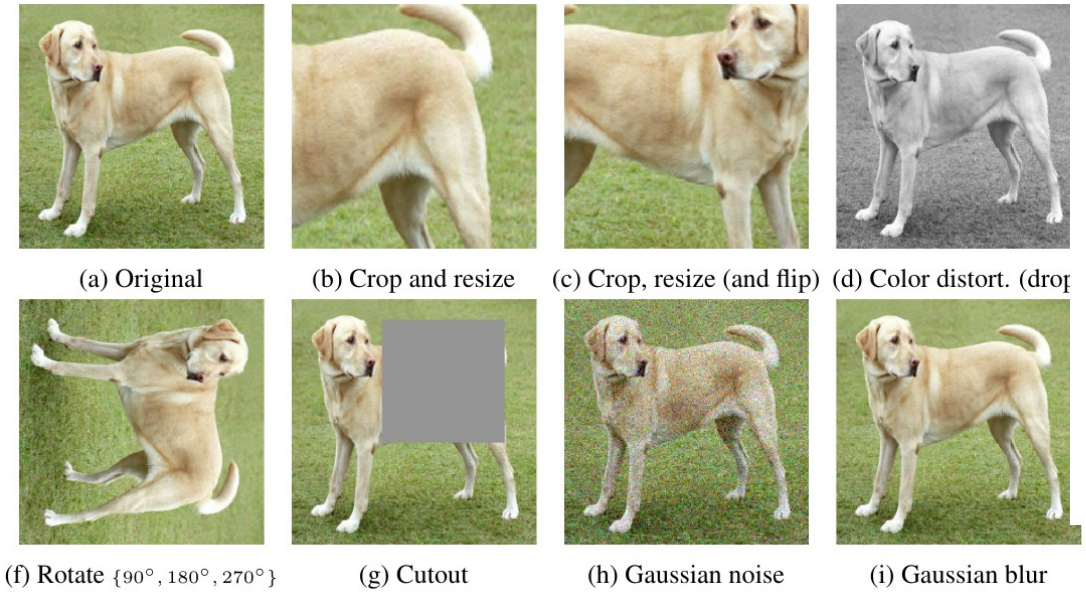
- Choice of transformations is...
- ...data-type-dependent
- ...important

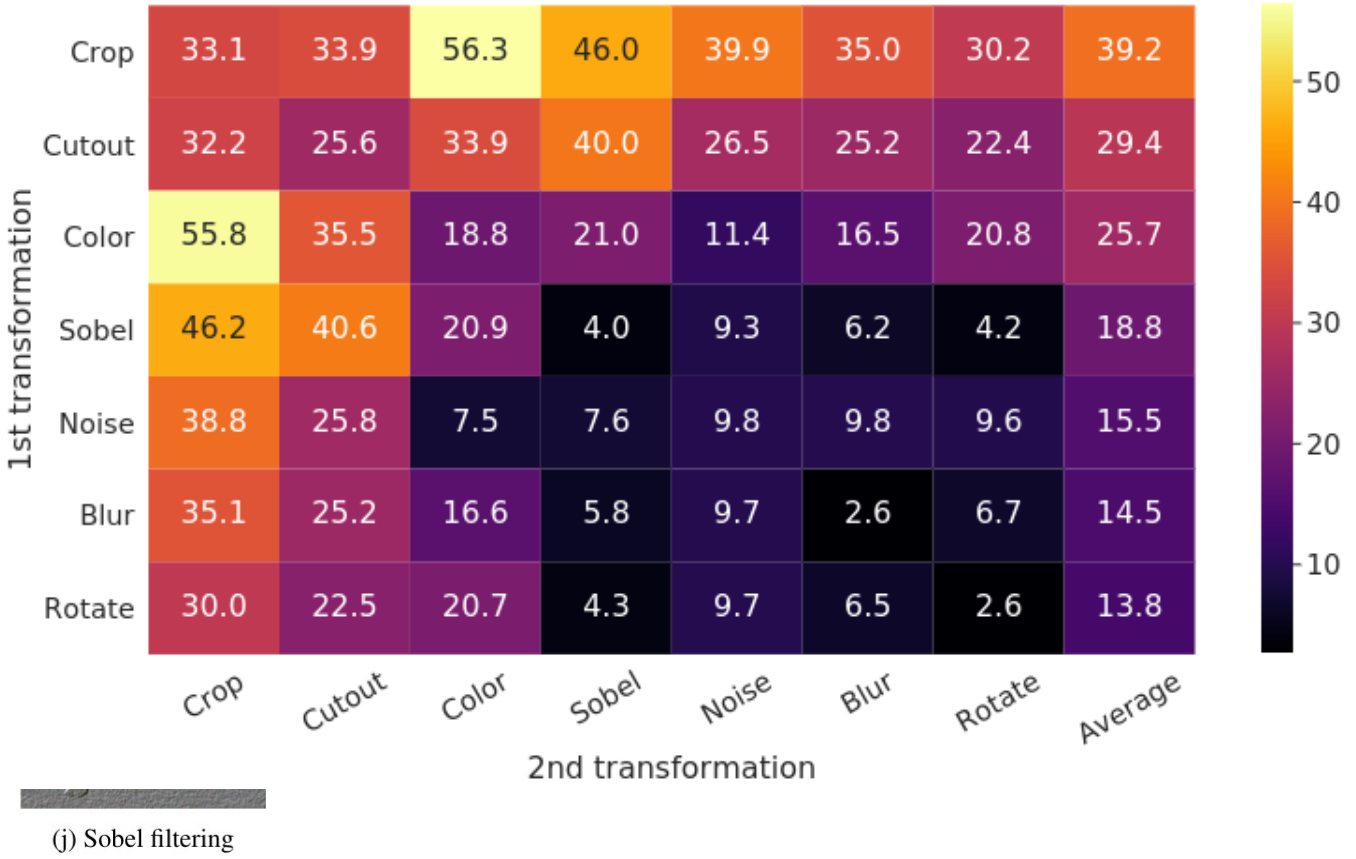


Chen+20

Deep Dive: Augmentor

- Choice of transformations is...
- ...data-type-dependent
- ...important

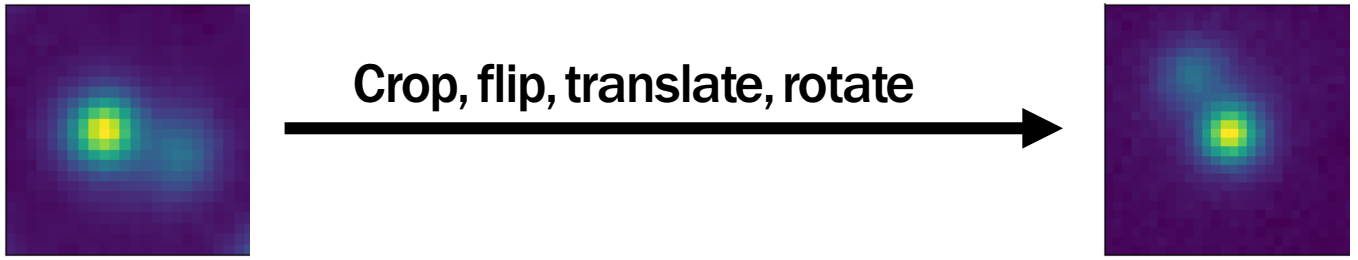




Chen+20

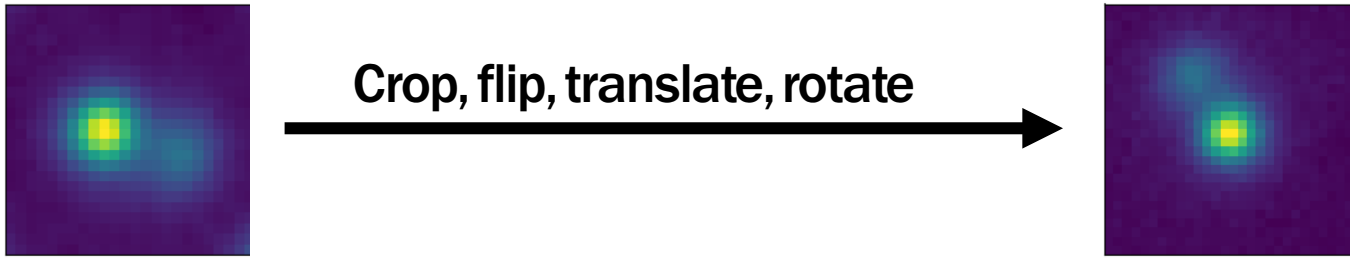
Deep Dive: Augmentor

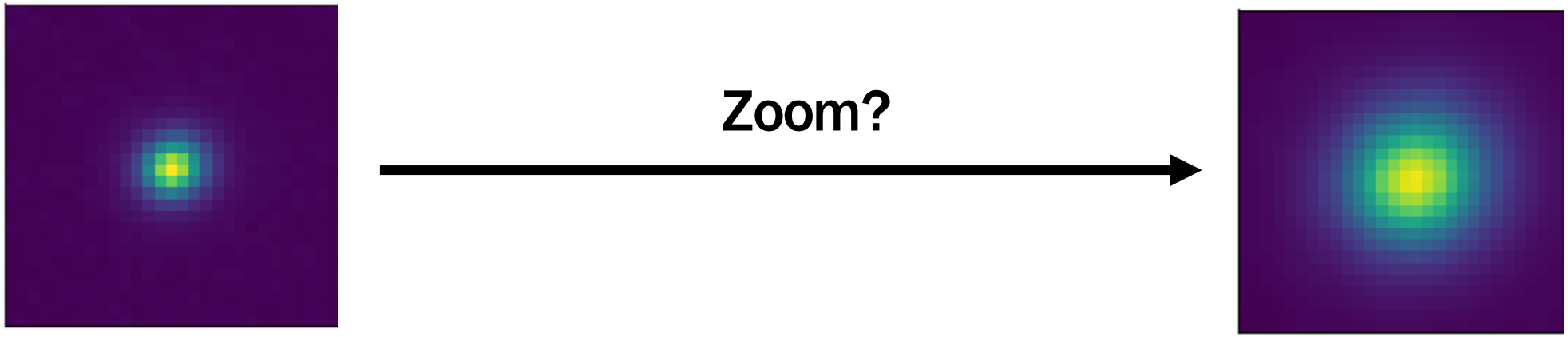
- Choice of transformations is...
- ...data-type-dependent
- ...important
- ...case-dependent



Deep Dive: Augmentor

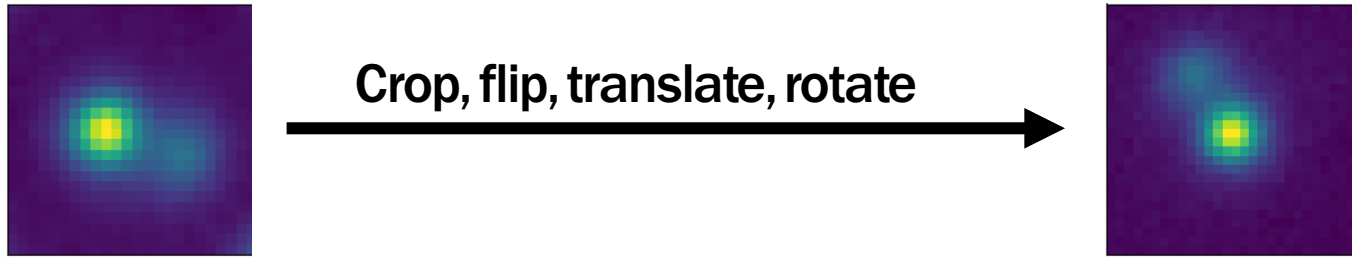
- Choice of transformations is...
- ...data-type-dependent
- ...important
- ...case-dependent

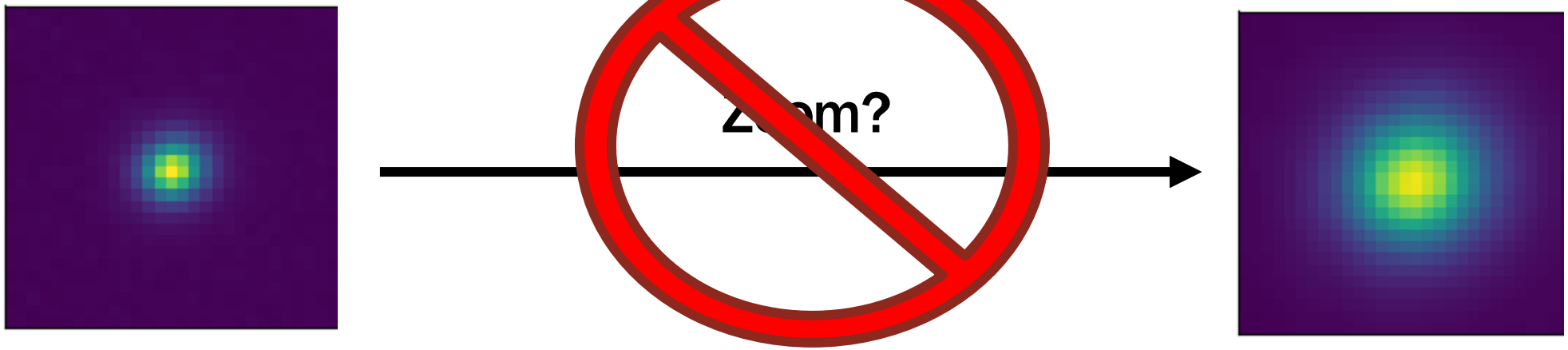




Deep Dive: Augmentor

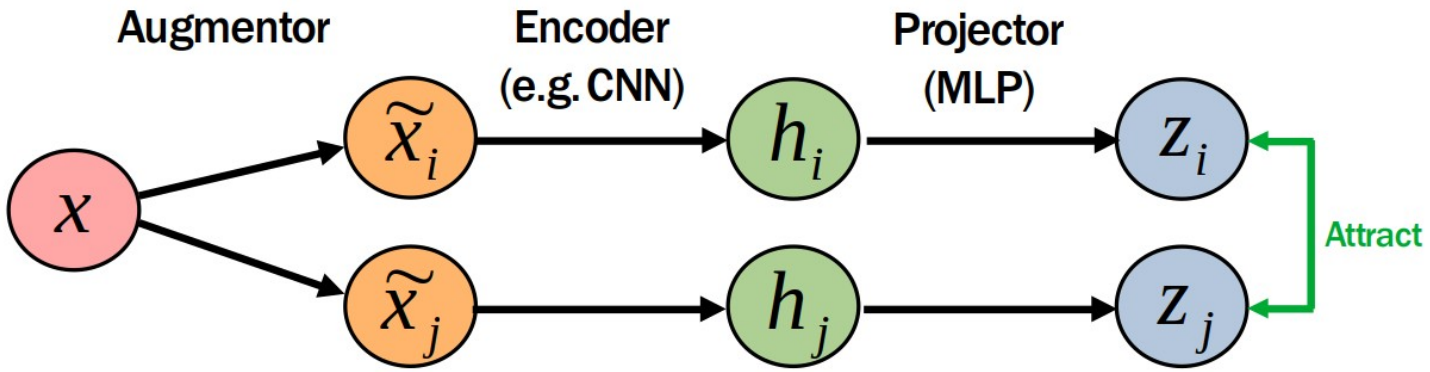
- Choice of transformations is...
- ...data-type-dependent
- ...important
- ...case-dependent





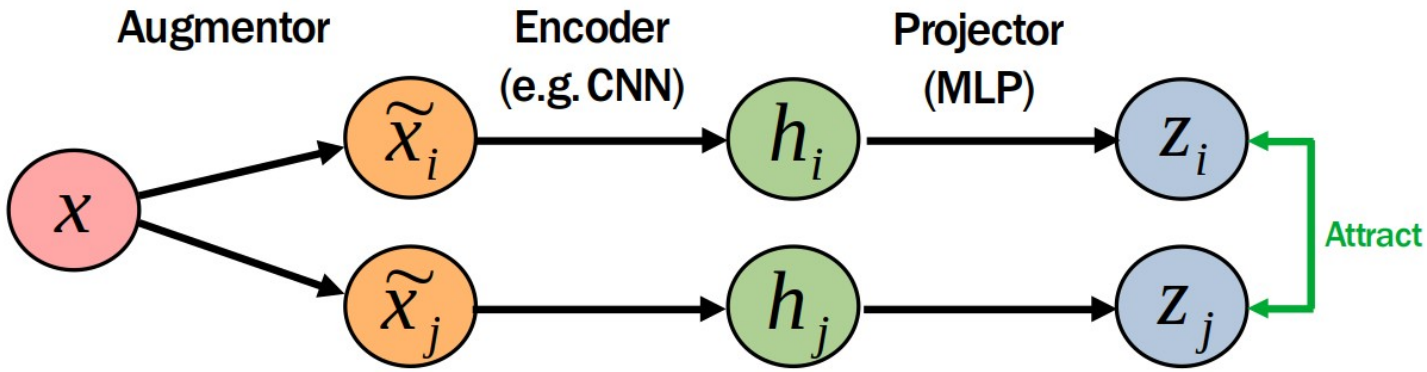
Deep Dive: Encoder & Projector

Training:

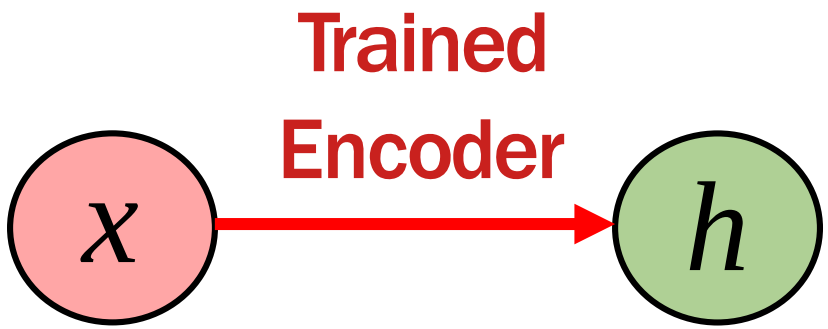


Deep Dive: Encoder & Projector

Training:



Clustering:



>10% better!

How good is Contrastive Learning?

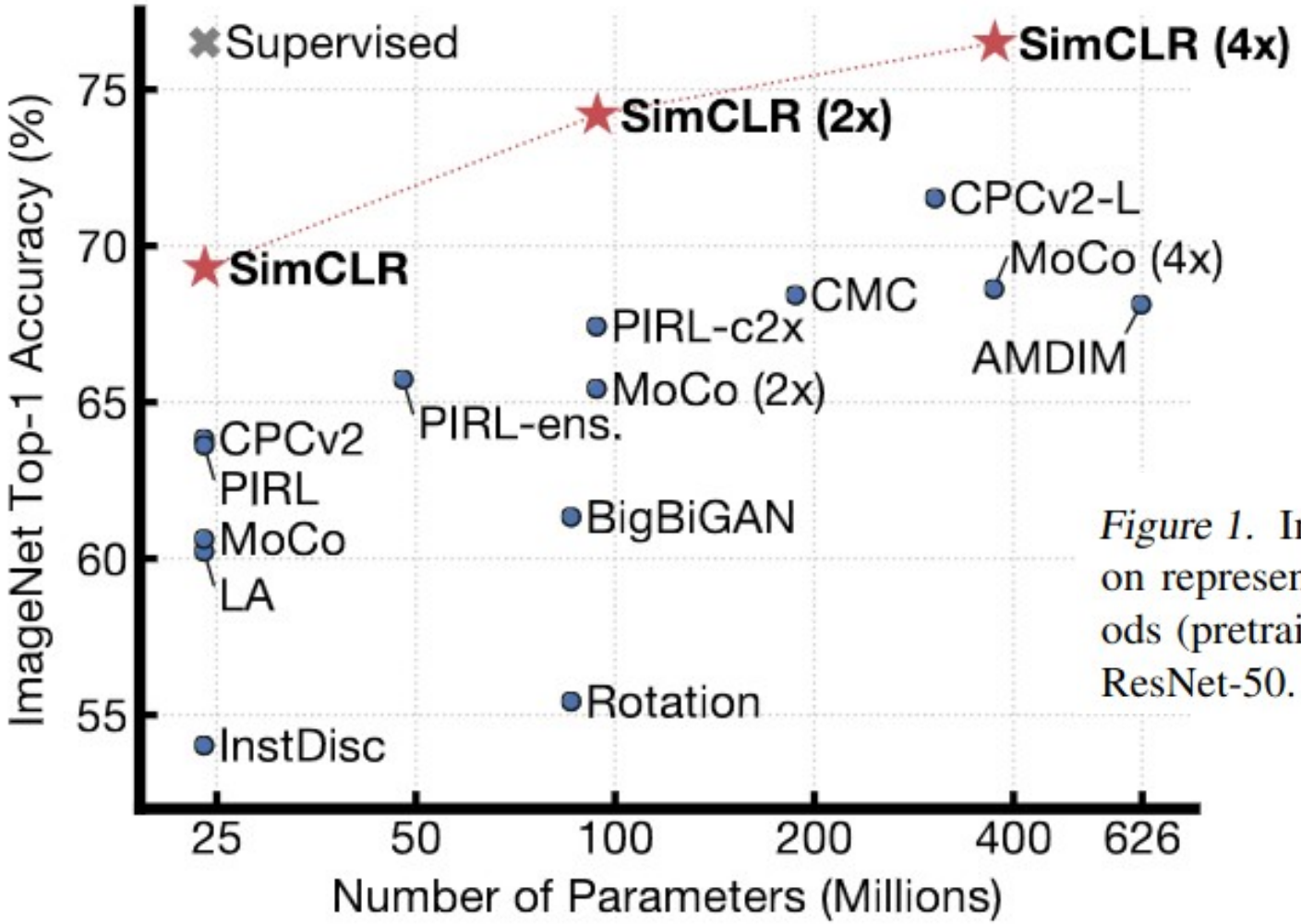
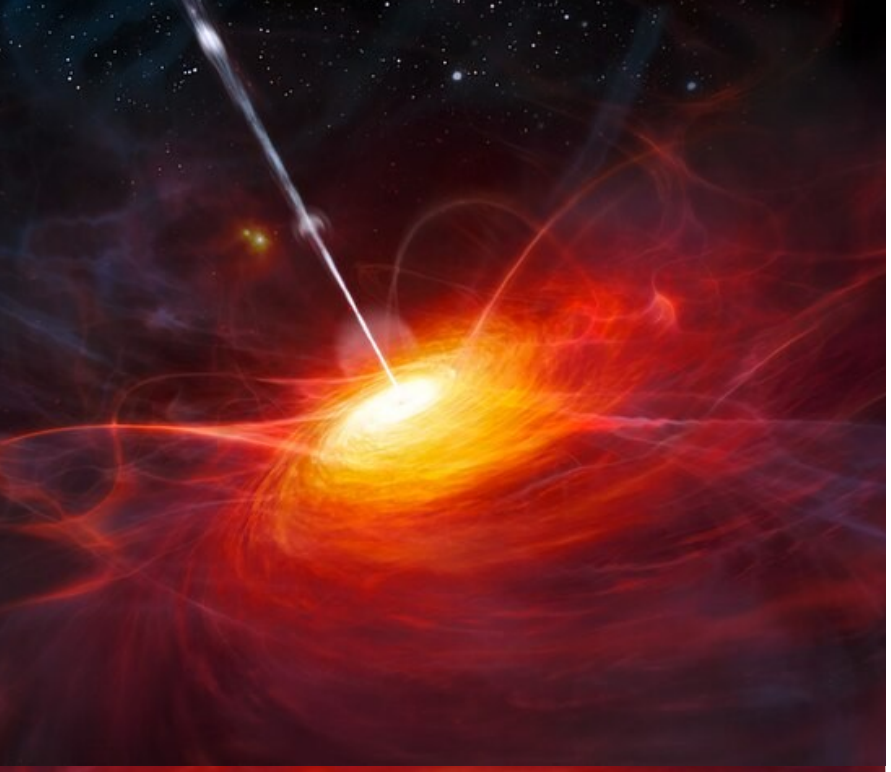


Figure 1. ImageNet Top-1 accuracy of linear classifiers trained on representations learned with different self-supervised methods (pretrained on ImageNet). Gray cross indicates supervised ResNet-50. Our method, SimCLR, is shown in bold.

Chen+20



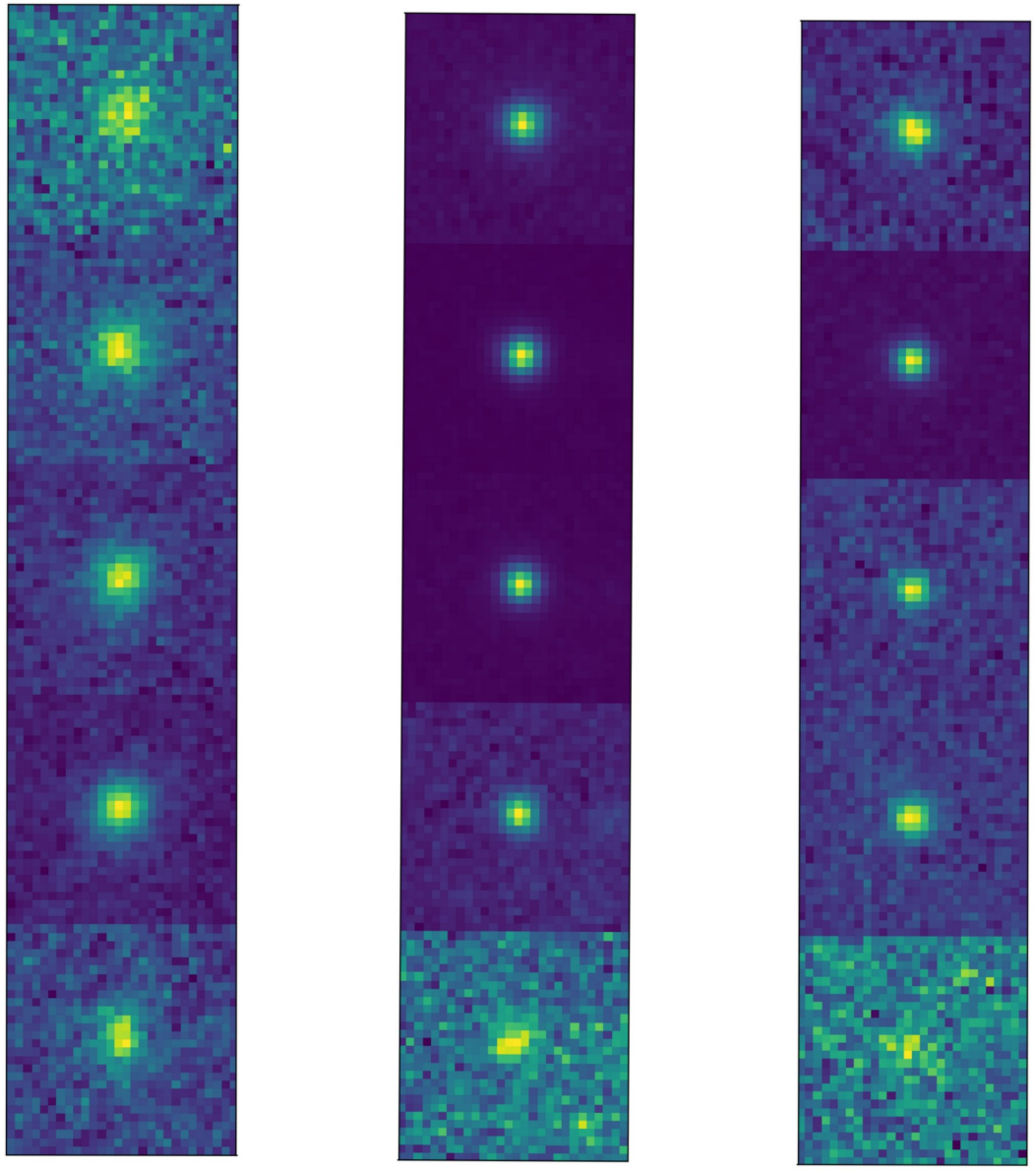
Conclusions

- Contrastive Learning is a useful framework for unsupervised clustering
- Trains a NN to ignore trivial transformations of training data
- Applicable to a range of use cases
- Questions?

ajnb3@cam --- Xander Byrne --- xbyrne.github.io

After Training

- Training data then fed into trained NN
- Data clustered in (high-D) latent space
- Dimensionality reduction then employed (PCA, *t*-SNE, etc.)



Data Augmentation

- Supervised learning constrained by size of training set
- But data collection is expensive
- Data augmentation adds new ‘obvious’ training examples