INSTITUTE FOR ARTIFICIAL INTELLIGENCE

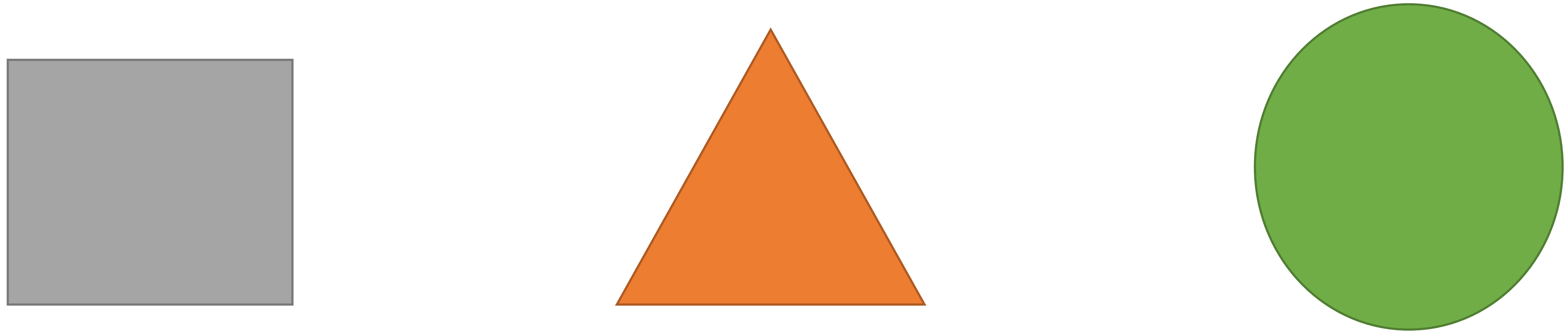
UNIVERSITY OF SOUTH FLORIDA

Learning useful representations of inputs, without labels

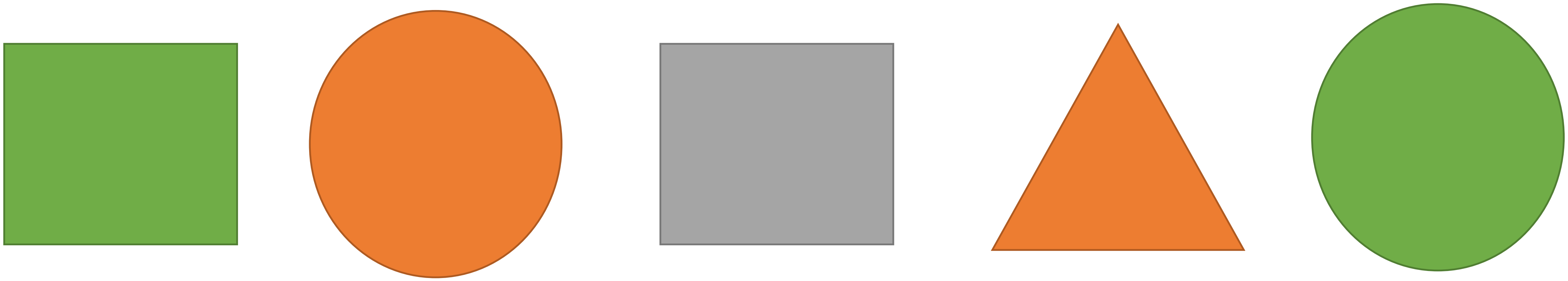
by

Ramy Mounir

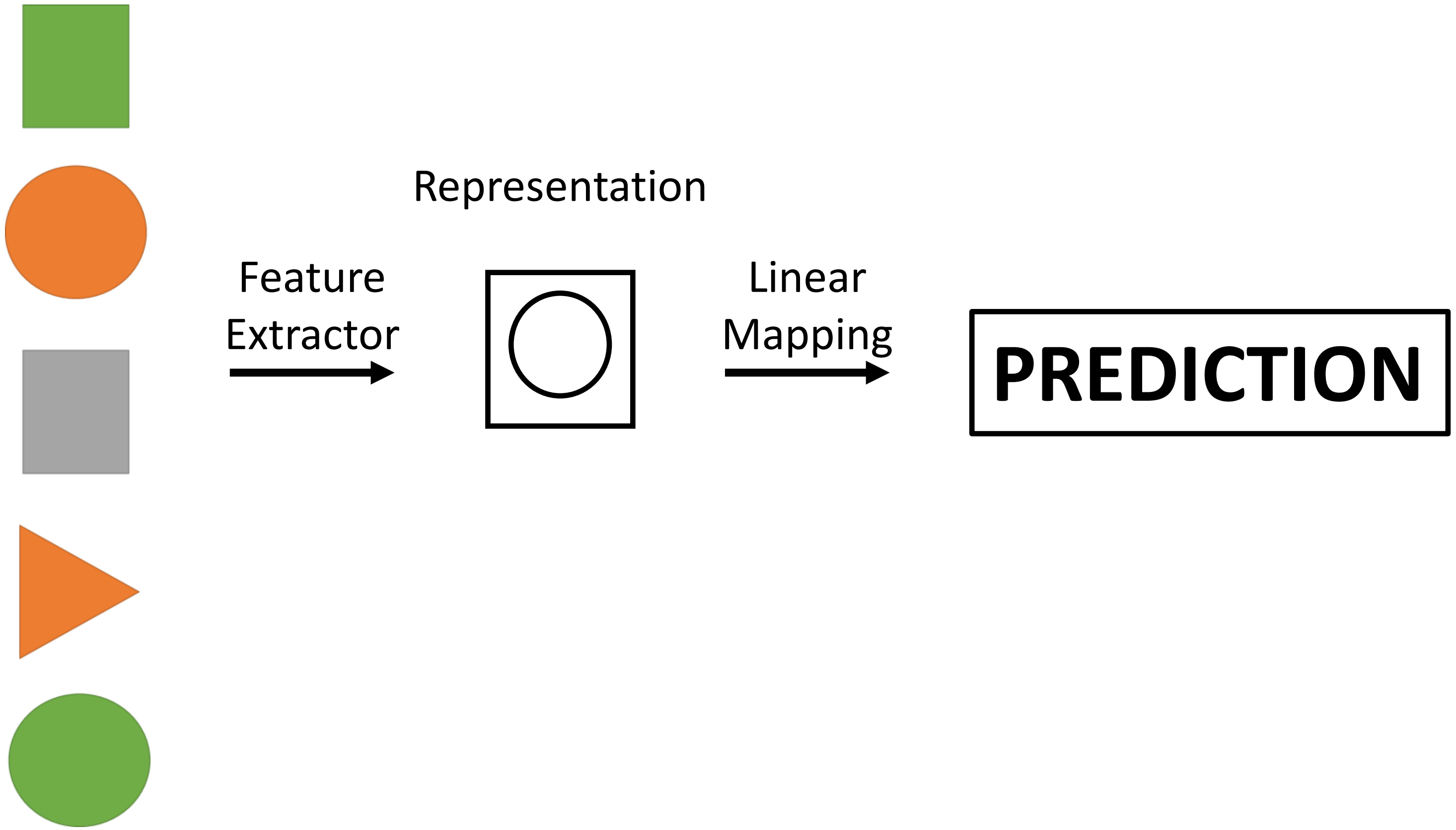
What is a Representation?



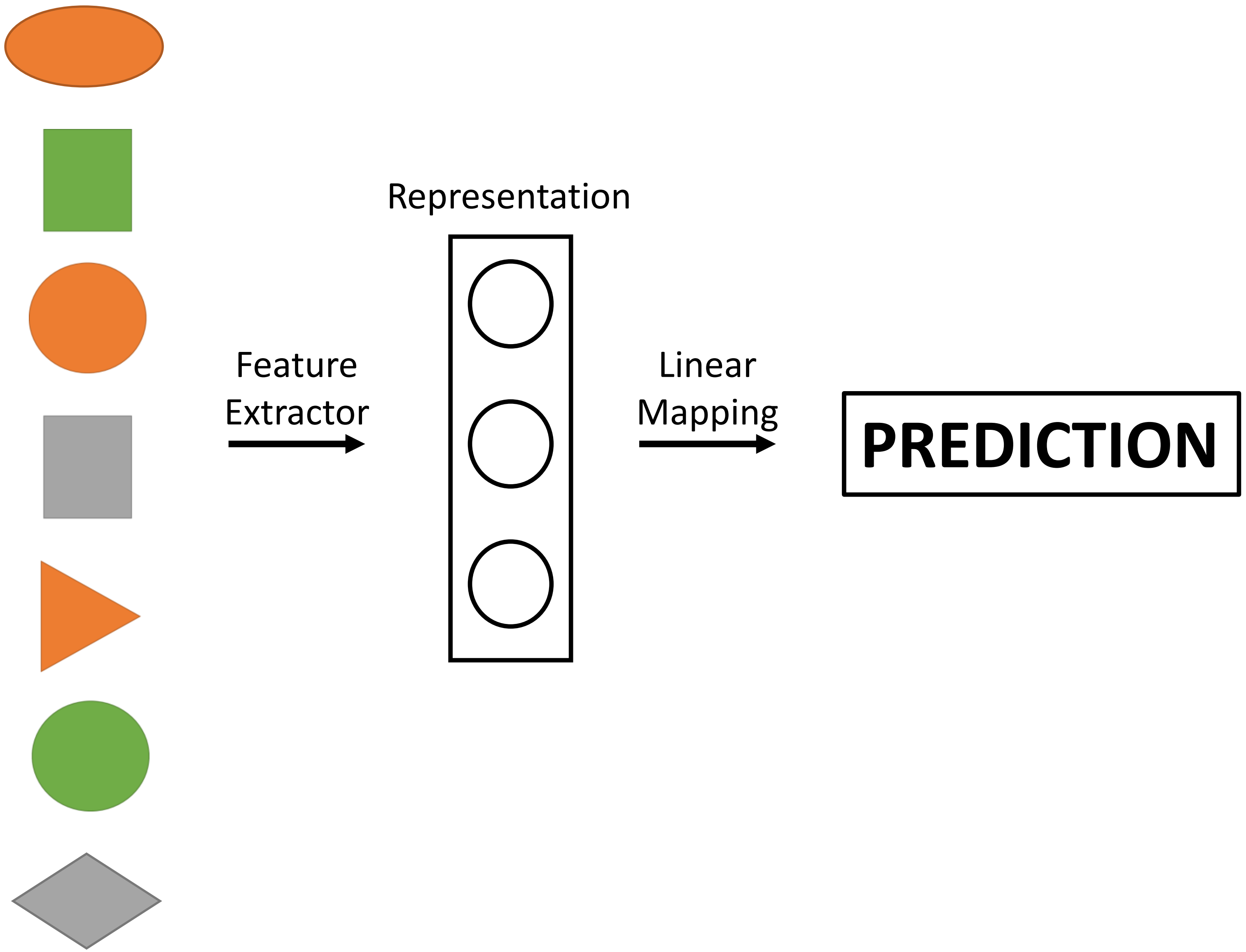
What is a Representation?



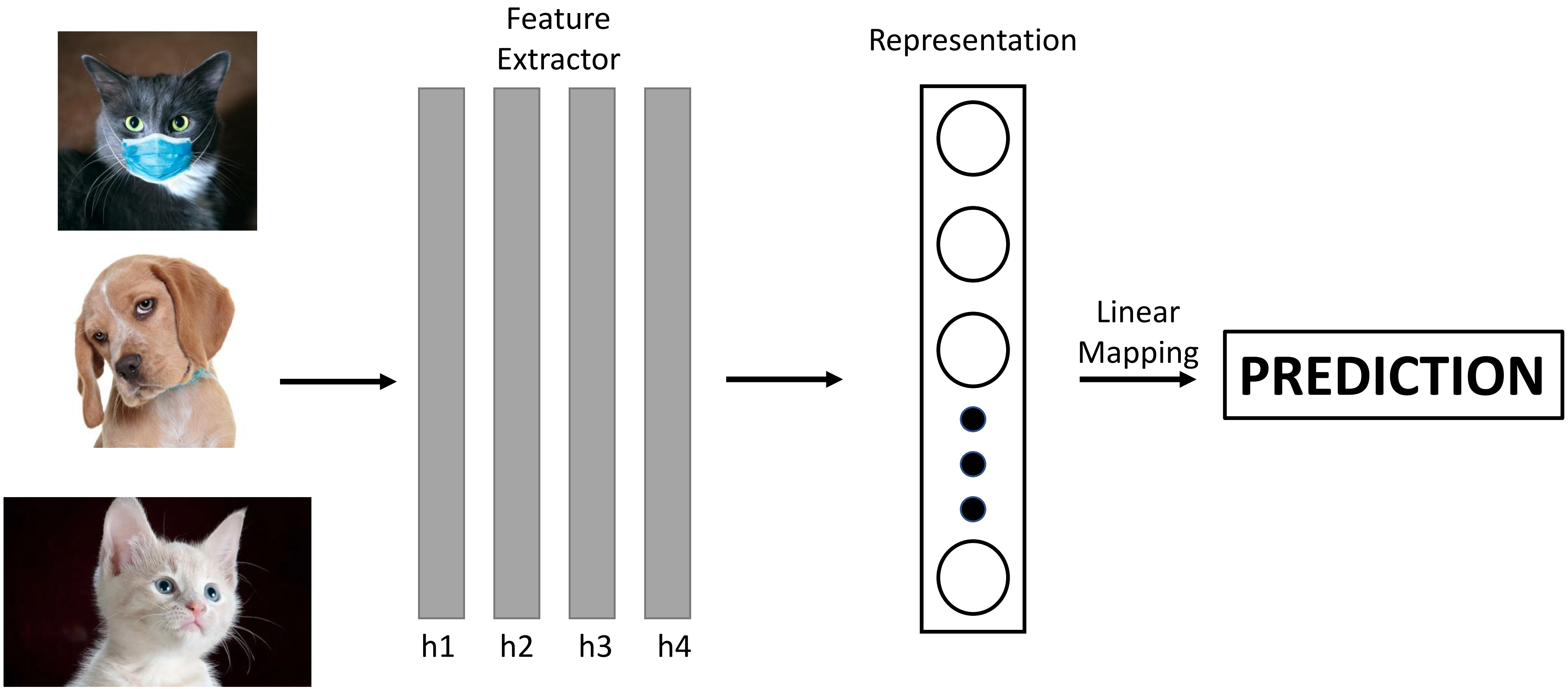
What is a Representation?



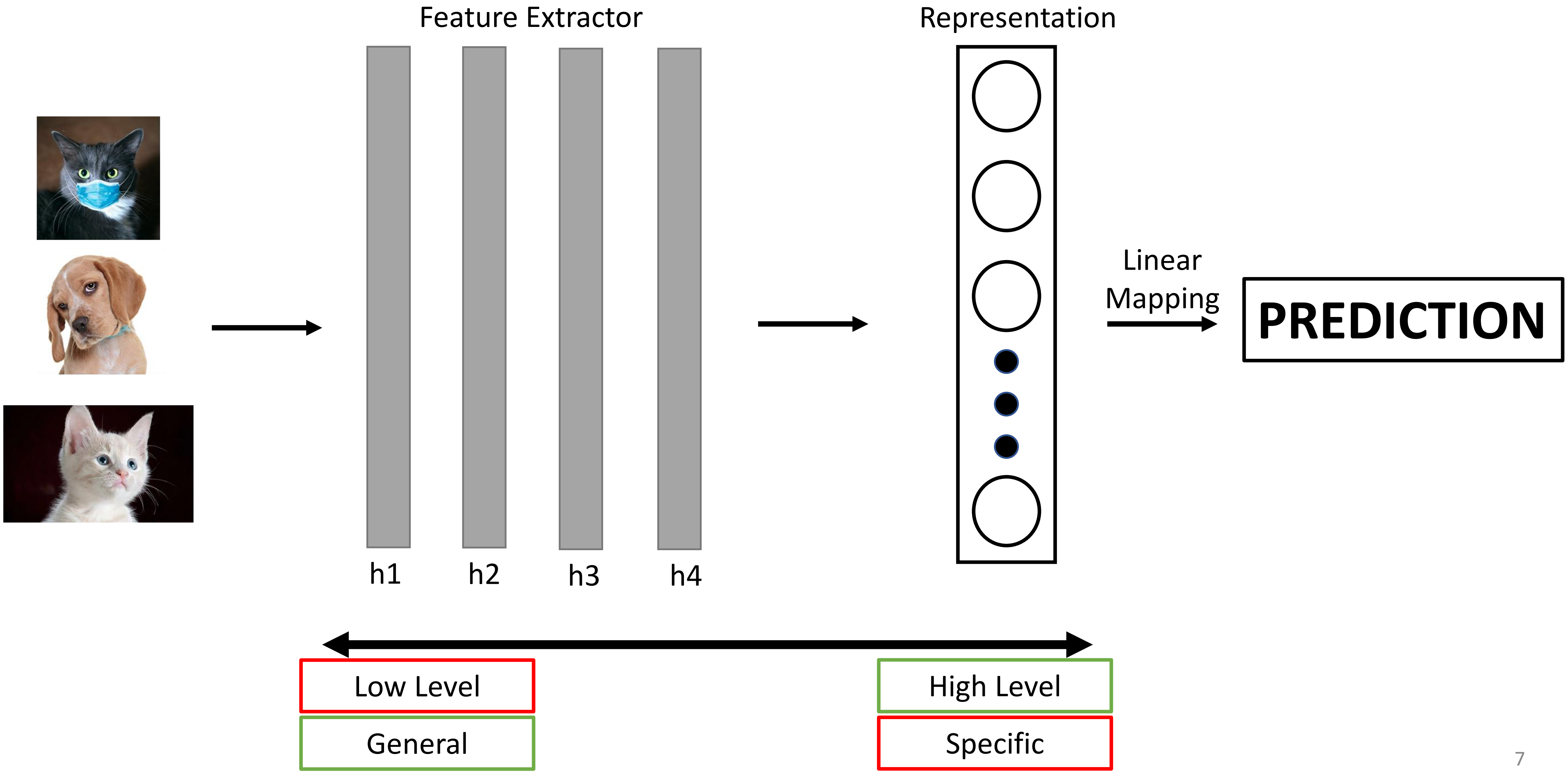
What is a Representation?



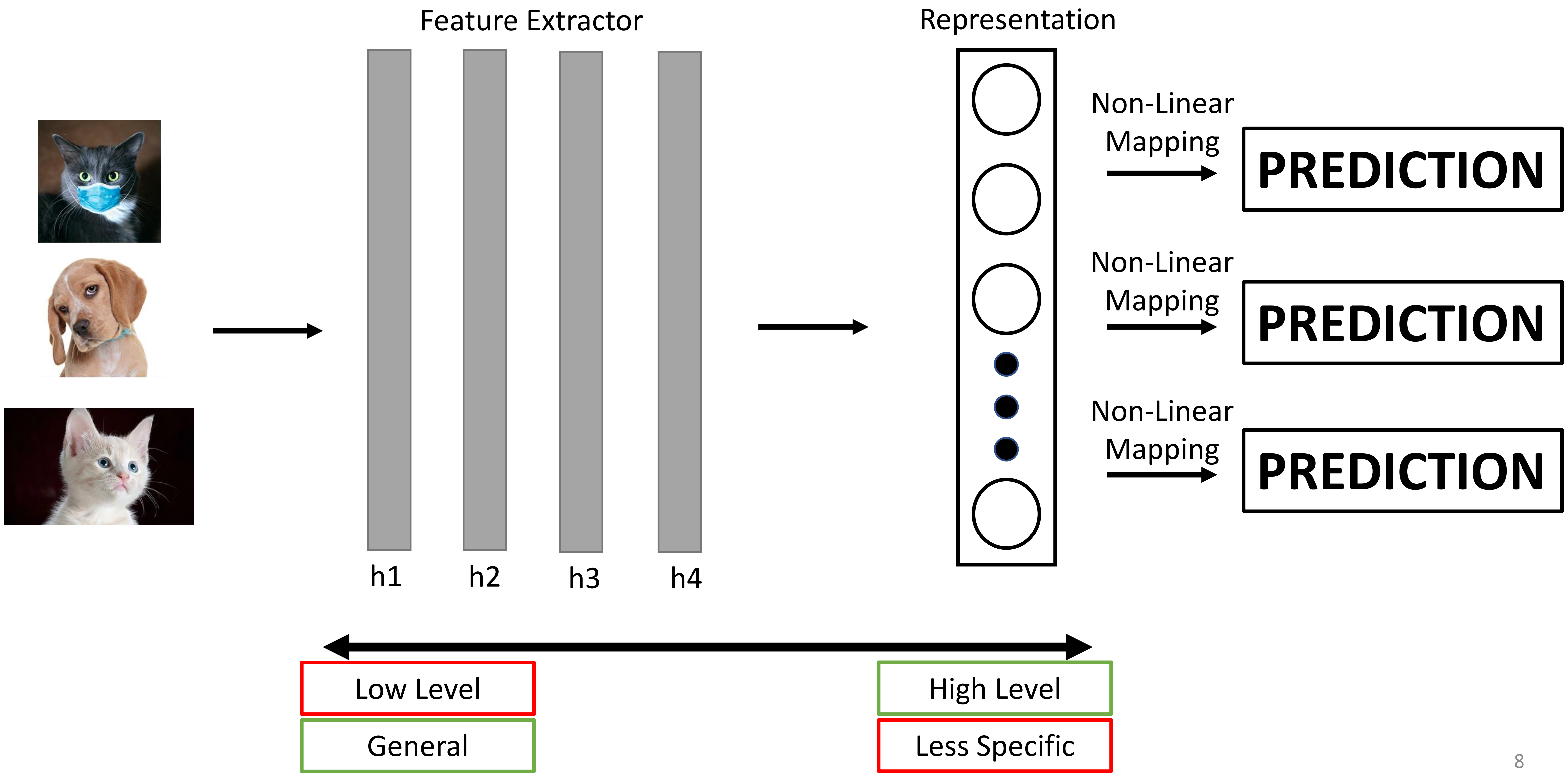
What is a Representation?



How to extract good features?

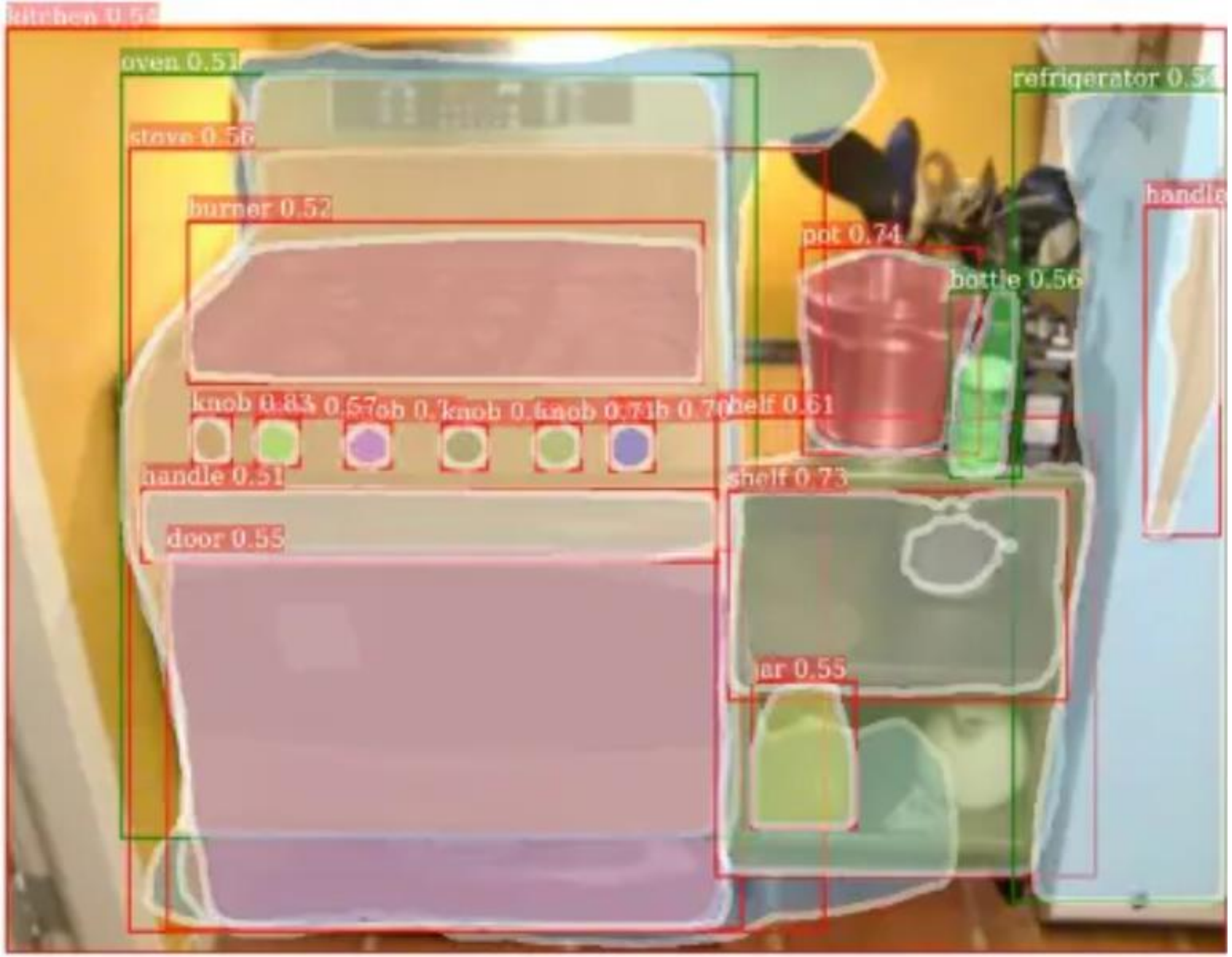


How to extract better features?



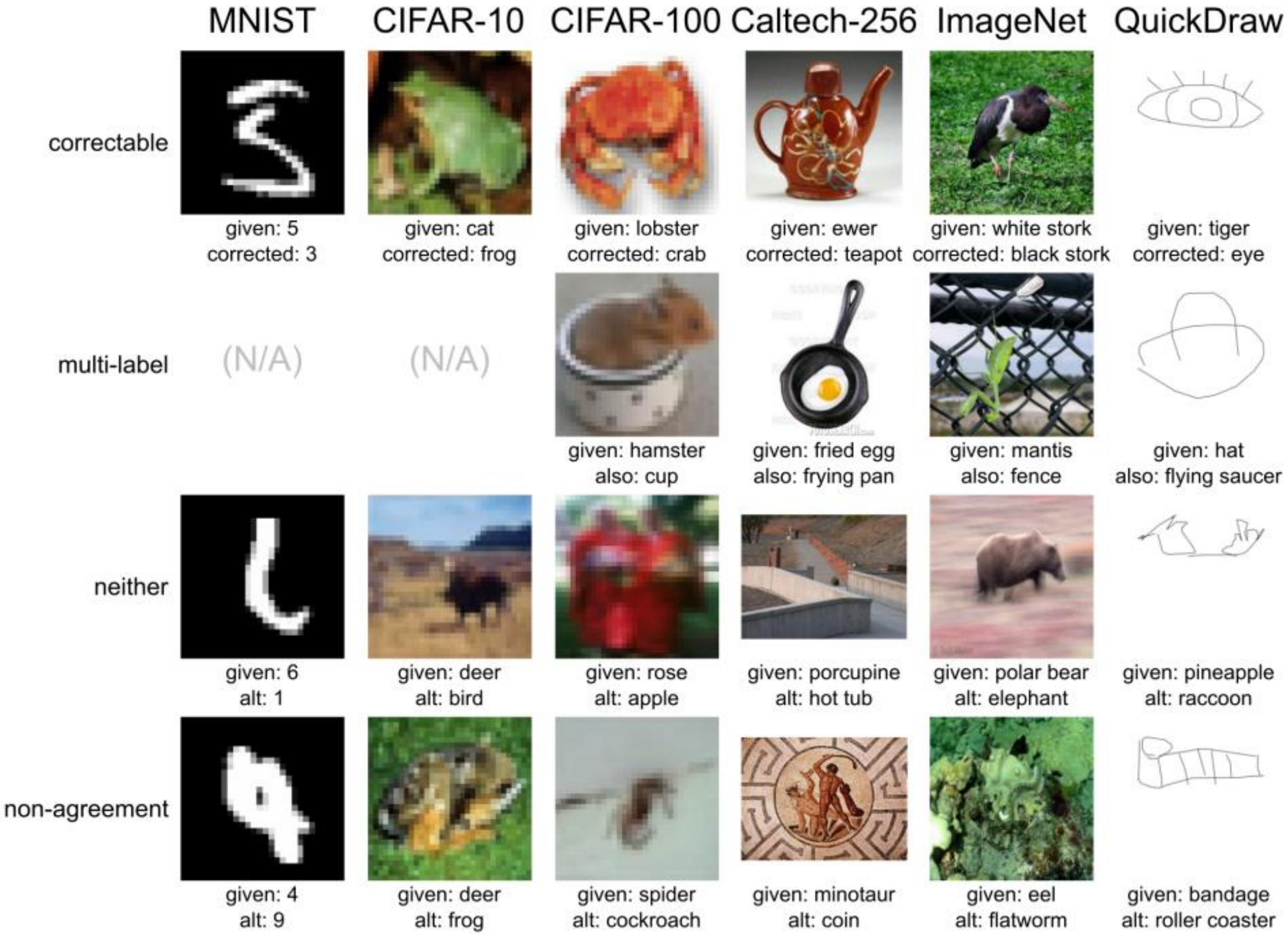
What avoid supervision?

- Labels are expensive.



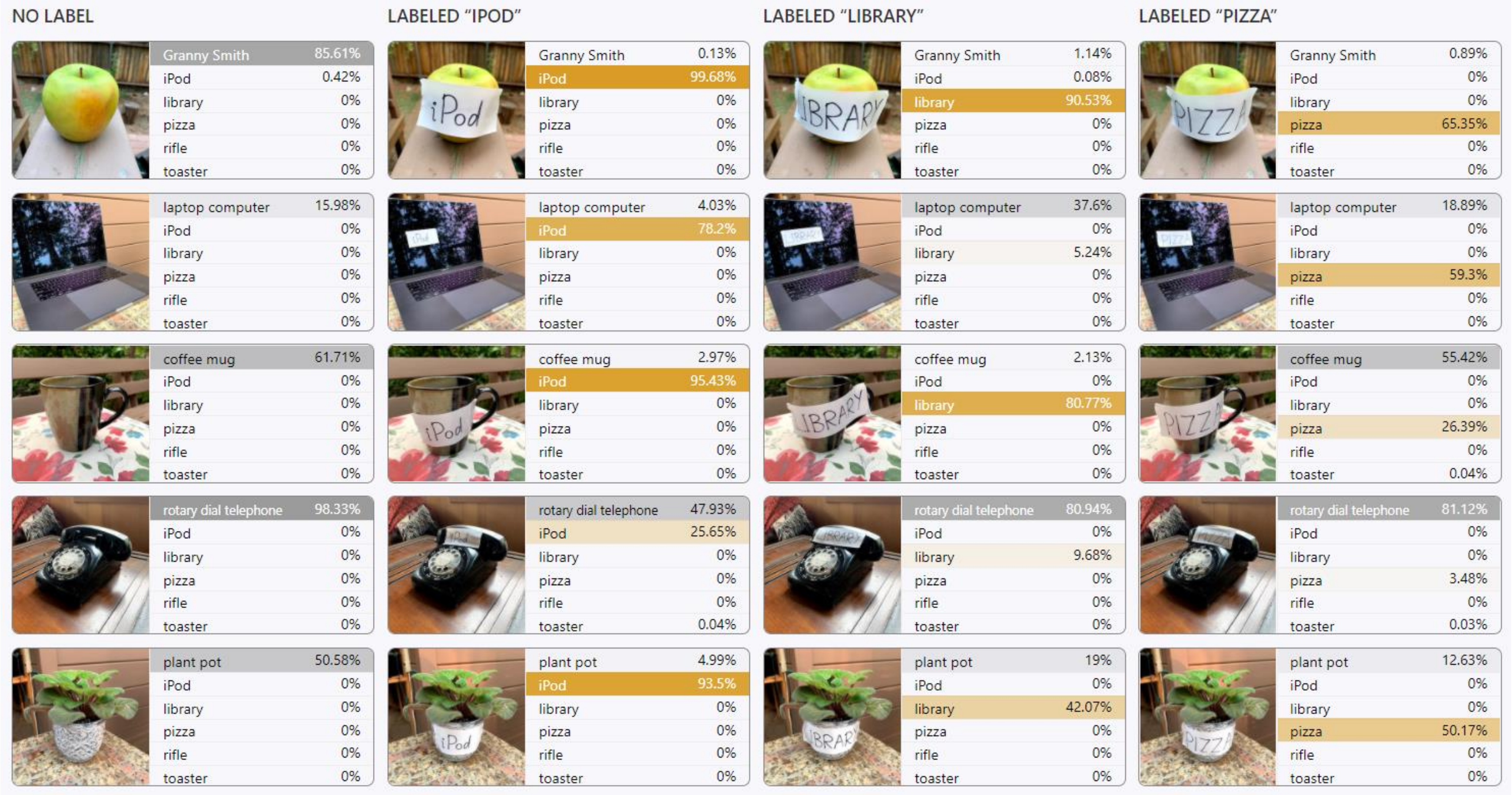
What avoid supervision?

- Labels are expensive.
- *Labels are inaccurate*



What avoid supervision?

- Labels are expensive.
- *Labels are inaccurate*

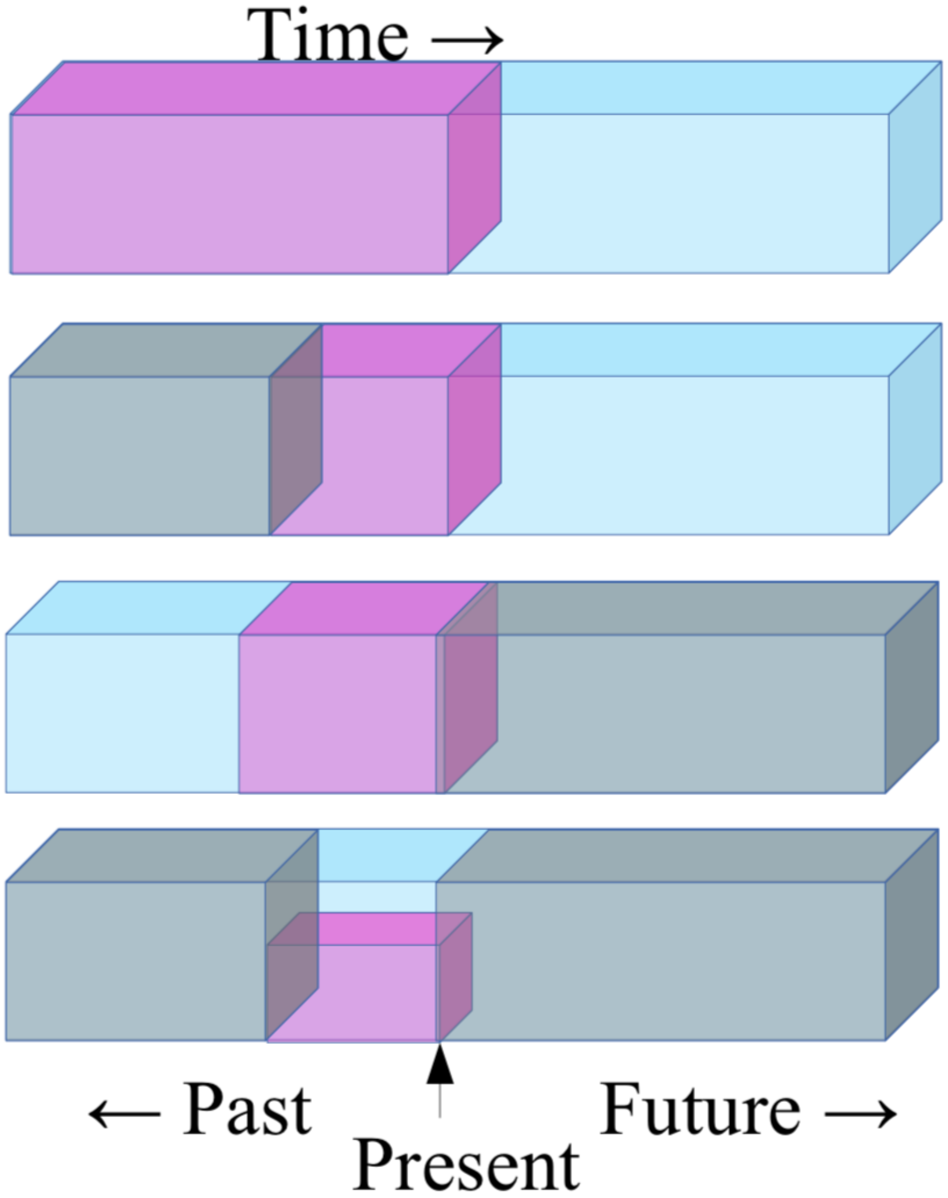


What avoid supervision?

- Labels are expensive.
- *Labels are inaccurate*
- *Labels encourage tailored and task-dependent features*
  - *e.g., Using human detector for tracking*

Self Supervised Representation Learning

- Past predicts the future
- Recent past predicts future
- Present predicts the past
- Visible predicts occluded



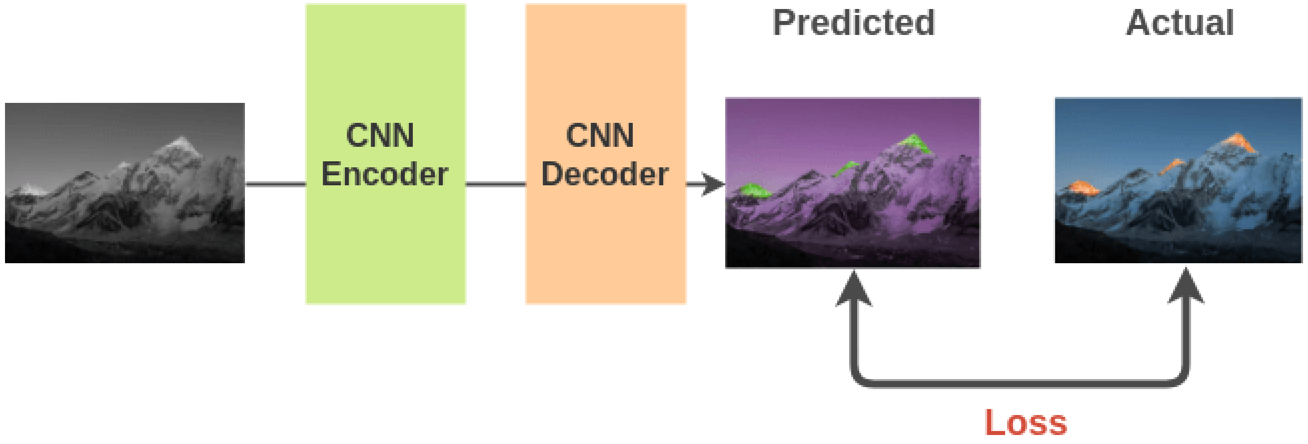
Adopted from Yann Lecun Presentation on self-supervised learning

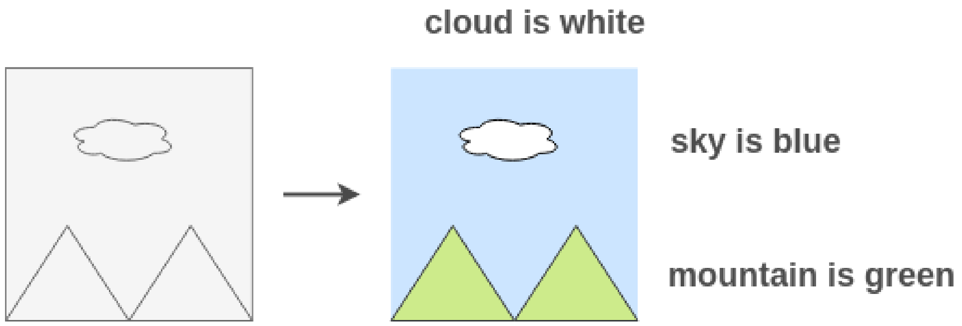
Self Supervised Representation Learning

Training Data Generation for Image Colorization

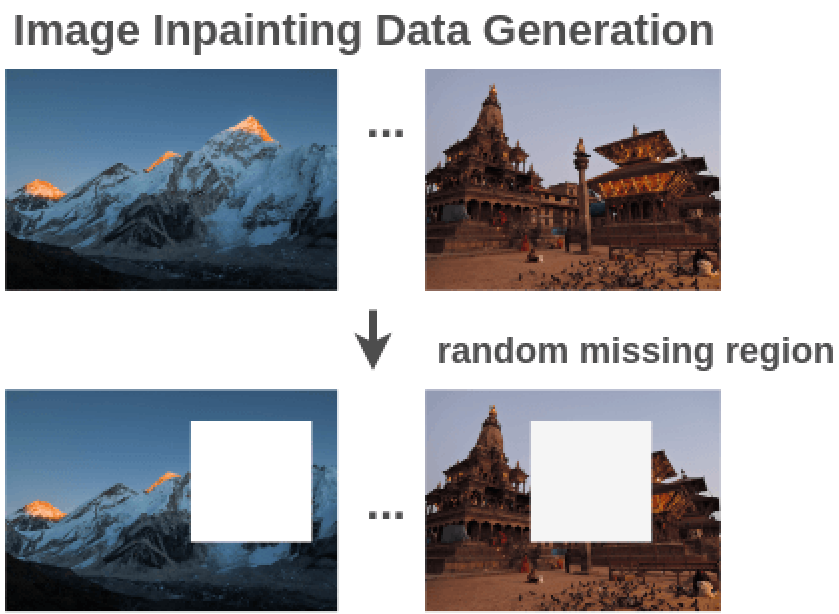
https://amitness.com/2020/02/illustrated-self-supervised-learning/

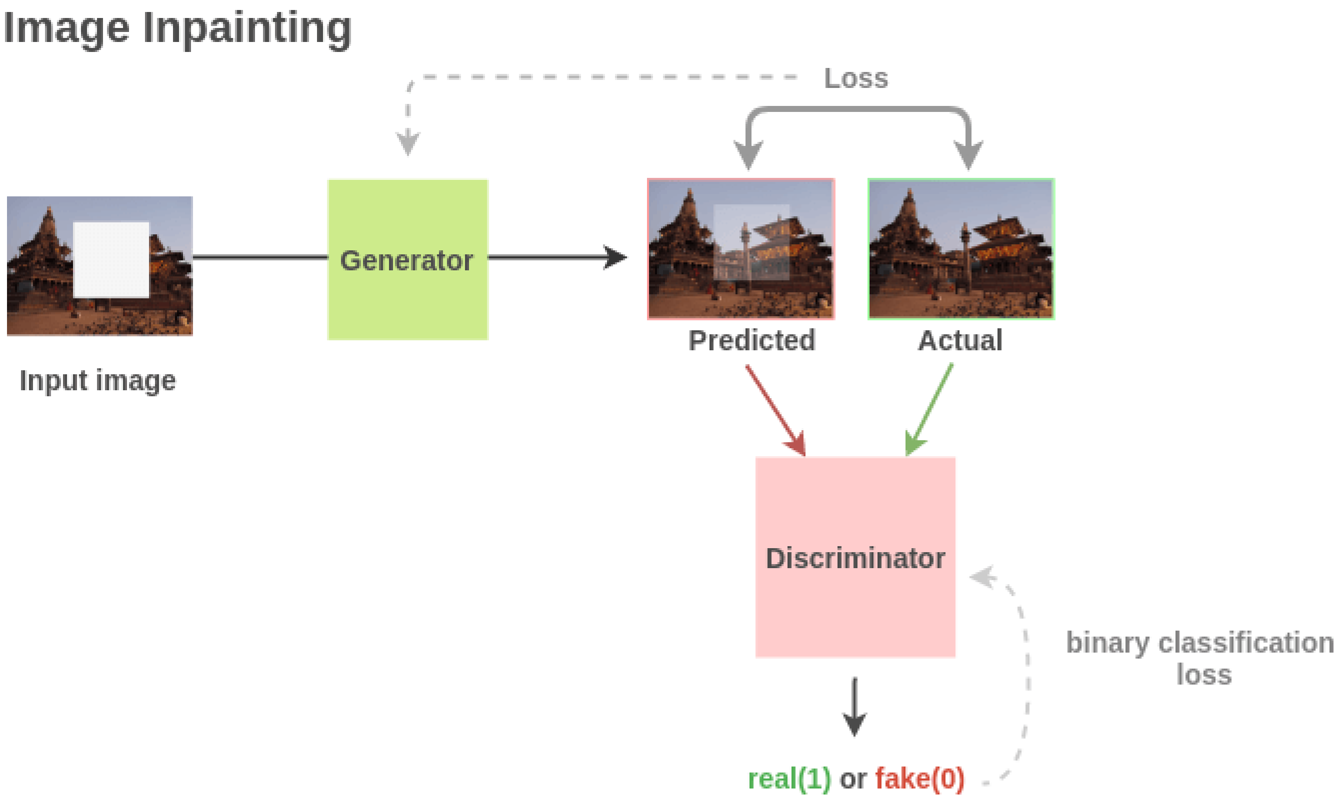
Image Colorization





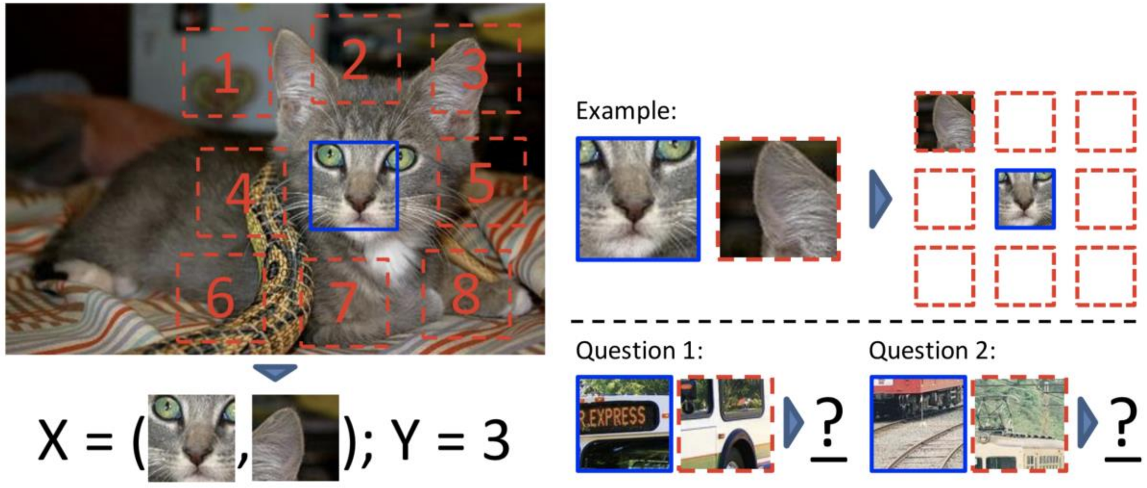
Self Supervised Representation Learning



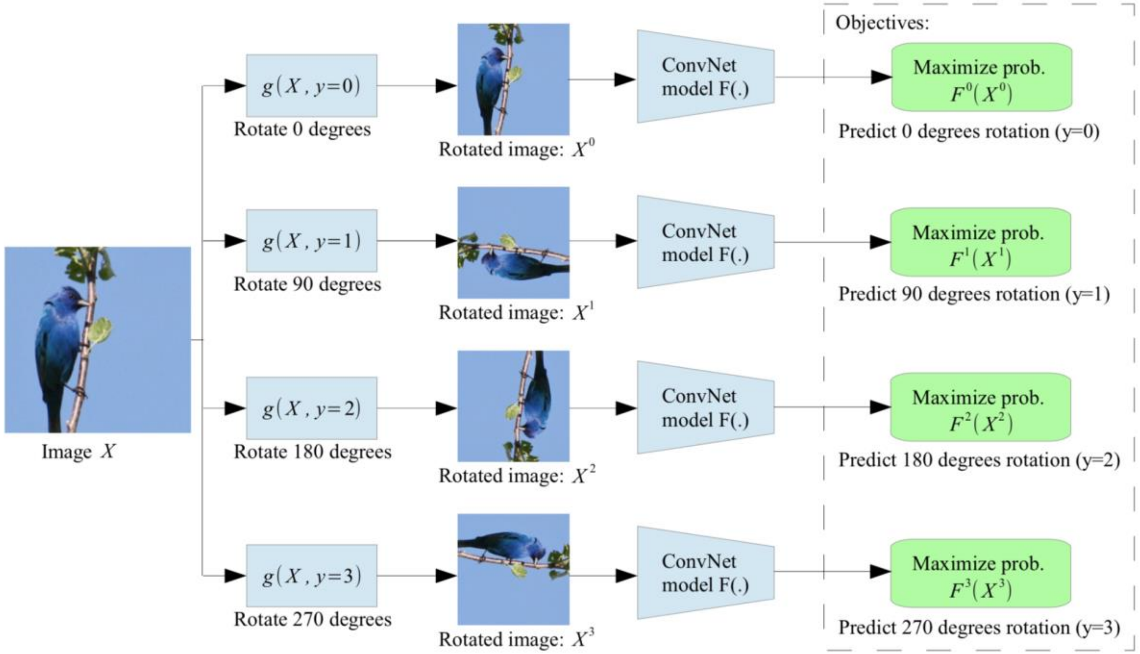


https://amitness.com/2020/02/illustrated-self-supervised-learning/

Self Supervised Representation Learning



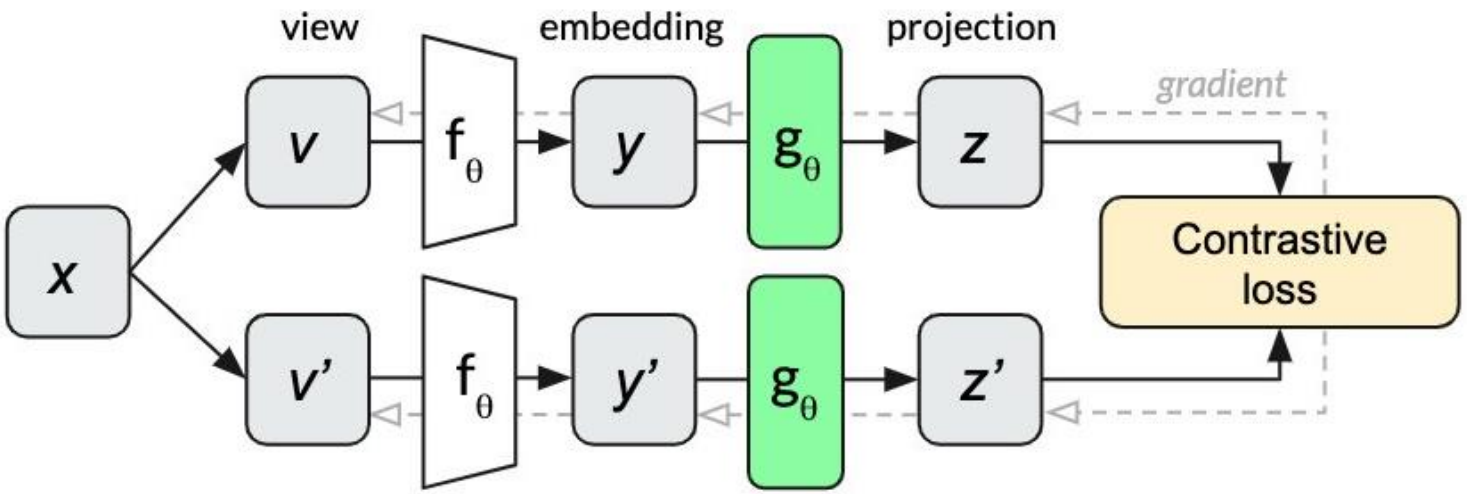
Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. "Unsupervised visual representation learning by context prediction." Proceedings of the IEEE international conference on computer vision. 2015.

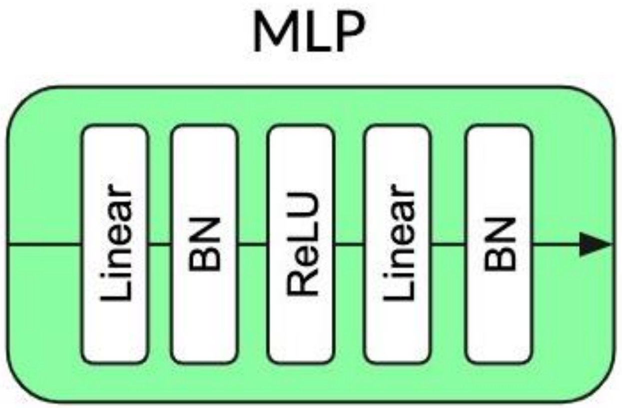


Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations." arXiv preprint arXiv:1803.07728 (2018).

Contrastive approaches

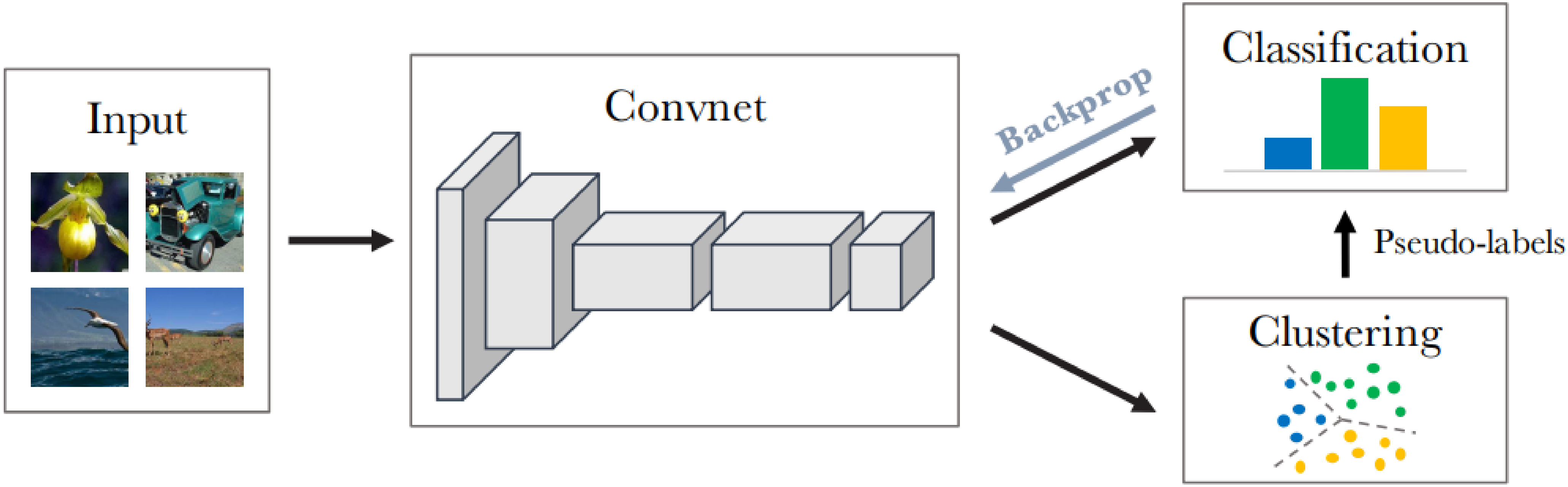
Contrastive approaches - SimCLR





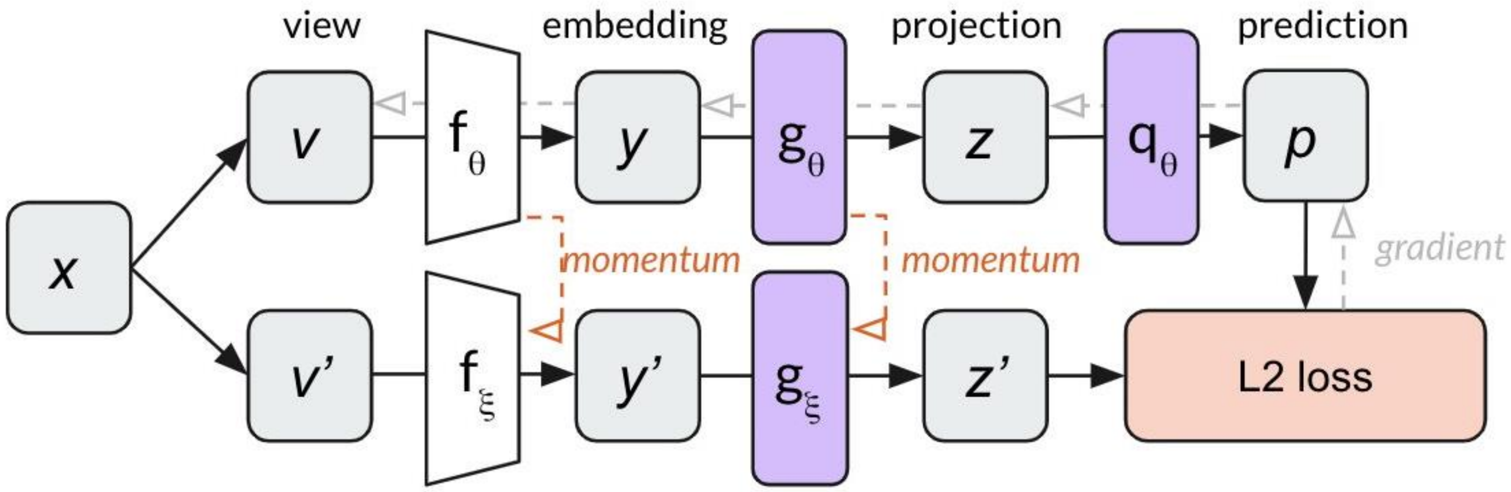
Chen et al. ICML 2020 [Image Source]

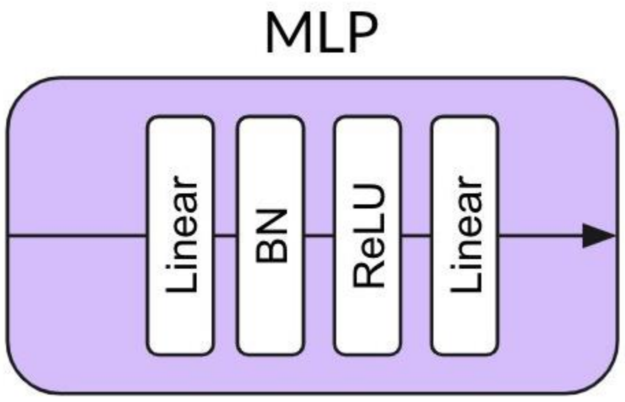
Clustering approaches



Caron et al. 2018

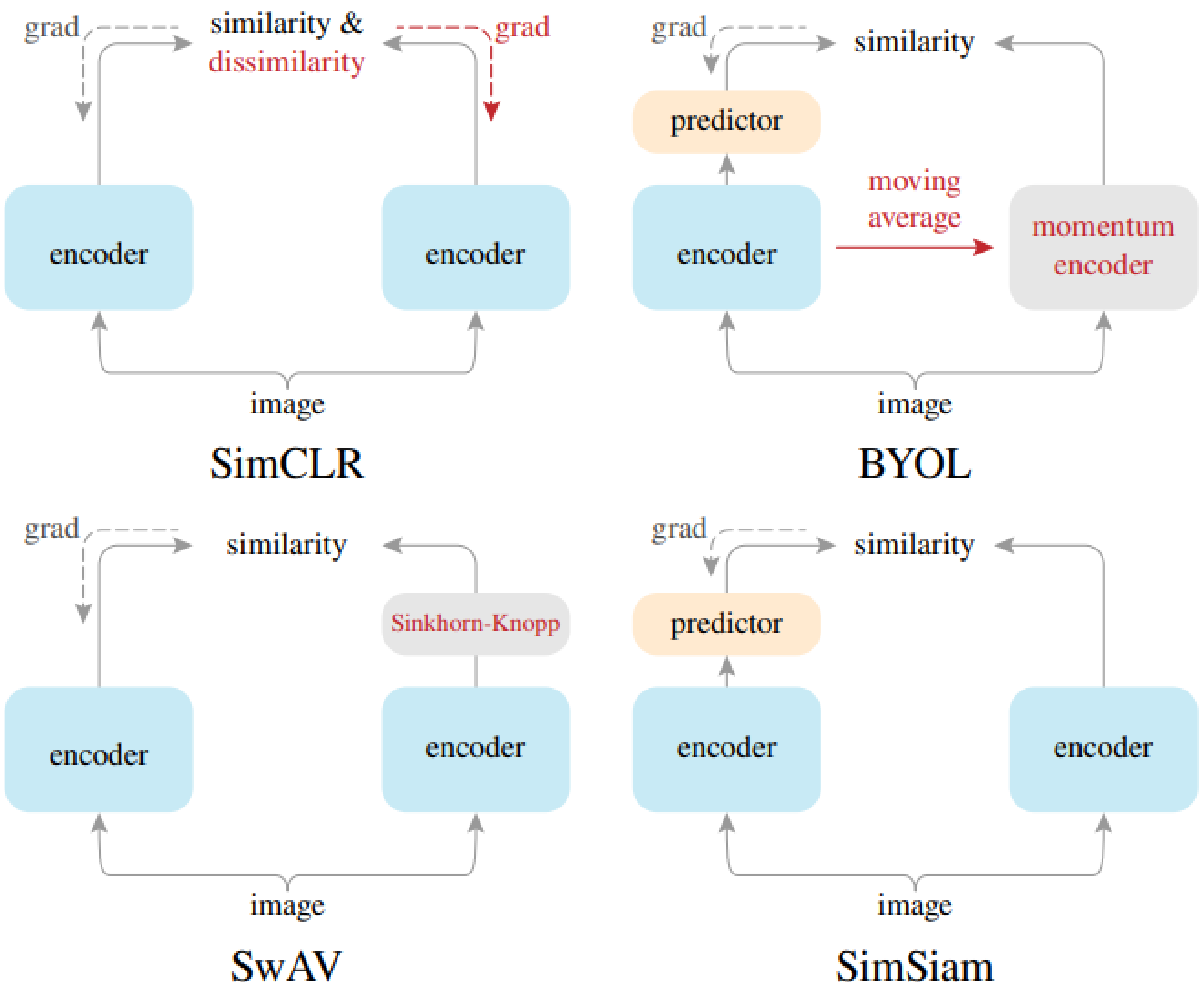
Momentum Encoder - BYOL





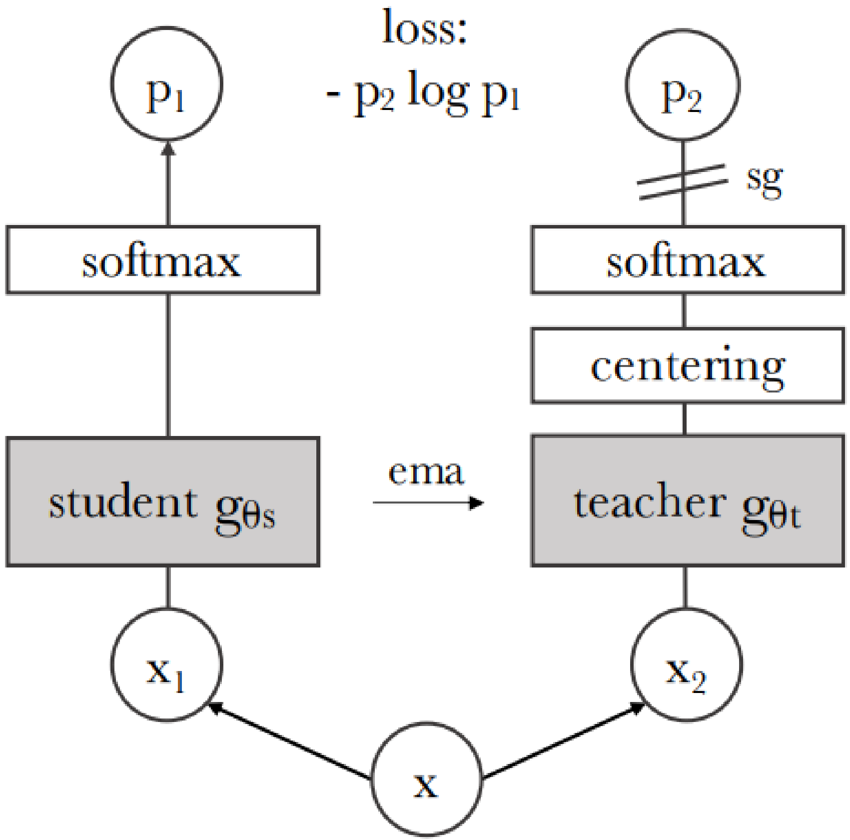
Grill et al. 2020 [Image Source]

Simple Siamese - SimSiam



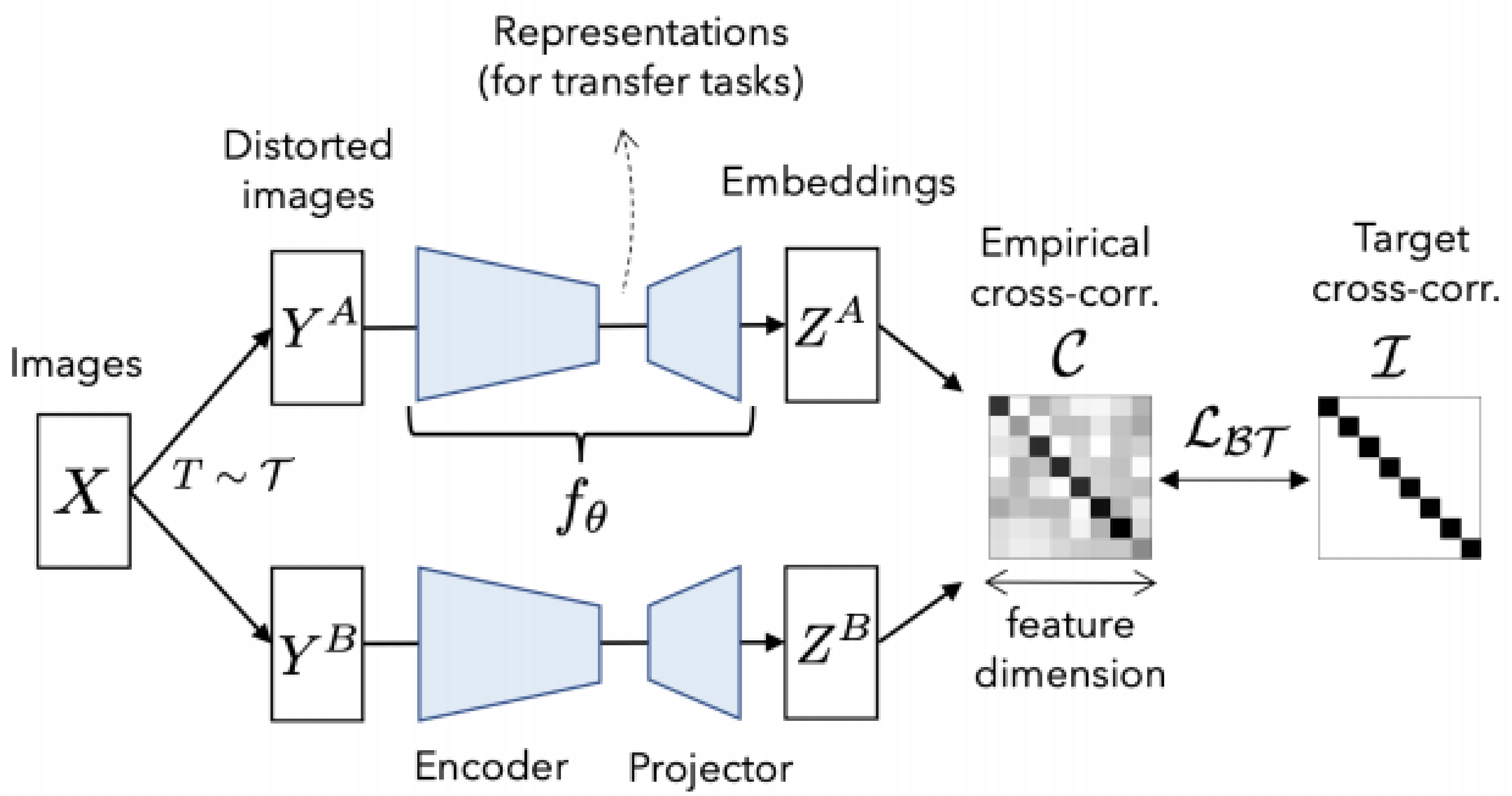
Chen et al. 2020

Momentum Encoder - DINO



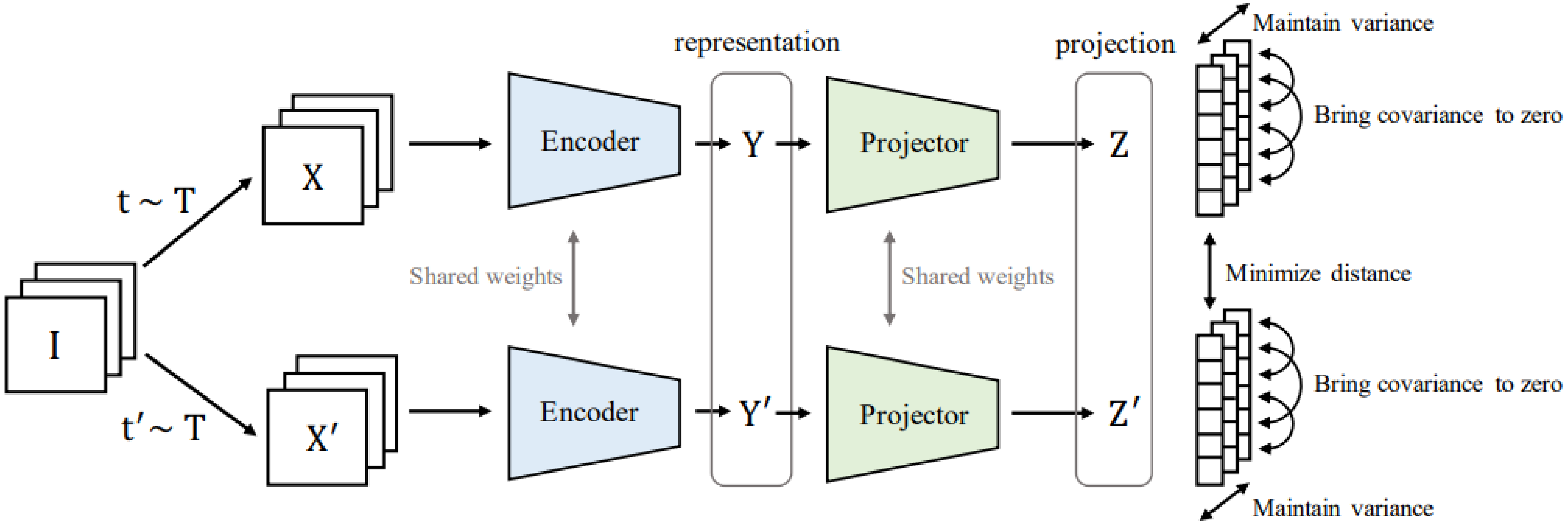
Caron et al. ICCV 2021

Symmetrical – Barlow Twins



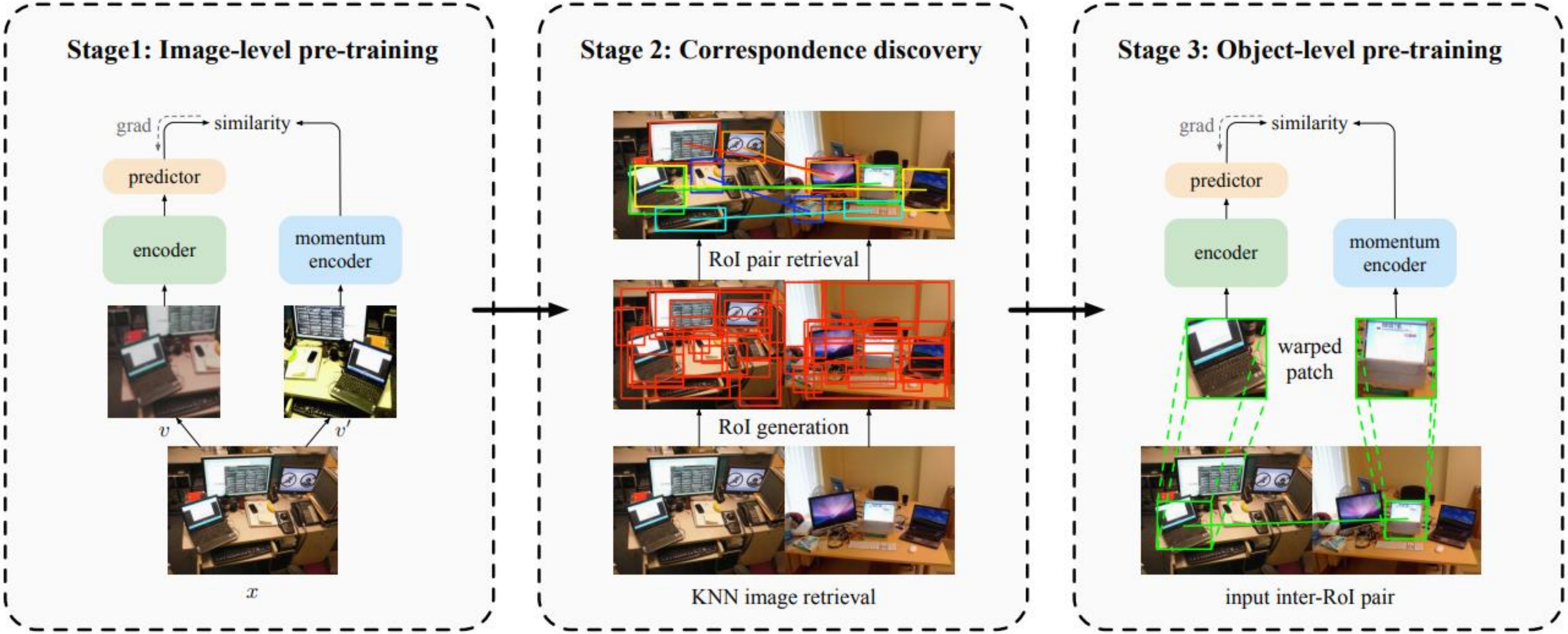
Zbontar et al. 2021

Symmetrical – VICReg



Bardes et al. 2021

Object Level - ORL



Xie et al. 2021

Object Level - ORL

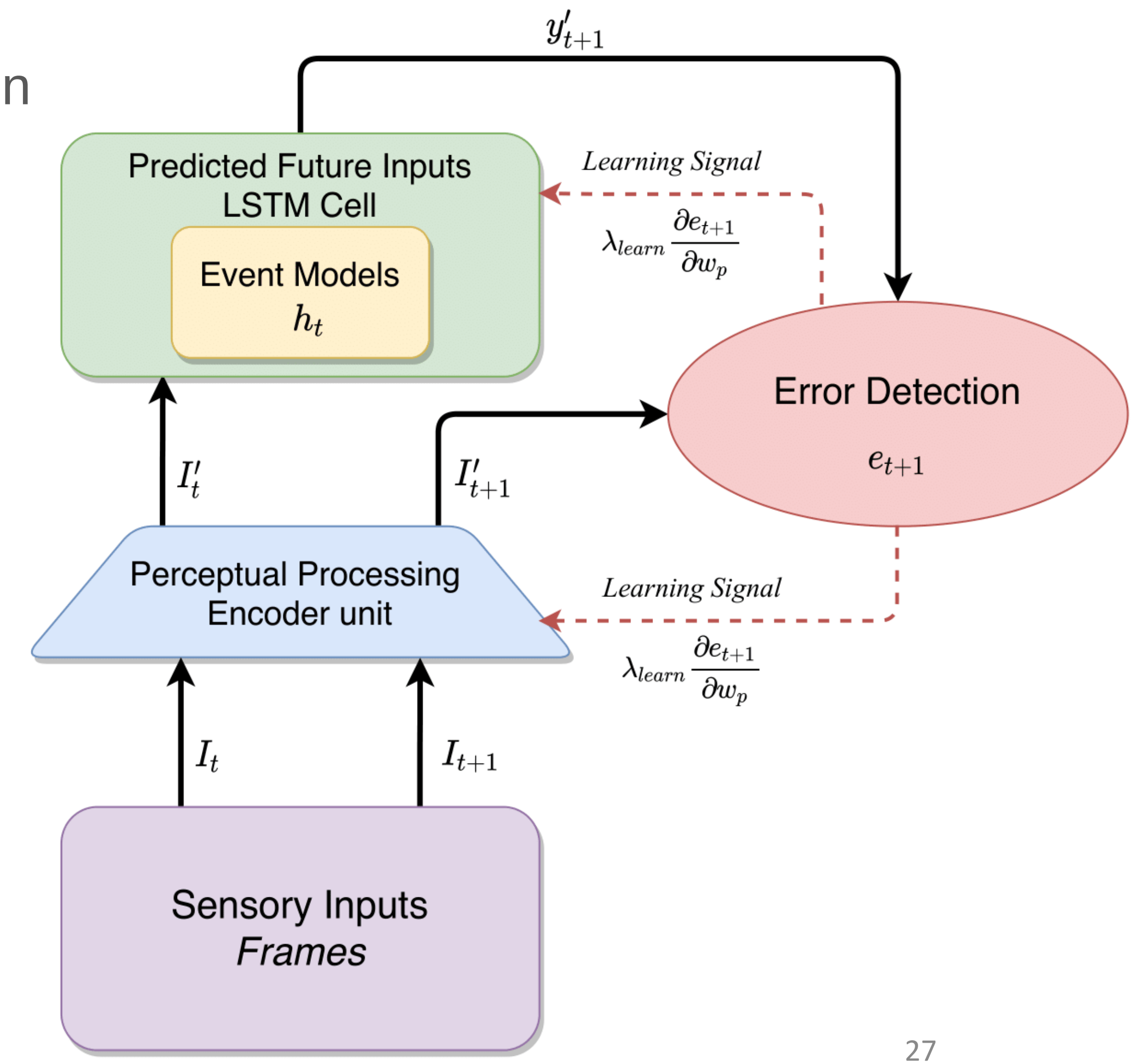
Xie et al. 2021

Self Supervised Event Segmentation

- LSTM cell for internal memory of event model.
- Error detection implemented as a low pass filter on prediction error running average.
- Gating signal triggered when error is above 1.5 times the running average.
- Adaptive learning controls learning rate

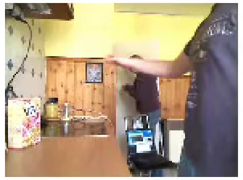
$$P_q(t) = P_q(t-1) + \frac{1}{n}(E_P(t) - P_q(t-1))$$

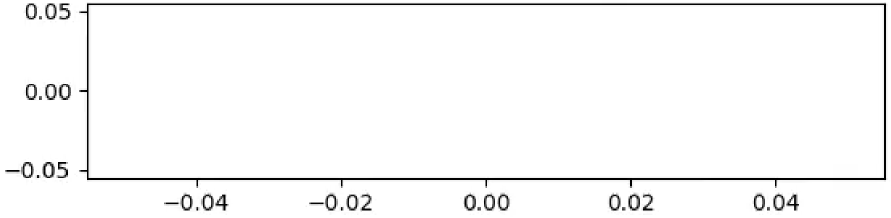
$$G(t) = \begin{cases} 1, & \frac{E_P(t)}{P_q(t-1)} > \psi_e \\ 0, & \text{otherwise} \end{cases}$$

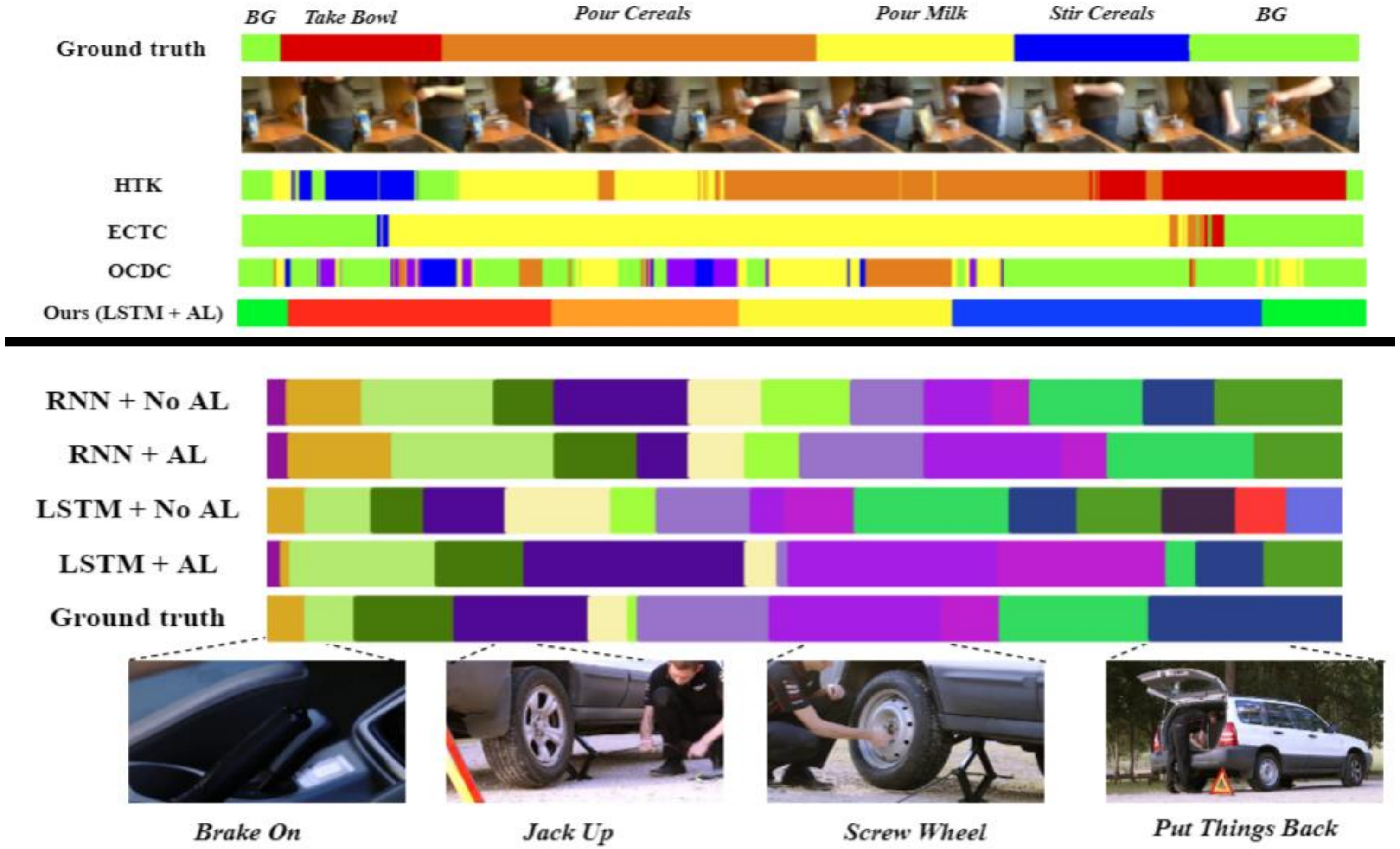


Aakur, S. N., & Sarkar, S. (2019). A perceptual prediction framework for self supervised event segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1197-1206).

Self Supervised Event Segmentation



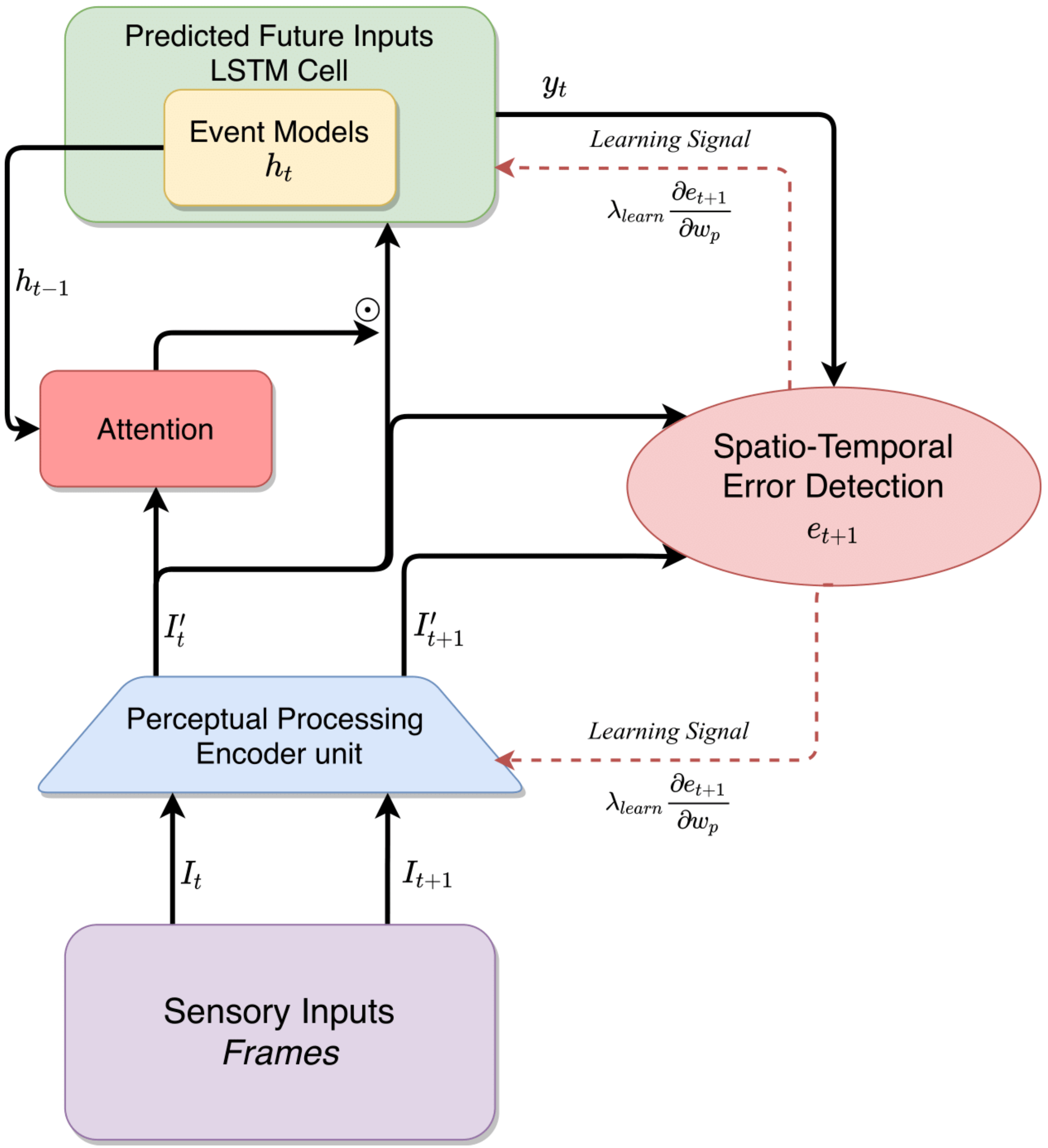




Wildlife Extended Videos

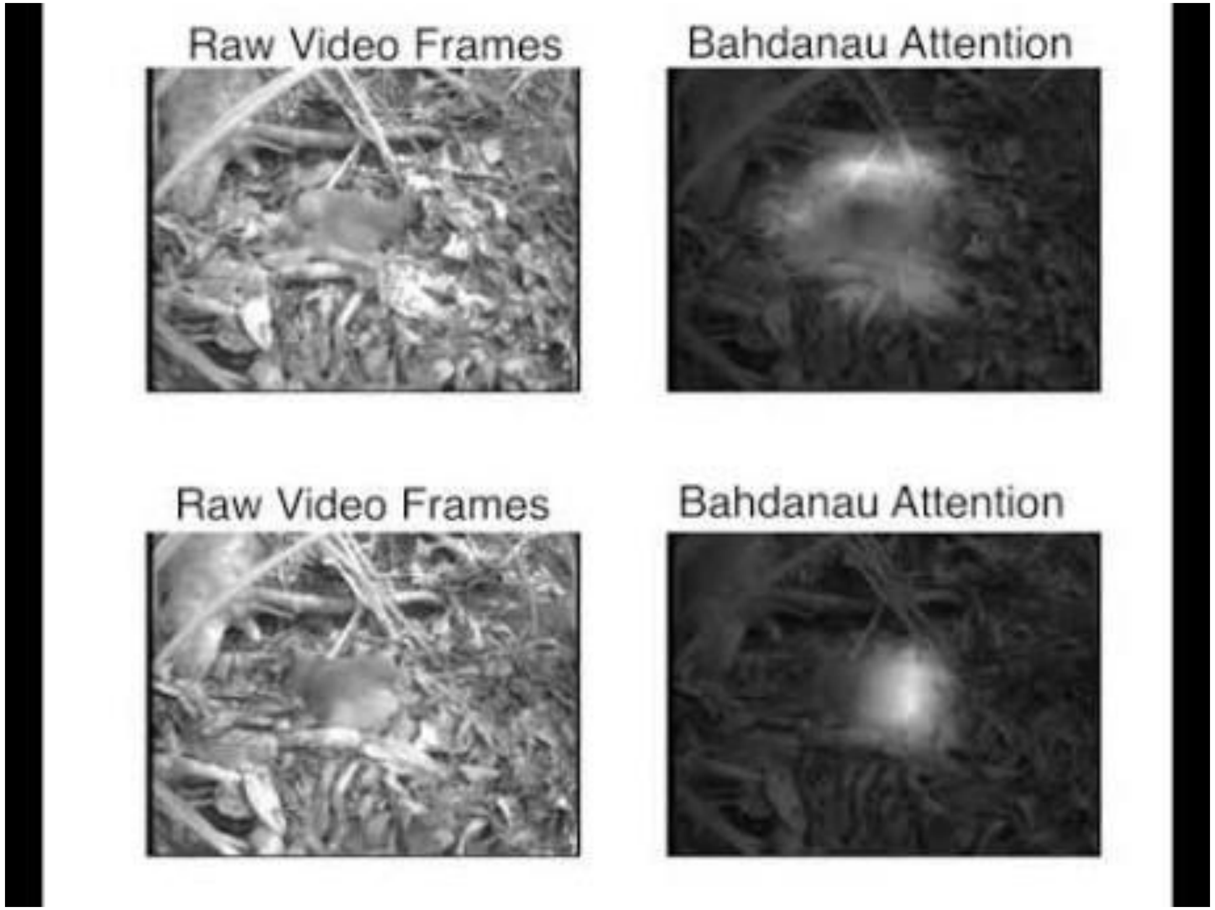
- Bahdanau attention is used to visualize the location of the bird.
- Motion-weighted loss is used instead of pure prediction loss.

$$e_t = ||(I'_{t+1} - y'_t)^{\odot 2} \odot (I'_{t+1} - I'_t)^{\odot 2}||^2$$

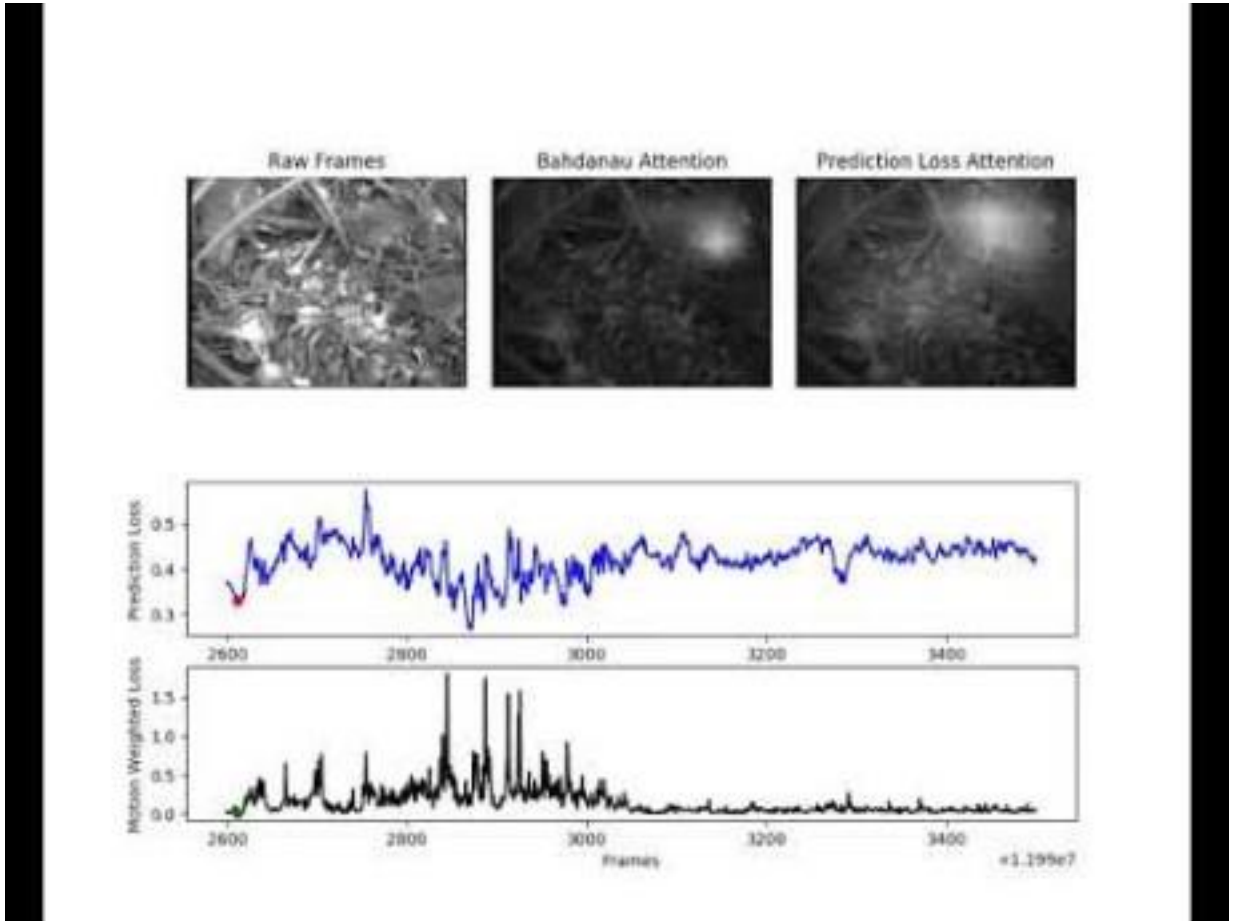


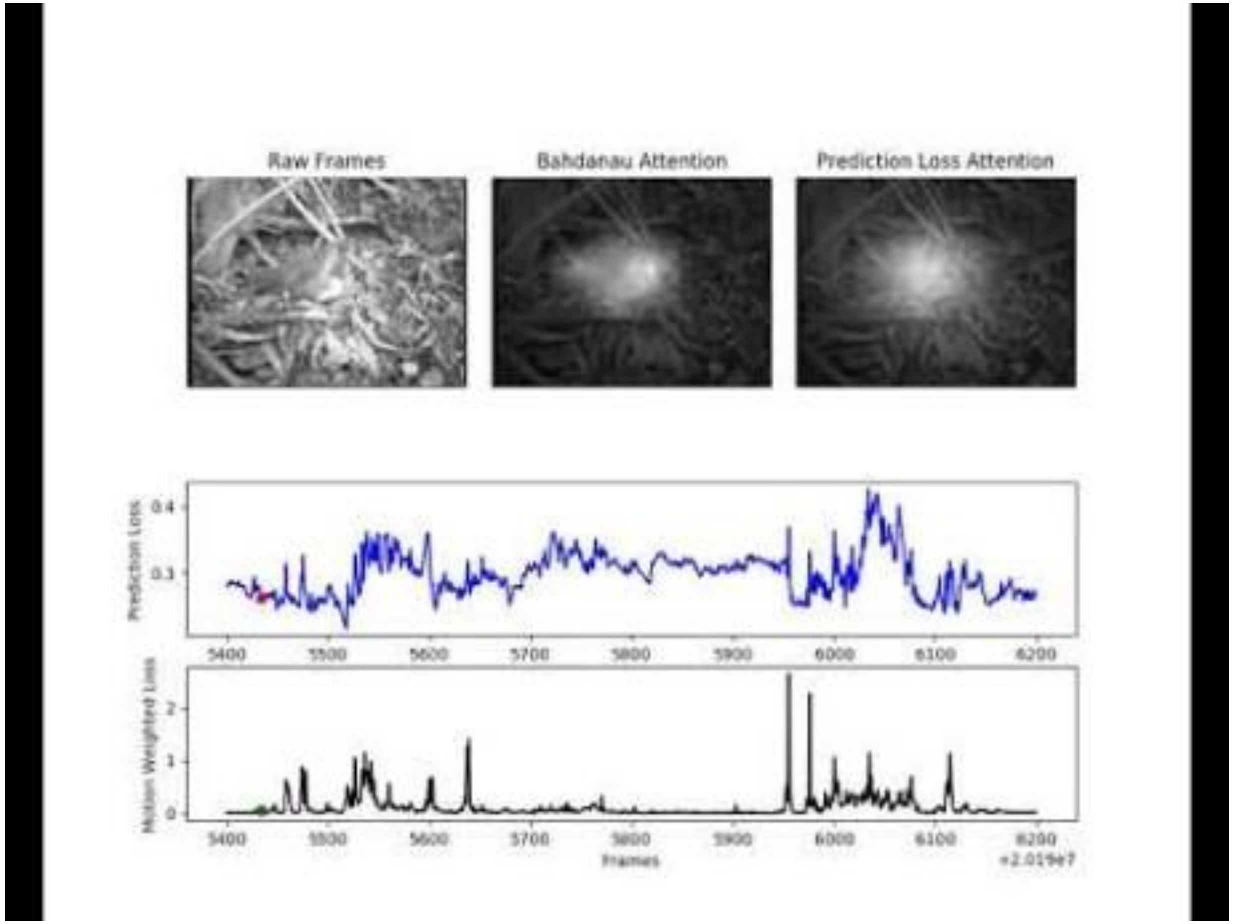
Mounir, R., Gula, R., Theuerkauf, J., & Sarkar, S. (2020). Temporal Event Segmentation using Attention-based Perceptual Prediction Model for Continual Learning. *arXiv preprint arXiv:2005.02463*.

Wildlife Extended Videos

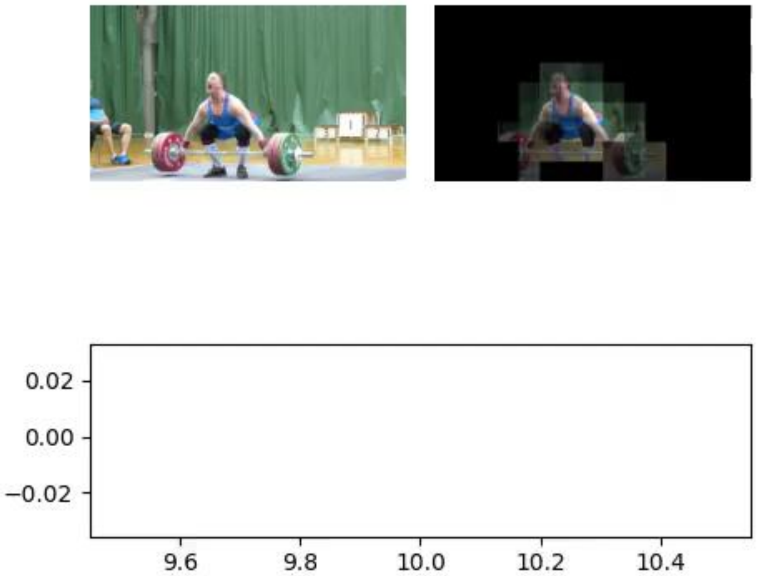


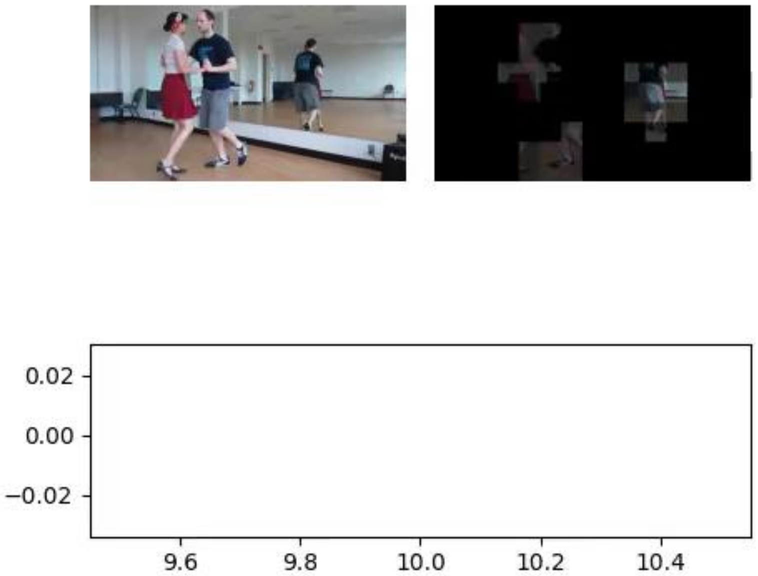
Wildlife Extended Videos

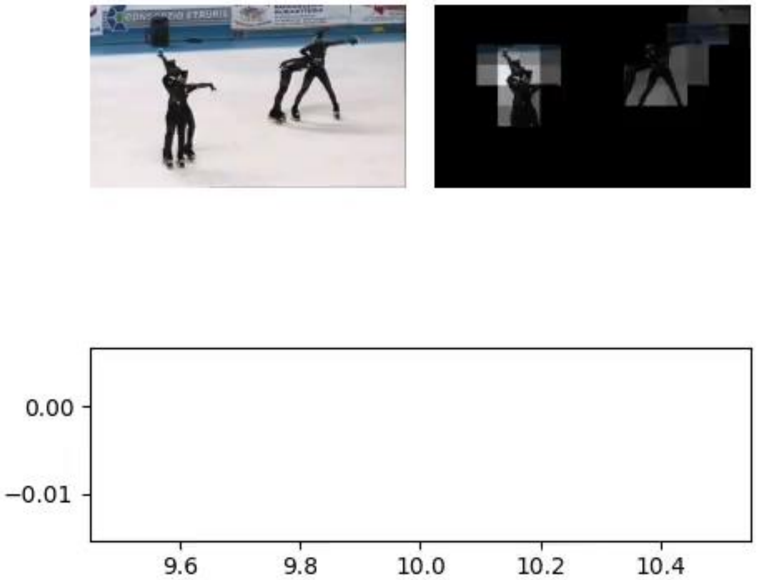




Other Domains







Thank you
Questions?

AI+X
INSTITUTE FOR ARTIFICIAL INTELLIGENCE

UNIVERSITY OF SOUTH FLORIDA