Recitation 12: Quiz 3 Review

Alex Singh and Robin Schmucker

Agenda

- Intelligent Exploration
- Offline RL
- Sim2Real
- Visual Imitation Learning
- Self-supervised Visual Learning

Intelligent Exploration

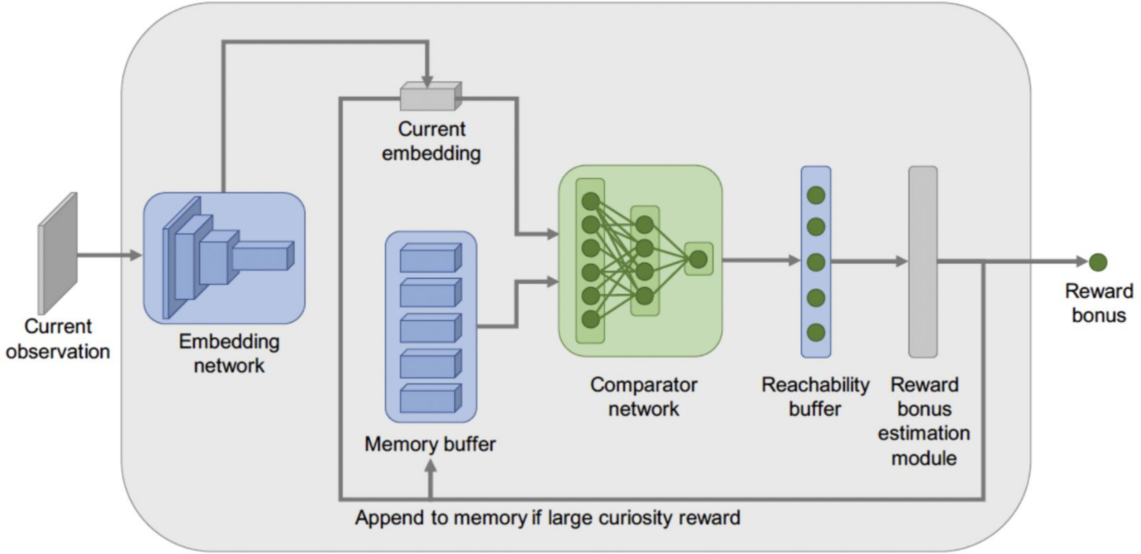
- Extrinsic Motivation
    - External reward, problem: sparse
- Intrinsic Motivation
    - Motivated by curiosity, enjoyability, etc.
    - Task independent, general, no supervision
- How to frame intrinsic motivation mathematically?
    - Q function ensembles, visit counts, reachability, etc.

Model Prediction Error as Intrinsic Motivation

- Add exploration bonus to states that will cause transition model to fail
- How to formulate exploration bonus?
  - Predict entire observation?
  - Predict latent state?
- Limitations of prediction error
  - Noisy TV

Curiosity Through Reachability

- Store non-parametric memory structure of past image embeddings
-



Offline RL

- Great summary here: https://arxiv.org/abs/2005.01643
- How can we extract effective policies from previously collected data, without additional experience collection?
  - Would facilitate usage of large datasets collected under some different policy
- How does this differ from the “off-policy” algorithms we have seen before?

Extrapolation Error and Batch-Constrained RL

- Q function on fixed experience has bad estimates on actions not in buffer
    - Leads to poor Q estimates
- Solution?
    - Only traverse transitions contained in batch
    -

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(r + \gamma \max_{a' \text{ s.t. } (s',a') \in \mathcal{B}} Q(s',a')).$$

BCQ

- Train model to generate actions that are contained within the batch
- Four key components
  - cVAE to generate actions conditioned on state
  - Perturbation model to add diversity to actions
  - Two Q networks as in clipped double Q learning
- Go through paper/algorithm?

IRIS

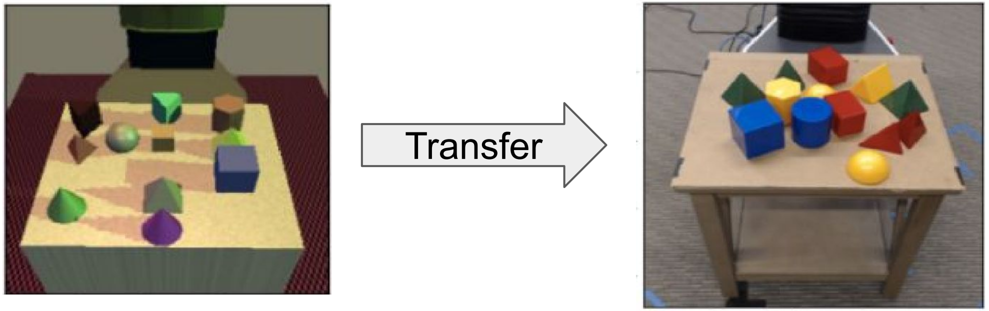
- High-level Idea: Train High-Level Goal Proposal, Low-Level Controller
    - Low-level controller trained via imitation learning
    - High-level Goal Proposal trained as cVAE
    - Generate proposals for the low level controller to reach
- Go through paper/algorithm?

Sim2Real

- The very large number of samples required by many model-free RL algorithm is often only possible in simulations (simulator acts as “model”)

- **Idea**: We can train our policies in simulation then transfer to the real world

Simulation Real World



Sim2Real

Idea: We can train our policies in simulation then transfer to the real world

Pros:

- We can afford many samples
- Exploration is safe
- Avoids wear and tear on robot
- Can explore with different configs

Cons:

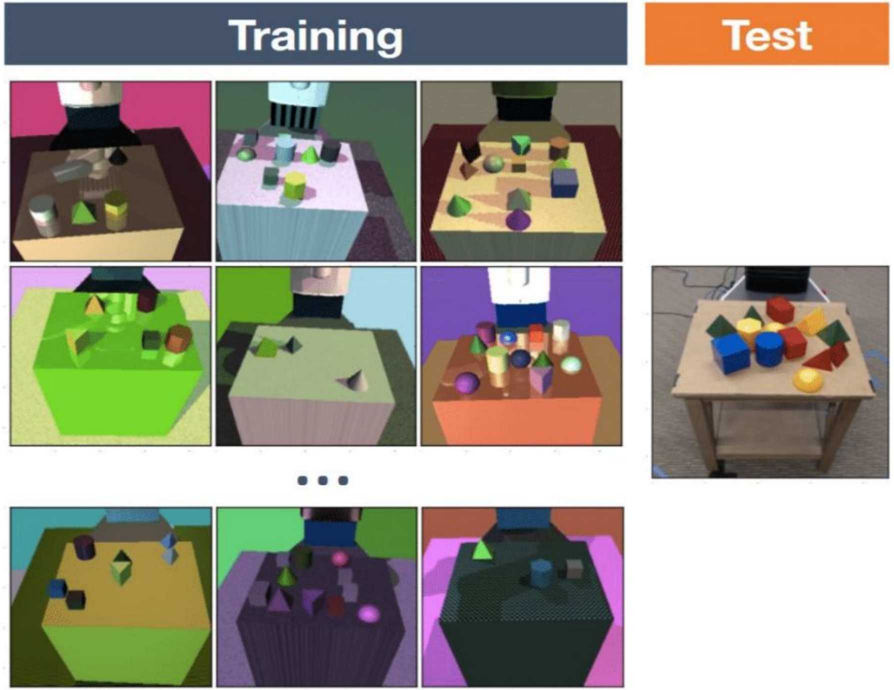
- Creating simulators is expensive
- Discrepancy in observations
- Discrepancy in dynamics

Result: Policies learnt in simulation usually do not transfer well…

Sim2Real: Domain Adaptation

- **Idea**: Sample from a large set of simulation environments by randomizing the simulator parameterization (dynamics, visuals)

- By learning from many different environments we hope to improve transfer performance



Sim2Real: Automatic Domain Randomization

Algorithm 1 ADR

Require: ϕ^0 $\triangleright$ Initial parameter values
Require: $\{D_i^L, D_i^H\}_{i=1}^d$ $\triangleright$ Performance data buffers
Require: m, t_L, t_H, where $t_L < t_H$ $\triangleright$ Thresholds
Require: Δ $\triangleright$ Update step size

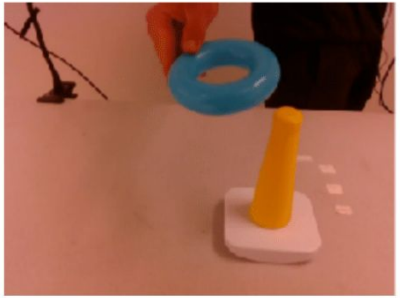
$\phi \leftarrow \phi^0$
repeat
  $\lambda \sim P_\phi$ — sample environment config
  $i \sim U\{1, \ldots, d\}, x \sim U(0,1)$
  if $x < 0.5$ **then**
    $D_i \leftarrow D_i^L, \lambda_i \leftarrow \phi_i^L$ $\triangleright$ Select the lower bound in "boundary sampling"
  else
    $D_i \leftarrow D_i^H, \lambda_i \leftarrow \phi_i^H$ $\triangleright$ Select the higher bound in "boundary sampling"
  end if
  $p \leftarrow \text{EVALUATEPERFORMANCE}(\lambda)$ $\triangleright$ Collect model performance on environment parameterized by λ
  $D_i \leftarrow D_i \cup \{p\}$ $\triangleright$ Add performance to buffer for λ_i, which was boundary sampled
  if $\text{LENGTH}(D_i) \geq m$ **then**
    $\bar{p} \leftarrow \text{AVERAGE}(D_i)$
    $\text{CLEAR}(D_i)$
    if $\bar{p} \geq t_H$ **then**
      $\phi_i \leftarrow \phi_i + \Delta$ — increase configuration space
    else if $\bar{p} \leq t_L$ **then**
      $\phi_i \leftarrow \phi_i - \Delta$
    end if
  end if
until training is complete

Resolves need for manual configuration

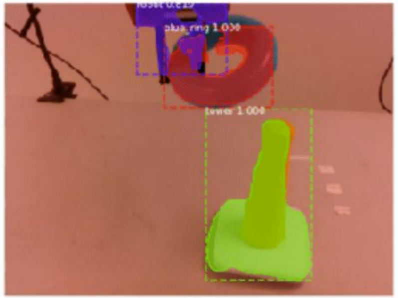
Visual Imitation Learning

- Learning skills by watching people or other agents performing the skill

human demonstration



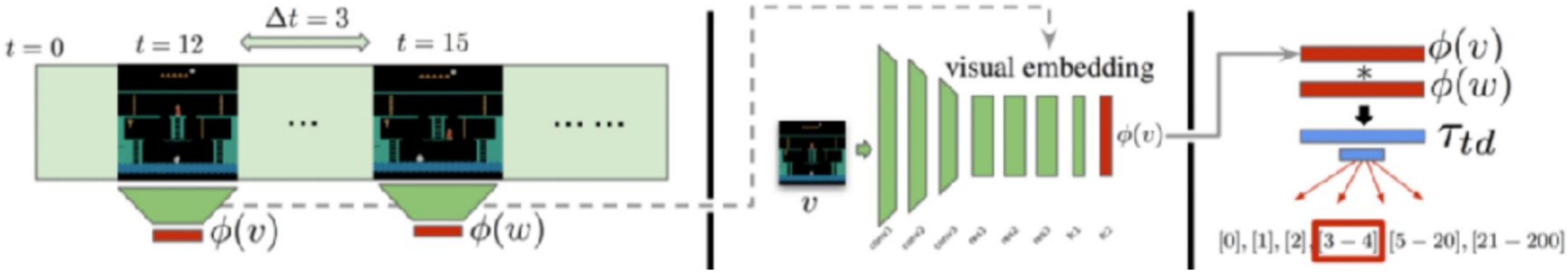
robot’s imitation



- Central difficulty in visual imitation is perceiving the world state: where are the objects, in which pose, what velocities, etc.
- We use Computer Vision to learn a suitable representation

Visual Imitation Learning

Paper: Playing exploration games by watching YouTube



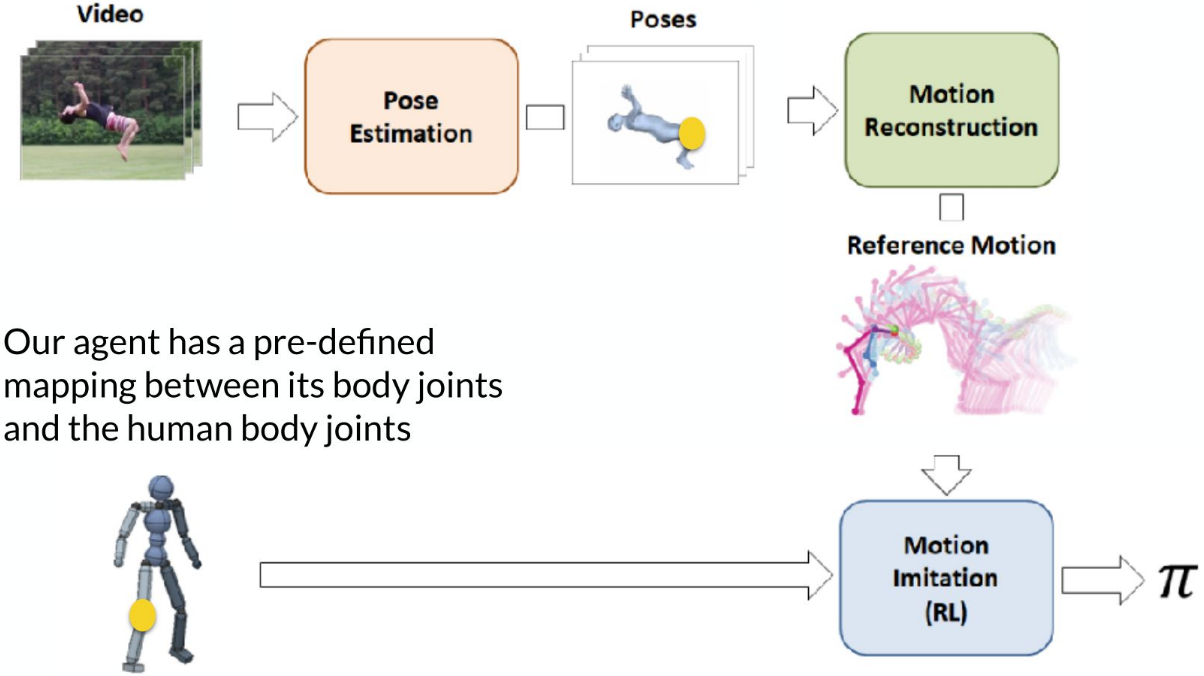
- Temporal distance classification: given two frames, clarify their temporal distance into one of k intervals, e.g., {[0],[1],[2],[3−4],[5−20],[21−200]}
- Given one video demo, use visual similarity encoded as frame embedding distance asimitation reward

Visual Imitation Learning

Paper: SFV: Reinforcement Learning of Physical Skills from Videos

Visual Imitation Learning

<u>Paper</u>: SFV: Reinforcement Learning of Physical Skills from Videos



Our agent has a pre-defined mapping between its body joints and the human body joints

Self-supervised Visual Learning

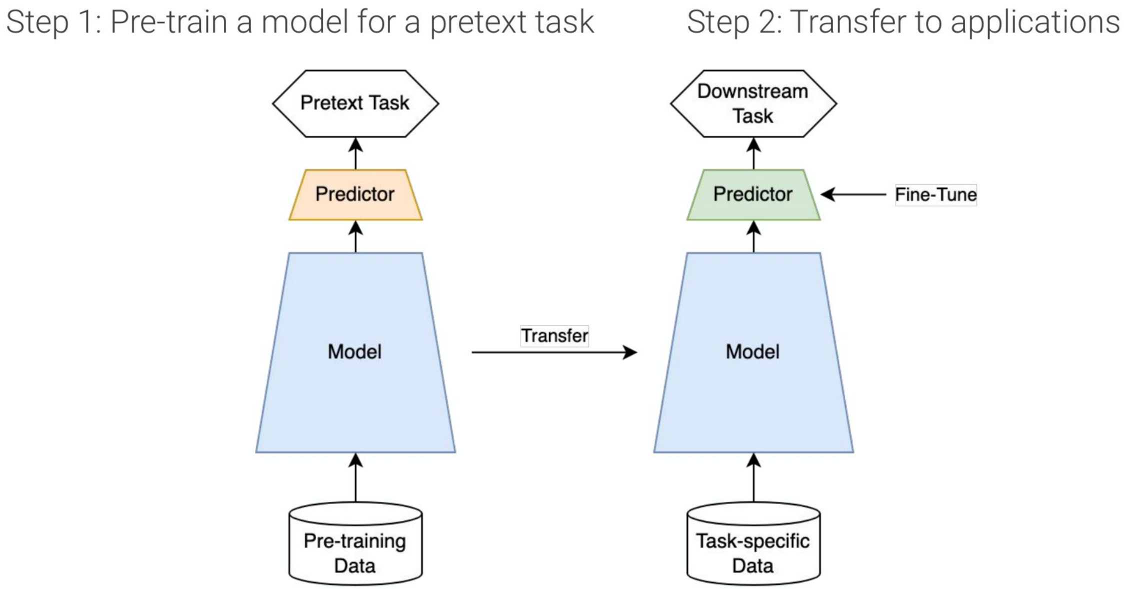
- Try to learn good representation from unlabelled data
- **Idea**: Construct supervised learning tasks out of unsupervised datasets. We call these tasks **pretext tasks**.

Why do we want to this?

- Data labeling is expensive and high-quality labeled datasets are limited
- Learning good representation makes it easier to transfer useful information to downstream tasks (few-shot, zero-shot learning)

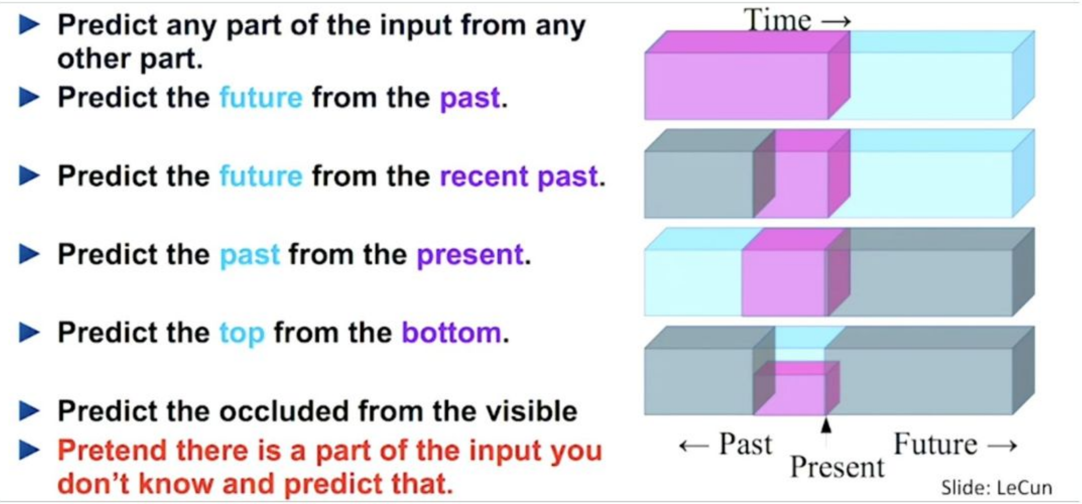
Self-supervised Visual Learning

Idea: Construct supervised learning tasks out of unsupervised datasets. We call these tasks **pretext tasks**.



Self-supervised Visual Learning

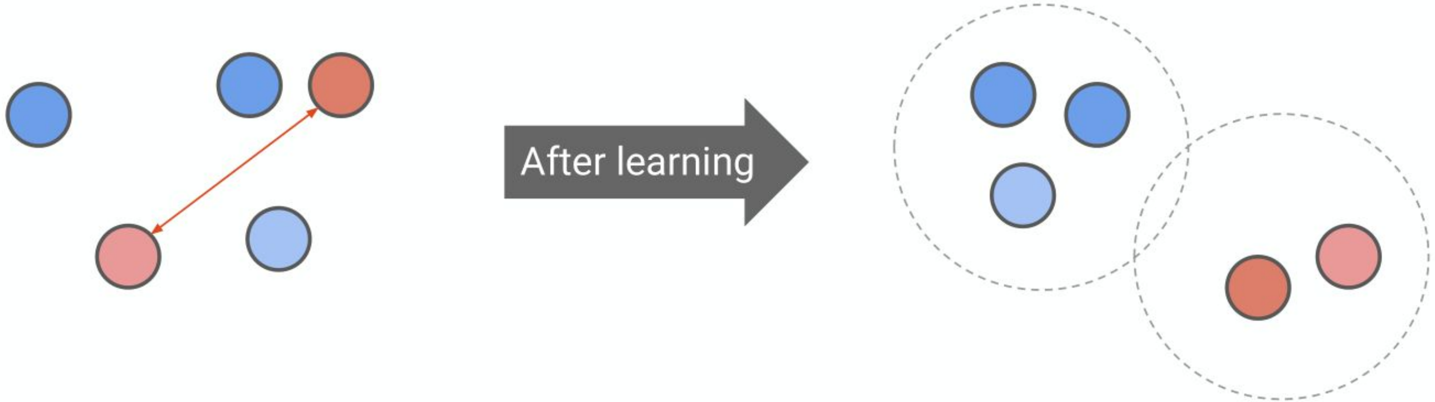
Self-prediction: Given one individual data sample, the task is to predict one (unseen) part of the sample given the other part.



(Famous illustration from Yann LeCun)

Self-supervised Visual Learning

Contrastive Learning: Learn representations such that embeddings of similar sample pairs are close to each other while dissimilar ones are far apart



Self-supervised Visual Learning

Contrastive Learning: Learn representations such that embeddings of similar sample pairs are close to each other while dissimilar ones are far apart

- **Contrastive Loss**: Given two labeled samples $(\mathbf{x}_i, y_i)$ and $(\mathbf{x}_j, y_j)$

$$\mathcal{L}_{\text{cont}}(\mathbf{x}_i, \mathbf{x}_j, \theta) = \mathbb{1}[y_i = y_j] \underbrace{\boxed{\|f_\theta(\mathbf{x}_i) - f_\theta(\mathbf{x}_j)\|_2^2}}_{\text{minimize}} + \mathbb{1}[y_i \neq y_j] \max(0, \epsilon - \underbrace{\boxed{\|f_\theta(\mathbf{x}_i) - f_\theta(\mathbf{x}_j)\|_2}}_{\text{maximize}})^2$$

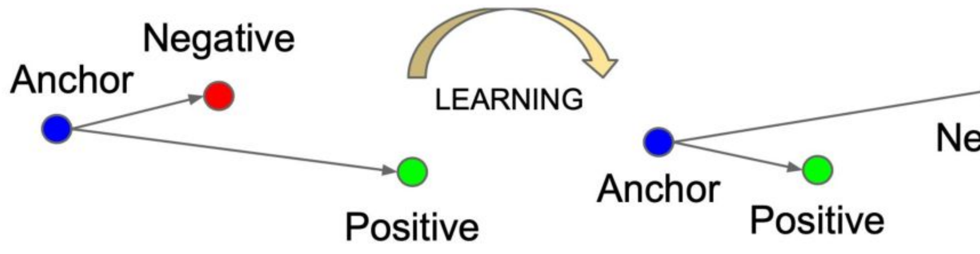
Chopra et al. 2005

Self-supervised Visual Learning

<u>**Contrastive Learning**</u>: Learn representations such that embeddings of similar sample pairs are close to each other while dissimilar ones are far apart

- **Triplet Loss**: Minimize distance between anchor **x** and positive example **x+** and maximize distance between anchor **x** and negative example **x-**

$$\mathcal{L}_{\text{triplet}}(\mathbf{x}, \mathbf{x}^+, \mathbf{x}^-) = \sum_{\mathbf{x} \in \mathcal{X}} \max\left(0, \|f(\mathbf{x}) - f(\mathbf{x}^+)\|_2^2 - \|f(\mathbf{x}) - f(\mathbf{x}^-)\|_2^2 + \epsilon\right)$$

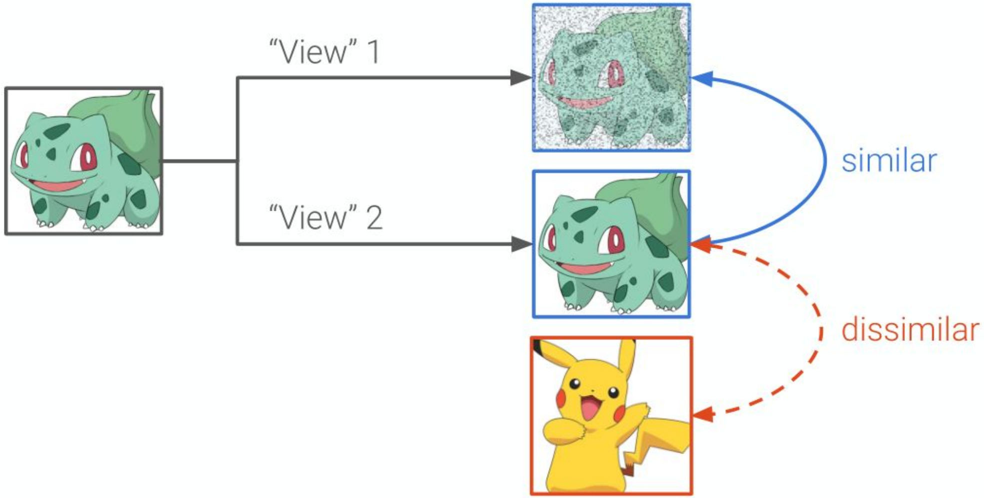


(Schroff et al. 2015)

Self-supervised Visual Learning

Contrastive Learning: Learn representations such that embeddings of similar sample pairs are close to each other while dissimilar ones are far apart

- **Visual Pretext**: Use data augmentation to each image and consider its distorted versions as similar pairs



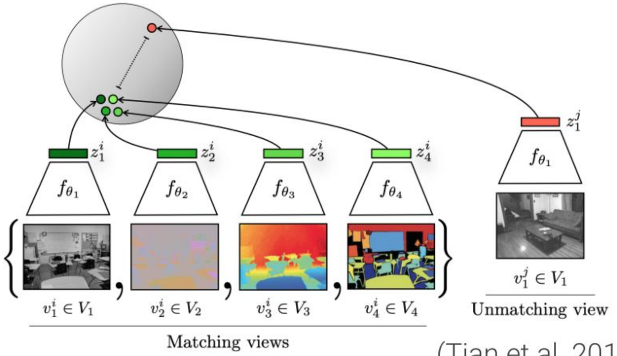
Self-supervised Visual Learning

- **Visual Pretext**:

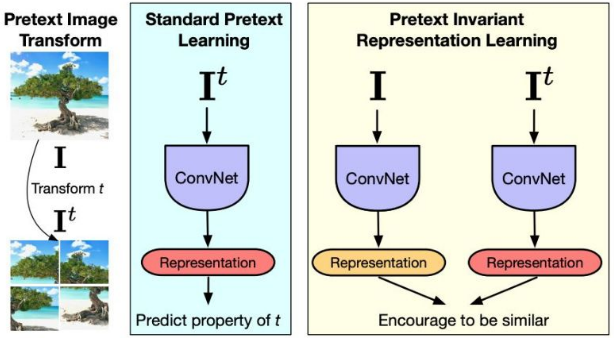
  Augmented Multiscale Deep InfoMax
  (**AMDIM**; Bachman et al. 2019)
  - Views from different **augmentations**

  Contrastive Multiview Coding
  (**CMC**; Tian et al. 2019)
  - Multiple views from different **channels**

  Pretext-Invariant Representation Learning
  (**PIRL**; Misra et al. 2019)
  - Jigsaw transformation



(Tian et al. 2019)



(Misra et al. 2019)

Questions?