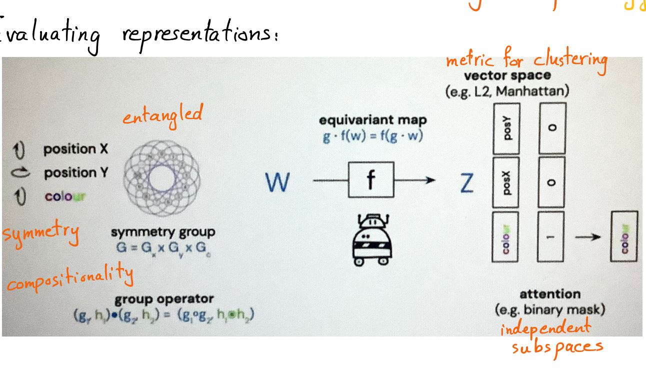
Deepmind - Unsupervised representation learning

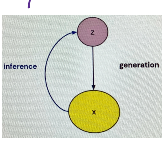
Thursday, 25 February 2021 22:38

- Unsupervised learning: find structure in provided data.
  Use cases: - Clustering
  - Dimensionality reduction.
- Challenges: data efficiency, Robustness, Generalization, Transfer, Common sense.
- Benchmarks:
  Semi-supervised learning: Visual task adaptation Benchmark (Google)
  Generalization: Coin Run (Open AI)
  Transfer: DM Lab-30 (DeepMind)

- Representation:
  - Representational form orthogonal to information content.
  - Useful abstraction to make different computations more efficient.
  - defined by the shape of the manifold on which the data lie within the representational space.
  - Good representation specifications from Neuroscience:
    1. Untangled.
    2. Attention. (things not relevant)
    3. Clustering.
    4. Latent states (beyond perception)
    4. Compositionality.
  - Compositionality: Leads to open-endedness; can construct arbitrary large number of meaningful complex expressions from a finite number of constituent expressions and combination rules.
  - Representation specification from Physics:
    Symmetry transformations
  - Representations in AI:
  - Information bottleneck.
  - Invariance: representation remains unchanged when a certain transformation is applied to the input: $f(g.x) = f(x)$
  - Equivariance: representation reflects the transformation applied to the input: $f(g.x) = g.f(x)$
  - Disentangled representation learning: invert a generative process to inference process
    DL of representations: looking forward [Bengio'13]
    - Symmetry: given a symmetry group of transformation G affects an abstract state W
      → we want to find an equivariant map f:
      $g.f(w) = f(g.w)$; $\forall g \in G, w \in W$
      Towards a definition of Disentangled Repr. [Higgins'19]
- Evaluating representations:

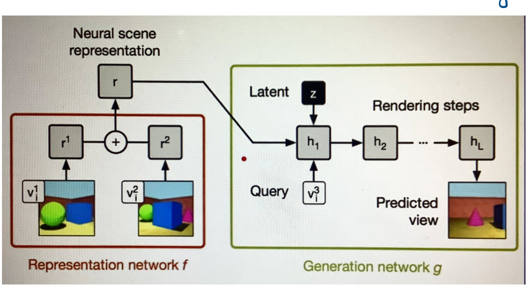


- Representation learning techniques with DNN:
  1. Generative modeling: learn the distribution using generative modeling, often through reconstruction.
  2. Contrastive loss: use classification losses to learn representations that preserve temporal or spatial data consistency.
  3. Self-supervision: Exploit knowledge of data to design learning tasks which lead to useful representations.
- Generative modeling for representation learning:
  - Model the underlying data distribution.
  - Unsupervised learning (task agnostic).
  - Intuition: most efficient way to model a distribution is to extract common patterns (representations).
  - Latent variable models: model the data generation process as a mapping from a low dimensional unknown (latent) space to the data distribution.
  - Inference in latent variable models: find $p(z|x)$
    Intuition: find the underlying factors which generated the data (with uncertainty estimates).

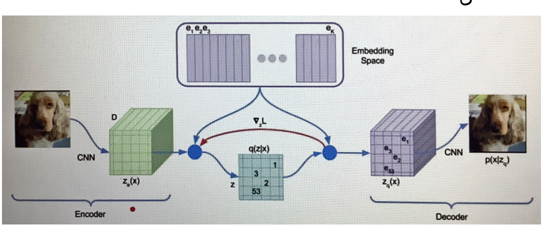


  - Variational autoencoder:
    Maximum likelihood $\mathbb{E}_{p^*(x)}[\log p_\theta(x)]$
    LVM $\log p_\theta(x) = \log \int p_\theta(x|z)p(z)\,dz$
    Lower bound on maximum likelihood (ELBO):
    $$\log p_\theta(x) \geq \underbrace{\mathbb{E}_{q_n(z|x)} \log p_\theta(x|z)}_{\text{reconstruct}} \underbrace{- KL(q_n(z|x) \| p(z))}_{\text{stay close to prior}}$$
    $q_n(z|x)$ - approximate posterior

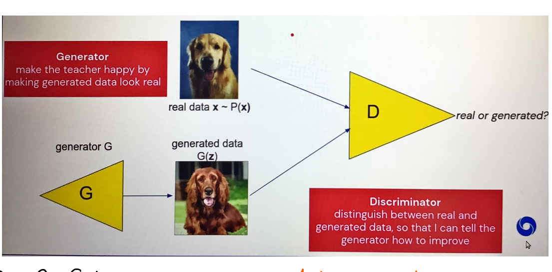
  - Beta-VAE: learn disentangled continuous representations encoding semantic information.
    - Beta-VAE in RL: improves geralization & transfer.
      DARLA: improving zero-shot transfer in RL [Higgins'17]
  - Sequential VAEs - Conv Draw.
    Towards conceptual compression [Gregor'18]
  - Layered models - Monet
    MONet: unsupervised scene Decom. [Burgess'19]
    - Using attention in a multi-level process leads to a generative model which learns concepts (objects) in an unsupervised way.
    - Latent traversals show that Monet learns to encode the position of an object into a latent.
      → learns tasks in RL quickly
- Generative Query Networks (GQN):
  - Learn a representation by providing examples angle scene pairs → use the model to generate how the scene looks like from a different angle.



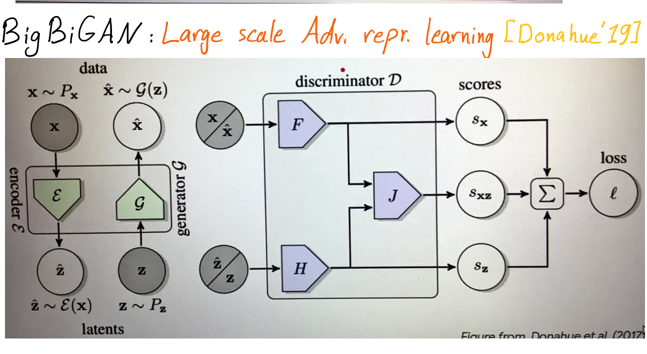
- Vector Quantized VAEs (VQ-VAE):
  - Learning discrete latent variable is challenging (high variance gradient estimation).
    - Solution: reconstruct by using discrete latent variables to index into a learned continuous embedding space.



  - Discrete latent variables can be used to capture high and low level information from data.
    Neural Discrete repr. learning [von den Oord '17]
- Generative Adversarial Networks (GAN):

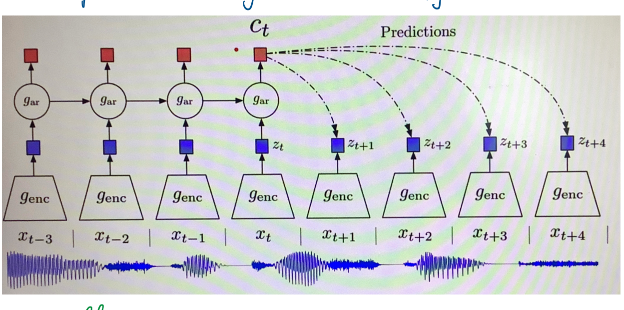


- BigBiGAN: Large scale Adv. repr. learning [Donahue'19]

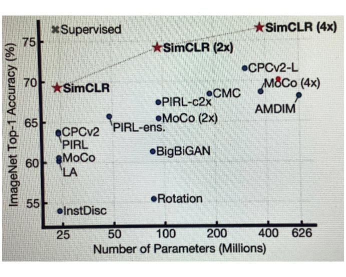


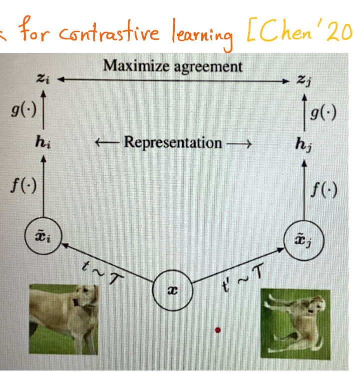
  - No pixel loss reconstruction → reconstructions capture high level information → latents capture high level information.

- Contrastive learning:
  - Remove the need for a generative model.
  - Ensure model learns the right context.
  - Use classification losses.
  e.g: Word2vec
  - Contrastive Predictive Coding [van den Oord '18]
    - Maximize mutual information between data and representations.
    - Uses supervised learning to model density ratios.



  - More efficient in low data regime compared to learning from pixels.
    Data-efficient image recognition with CPC [Henaff'19]
  - SimCLR: A simple framework for contrastive learning [Chen'20]





- Self-supervised learning:
  - Design tasks based on data modalities.
  - Easy to obtain data.
  - Representations are deterministic, often features of NNs.
  e.g.: - Colorful image colorization.
    - Unsupervised visual representation learning by context prediction; unsupervised object discovery.
  - Unsupervised Representation learning by sorting sequences:
    - Learn temporal coherence from video.
    - avoid data generation in pixel space.
  - Bert: Pretraining of Deep Bidirectional transformers for language understanding [Devlin'19]
    - Representation learned with multiple tasks (self-supervision meets generative modeling)

  - Self-supervised learning exploit domain knowledge to build tasks useful for representation learning.