Invariant-equivariant representation learning for multi-class data

Ilya Feige

Faculty

→

Invariant-equivariant representation learning

High-level introduction

Separating content from style

> *This work is about disentangling representations. We present a novel approach to an old problem.*

What?

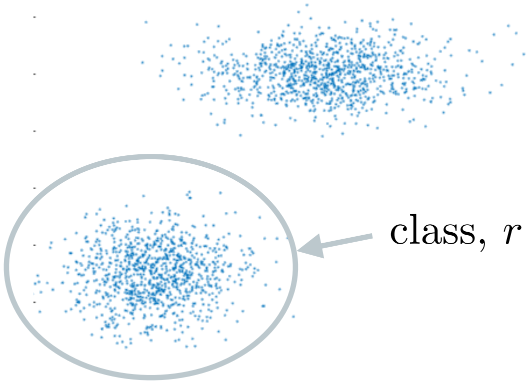
Want to represent the **class** and the **data instance** separately

Separating content from style

> *This work is about disentangling representations. We present a novel approach to an old problem.*

What?

Want to represent the **class** and the **data instance** separately

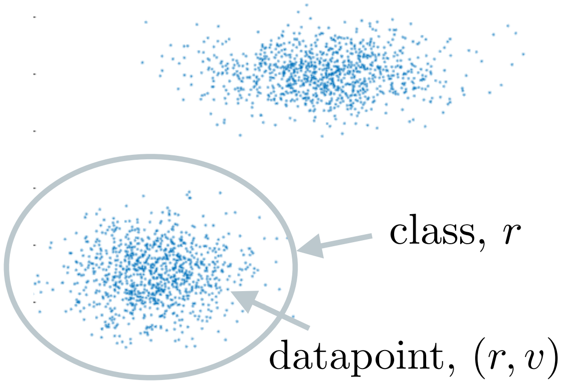


Separating content from style

> *This work is about disentangling representations. We present a novel approach to an old problem.*

What?

Want to represent the **class** and the **data instance** separately

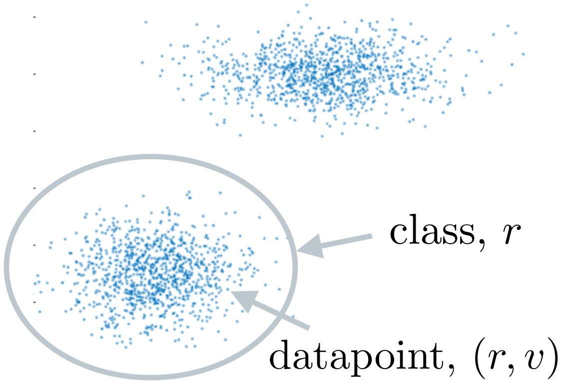


Separating content from style

> *This work is about disentangling representations. We present a novel approach to an old problem.*

What?

Want to represent the **class** and the **data instance** separately



Why?

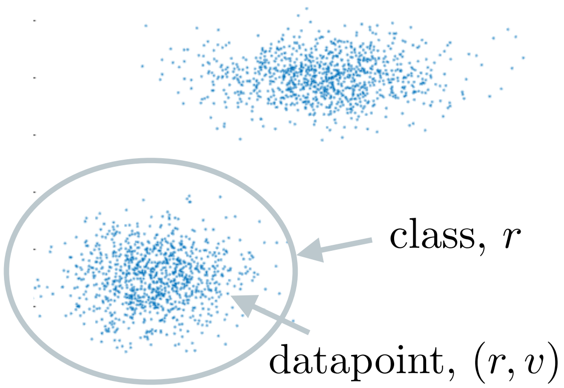
- Classification
- Interpretability
- Object detection
- Topic modelling
- Style transfer
- Face swap
- …

Separating content from style

> *This work is about disentangling representations. We present a novel approach to an old problem.*

What?

Want to represent the **class** and the **data instance** separately



Why?

- Classification
- Interpretability
- Object detection
- Topic modelling
- Style transfer
- Face swap
- …

What else?

This is not a new topic…

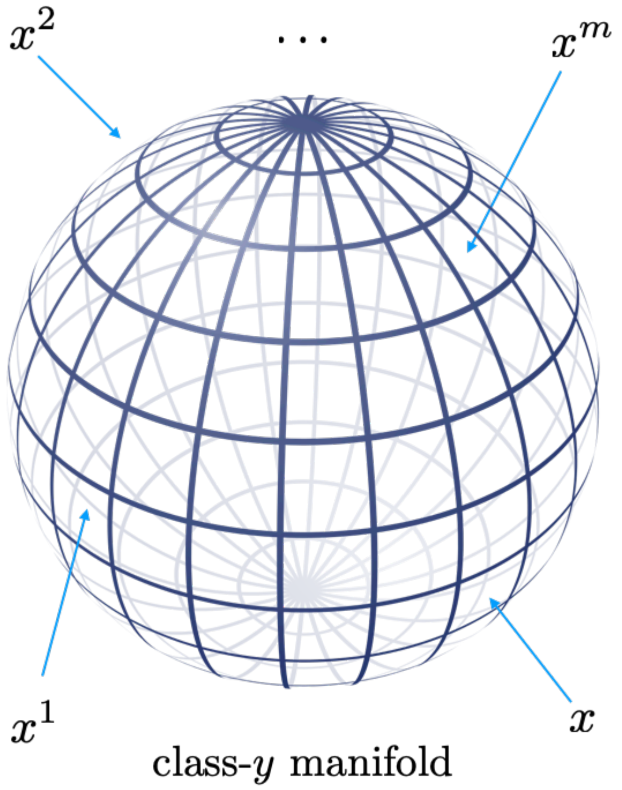
- Tenenbaum & Freeman. (2000)
- Reed et al. (2014)
- Cheung et al. (2014)
- Zhu et al. (2014)
- Radford et al. (2016)
- Chen et al. (2016)
- Makhzani et al. (2016)
- Siddharth et al. (2017)
- …

The main idea

Inferring the class latent using strategic data routing

Invariant (class) latent is deterministically calculated from "complementary" same-class examples:

$$\{x^1, x^2, \ldots, x^m\} \longrightarrow r_y$$



The main idea

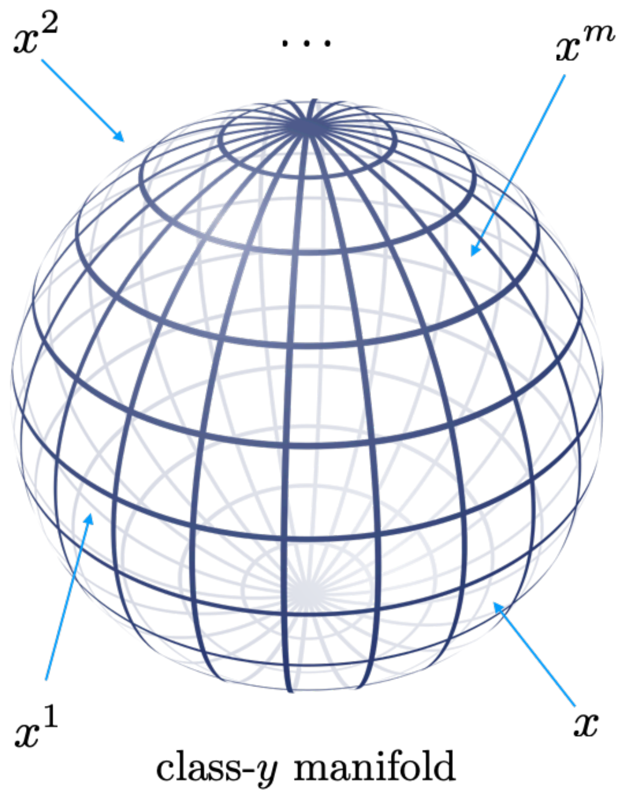
Inferring the class latent using strategic data routing

Invariant (class) latent is deterministically calculated from "complementary" same-class examples:

$$\{x^1, x^2, \ldots, x^m\} \longrightarrow r_y$$

Equivariant (instance) latent is stochastically inferred from both class and datapoint information:

$$\{r_y, x\} \longrightarrow v$$



The main idea

Inferring the class latent using strategic data routing

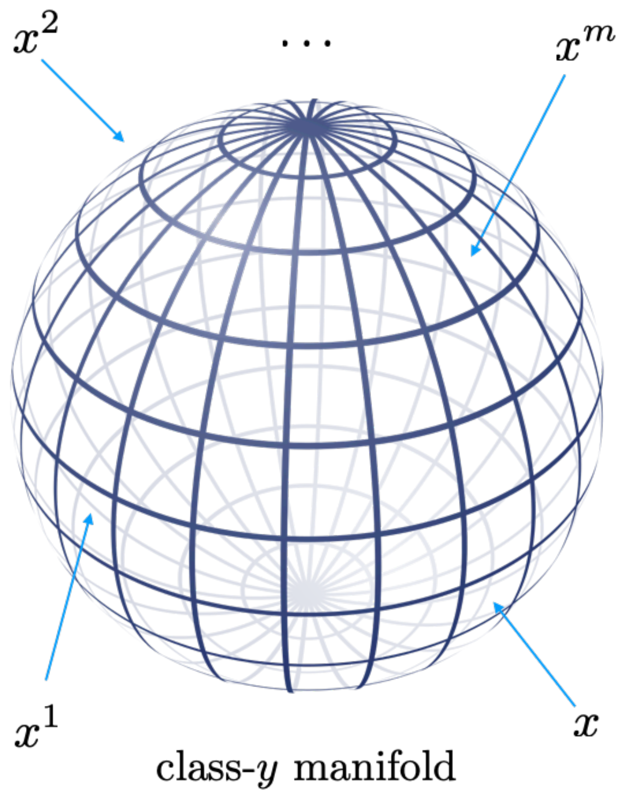
Invariant (class) latent is deterministically calculated from "complementary" same-class examples:

$$\{x^1, x^2, \ldots, x^m\} \longrightarrow r_y$$

Equivariant (instance) latent is stochastically inferred from both class and datapoint information:

$$\{r_y, x\} \longrightarrow v$$

Inspired by GQNs (Eslami et al., 2018)



Invariant-equivariant representation learning

Some detail

Flash through the math

Generative model

$$p\left(\{x_n, y_n\}_{n=1}^{N}\right) = \prod_{n=1}^{N} \int dv_n \, dr_n \, p_\theta(x_n | r_n, v_n) \, \delta\left(r_n - r_{y_n}\right) p(v_n) \, p(y_n)$$

Inference model

$$r_{y_n} = \frac{1}{m} \sum_{i=1}^{m} f_{\theta_{\text{inv}}}(x^i)$$

$$q_\phi(v | r_y, x) = \mathcal{N}\left(\mu_\phi(r_y, x), \sigma_\phi^2(r_y, x) I\right)$$

Objective

$$\mathcal{L}_{\text{lab}} = \mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y, x) \middle|\middle| p(v)\right] + \log p(y)$$

$$\mathcal{L}_{\text{unlab}} = \mathbb{E}_{q(y|x)}\left[\mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y, x) \middle|\middle| p(v)\right]\right] - \mathcal{D}_{\text{KL}}\left[q(y|x) \middle|\middle| p(y)\right]$$

Flash through the math

Generative model

$$p(\{x_n, y_n\}_{n=1}^N) = \prod_{n=1}^N \int dv_n \, dr_n \, p_\theta(x_n | r_n, v_n) \, \delta$$

Generative model
with 2 latent variables

Inference model

$$r_{y_n} = \frac{1}{m} \sum_{i=1}^m f_{\theta_{\text{inv}}}(x^i)$$

$$q_\phi(v | r_y, x) = \mathcal{N}\left(\mu_\phi(r_y, x), \sigma_\phi^2(r_y, x) I\right)$$

Objective

$$\mathcal{L}_{\text{lab}} = \mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y, x) || p(v)\right] + \log p(y)$$

$$\mathcal{L}_{\text{unlab}} = \mathbb{E}_{q(y|x)} \left[\mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y, x) || p(v)\right]\right] - \mathcal{D}_{\text{KL}}\left[q(y|x) || p(y)\right]$$

Flash through the math

Generative model

$$p(\{x_n, y_n\}_{n=1}^{N}) = \prod_{n=1}^{N} \int dv_n \, dr_n \, p_\theta(x_n | r_n, v_n) \, \delta$$

Generative model with 2 latent variables

Inference model

$$r_{y_n} = \frac{1}{m} \sum_{i=1}^{m} f_{\theta_{\text{inv}}}(x^i)$$

Deterministic, from same-class complementary data

$$q_\phi(v | r_y, x) = \mathcal{N}\big(\mu_\phi(r_y, x), \sigma_\phi^2(r_y, x) I\big)$$

Objective

$$\mathcal{L}_{\text{lab}} = \mathbb{E}_{q(v|r_y,x)} \log p(x|r_y, v) - \mathcal{D}_{\text{KL}}\big[q(v|r_y, x) || p(v)\big] + \log p(y)$$

$$\mathcal{L}_{\text{unlab}} = \mathbb{E}_{q(y|x)} \Big[\mathbb{E}_{q(v|r_y,x)} \log p(x|r_y, v) - \mathcal{D}_{\text{KL}}\big[q(v|r_y, x) || p(v)\big]\Big] - \mathcal{D}_{\text{KL}}\big[q(y|x) || p(y)\big]$$

Flash through the math

Generative model

$$p(\{x_n, y_n\}_{n=1}^{N}) = \prod_{n=1}^{N} \int dv_n \, dr_n \, p_\theta(x_n | r_n, v_n) \, \delta$$

Generative model with 2 latent variables

Inference model

$$r_{y_n} = \frac{1}{m} \sum_{i=1}^{m} f_{\theta_{\text{inv}}}(x^i)$$

Deterministic, from same-class complementary data

$$q_\phi(v | r_y, x) = \mathcal{N}\big(\mu_\phi(r_y, x), \sigma_\phi^2(r_y, x) I\big)$$

Standard VAE

Objective

$$\mathcal{L}_{\text{lab}} = \mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\big[q(v|r_y, x) || p(v)\big] + \log p(y)$$

$$\mathcal{L}_{\text{unlab}} = \mathbb{E}_{q(y|x)} \Big[\mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\big[q(v|r_y, x) || p(v)\big]\Big] - \mathcal{D}_{\text{KL}}\big[q(y|x) || p(y)\big]$$

Flash through the math

Generative model

$$p(\{x_n, y_n\}_{n=1}^{N}) = \prod_{n=1}^{N} \int dv_n \, dr_n \, p_\theta(x_n | r_n, v_n) \, \delta$$

Generative model with 2 latent variables

Inference model

$$r_{y_n} = \frac{1}{m} \sum_{i=1}^{m} f_{\theta_{\text{inv}}}(x^i)$$

Deterministic, from same-class complementary data

$$q_\phi(v | r_y, x) = \mathcal{N}\left(\mu_\phi(r_y, x), \sigma_\phi^2(r_y, x) I\right)$$

Standard VAE

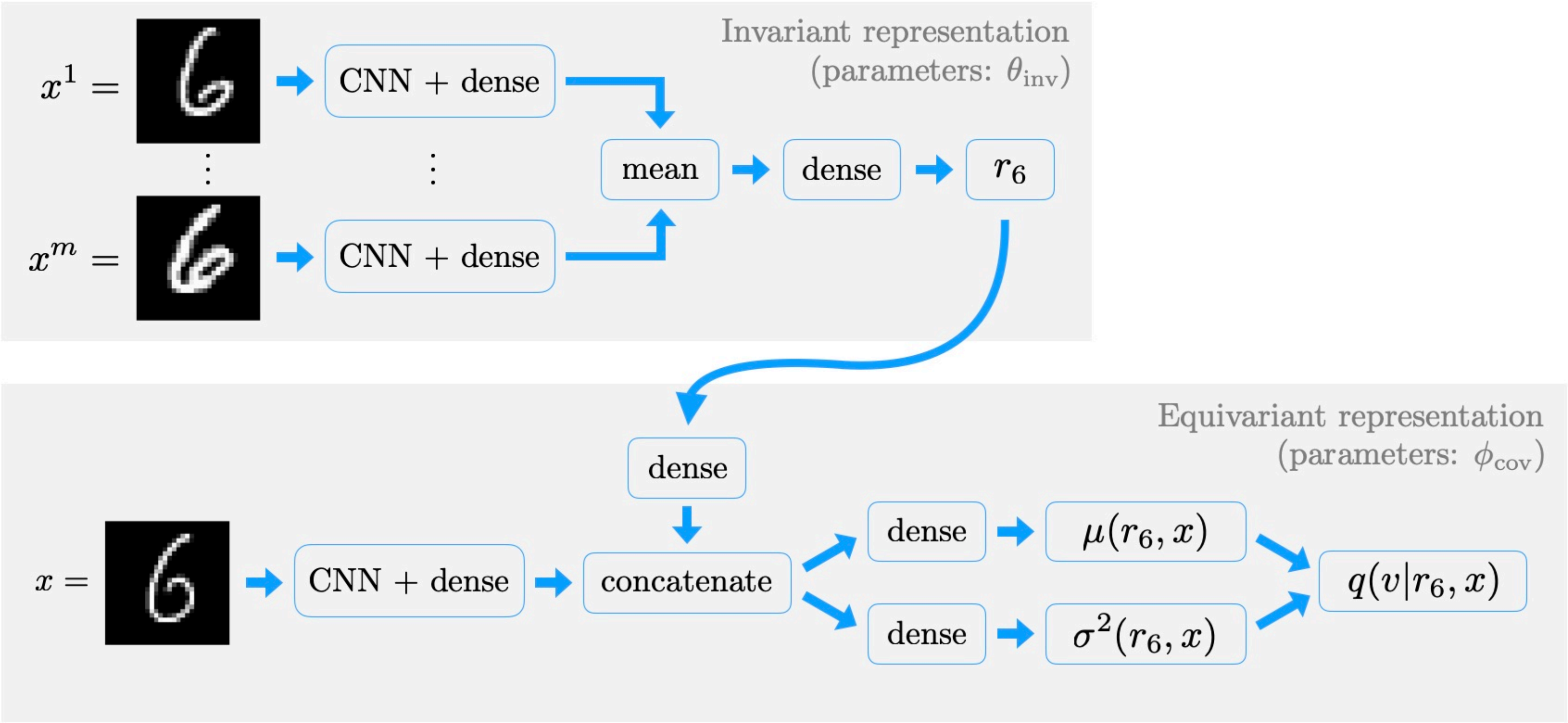
Objective

$$\mathcal{L}_{\text{lab}} = \mathbb{E}_{q(v|r_y,x)} \log p(x | r_y, v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y,x) || p(v)\right] + \log p(y)$$

$$\mathcal{L}_{\text{unlab}} = \mathbb{E}_{q(y|x)}\left[\mathbb{E}_{q(v|r_y,x)} \log p(x|r_y,v) - \mathcal{D}_{\text{KL}}\left[q(v|r_y,x)||p(v)\right]\right] - \mathcal{D}_{\text{KL}}\left[q(y|x)||p(y)\right]$$

Standard ELBOs

Inference in pictures



Invariant-equivariant representation learning

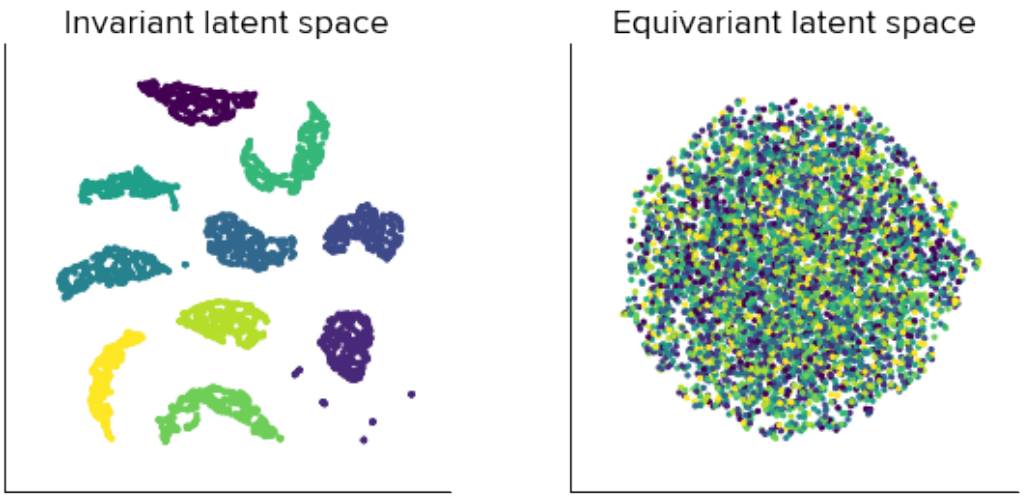
Results

Inferred latent space on MNIST

Latent spaces disentangle

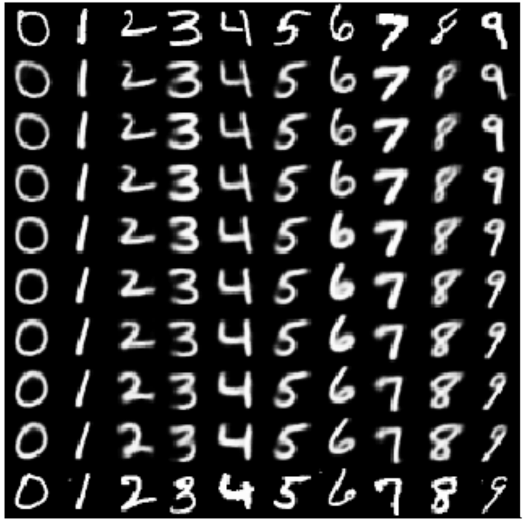
The invariant latent learns to separate the classes

The equivariant latent learns to ignore class information

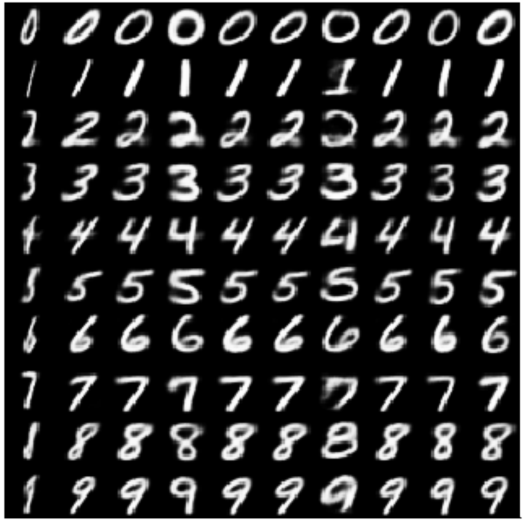


Generations from various latent configurations (MNIST)

Samples from each class



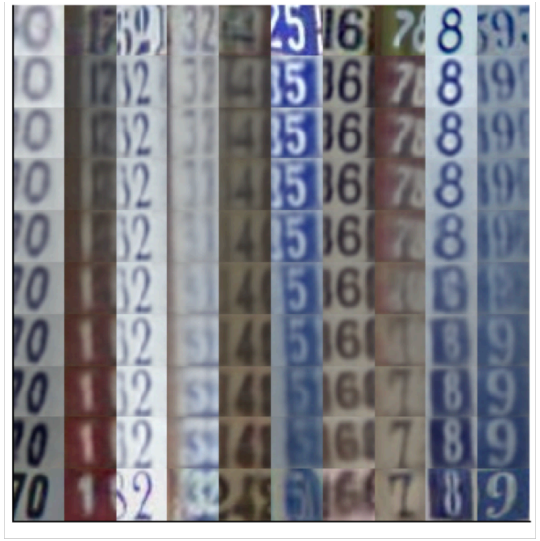
Equivariant interpolations, for multiple invariant latents



Invariant steps, for multiple equivariant latents

Generations from various latent configurations (SVHN)

Samples from each class



Equivariant interpolations, for multiple invariant latents

Invariant steps, for multiple equivariant latents

Classification

Fully supervised

(Using 0-parameter distance to nearest r_y)

	EquiVAE	Benchmark
MNIST	0.82 ± 0.05	0.84 ± 0.03
SVHN	12.30 ± 0.28	10.04 ± 0.14

Semi supervised

(Using label-inference distribution)

	Labels	EquiVAE	Benchmark
MNIST	100	8.90 ± 0.70	21.91 ± 0.66
	600	3.99 ± 0.17	6.64 ± 0.35
	1000	3.34 ± 0.17	5.43 ± 0.31
	3000	2.23 ± 0.14	2.96 ± 0.11
SVHN	1000	37.95 ± 0.66	39.64 ± 1.47
	3000	24.95 ± 0.57	25.50 ± 0.91

Benchmark is always the equivalent network, +2 dropout layers, trained to classify the labelled data

Invariant-equivariant representation learning

Outlook

Outlook

Advantages of this approach

A. Easy to implement

B. Very little tuning needed

C. Reasonably intuitive

D. Performs similarly to comparable approaches

Disadvantages

A. Requires some labels

B. Does not achieve state-of-the-art

At Faculty, we are implementing this technique into our generic object detection pipeline to extract invariant object representations

Thank you

If you're interested in finding out more about this work, or Faculty in general, please get in touch!

ilya@faculty.ai

Director of AI

info@faculty.ai

+44 20 3637 9415

54 Welbeck St, Marylebone,
London W1G 9XS, UK

Follow us: