Interpreting the Latent Space of Generative Adversarial Networks using Supervised Learning

Toan Van Pham[1] Tam Minh Nguyen[1] Ngoc N. Tran[1] Viet Hoai Nguyen[1] Linh Bao Doan[1] Huy Quang Dao[1] Ta Minh Thanh[2]

[1]R&D Department
Sun-Asterisk Inc.

[2]Faculty of Computer Science
Le Quy Don Technical University

The 14th The International Conference on Advanced COMPuting and Applications, November 2020

Table of Contents

Sun*

Table of Contents

Understanding GAN's latent space

2 branches of questions

- Are scalar variables of GAN's latent vector entangled? If they are, would it be beneficial to disentangle them? If yes, how to do that? And what feature of an image that those scalar variables capture?

Understanding GAN's latent space

2 branches of questions

- Are scalar variables of GAN's latent vector entangled? If they are, would it be beneficial to disentangle them? If yes, how to do that? And what feature of an image that those scalar variables capture?
- How to manipulate GAN's latent space to observe desired generated images?

Understanding GAN's latent space

2 branches of questions

- Are scalar variables of GAN's latent vector entangled? If they are, would it be beneficial to disentangle them? If yes, how to do that? And what feature of an image that those scalar variables capture?
- How to manipulate GAN's latent space to observe desired generated images?

In our paper, we focus on the 2nd questions for facial images

Table of Contents

Linear interpolation

Interpolate between 2 points in latent space (corresponding to 2 pictures) to observe the change in interpolated images.

- Pros: Easy implementation

Sun*

Linear interpolation

Interpolate between 2 points in latent space (corresponding to 2 pictures) to observe the change in interpolated images.

- Pros: Easy implementation
- Cons:

Linear interpolation

Interpolate between 2 points in latent space (corresponding to 2 pictures) to observe the change in interpolated images.

- Pros: Easy implementation
- Cons:
  - Does not make any implication about GAN's latent scalar variables.

Linear interpolation

Interpolate between 2 points in latent space (corresponding to 2 pictures) to observe the change in interpolated images.

- Pros: Easy implementation
- Cons:
  - Does not make any implication about GAN's latent scalar variables.
  - Can not manipulate GAN's latent space as desire

Vector Arithmetic

Add or subtract vectors in latent space to get desired generated images

- Pros:

Vector Arithmetic

Add or subtract vectors in latent space to get desired generated images

- Pros:
  - Easy implementation

Vector Arithmetic

Add or subtract vectors in latent space to get desired generated images

- Pros:
  - Easy implementation
  - Produce desired generated images

Vector Arithmetic

Add or subtract vectors in latent space to get desired generated images

- Pros:
  - Easy implementation
  - Produce desired generated images
- Cons:

Vector Arithmetic

Add or subtract vectors in latent space to get desired generated images

- Pros:
  - Easy implementation
  - Produce desired generated images
- Cons:
  - Hand-select generated images with desired characteristics used for manipulation

InfoGAN

Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

- Pros:

InfoGAN

Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

- Pros:
  - Unsupervised learning (saving human e ort)

InfoGAN

Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

- Pros:
  - Unsupervised learning (saving human e ort)
- Cons:

InfoGAN

Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

- Pros:
  - Unsupervised learning (saving human e ort)
- Cons:
  - Require training a new GAN model (di cult)

InfoGAN

Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

- Pros:
  - Unsupervised learning (saving human e ort)
- Cons:
  - Require training a new GAN model (di cult)
  - Cannot know in advance the characteristics encoded in hidden variables

Table of Contents

Idea

Create a linear mapping from GAN's latent space to the pre-de ned, semantically meaningful characteristic space.

$$y = zW^{>} + b; \qquad (1)$$

Idea

Create a linear mapping from GAN's latent space to the pre-de ned, semantically meaningful characteristic space.

$$y = zW^{>} + b; \qquad (1)$$

Image manipulation and obstacles

Image manipulation

$$z^{0} = z + \quad w_i$$

$$y^{0} = z^{0} W^{>} + b$$

$$, \quad y + \quad y = (z + \quad w_i) W^{>} + b$$

$$, \quad y + \quad y = (zW^{>} + b) + \quad w_i W^{>}$$

$$, \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \quad w_i \begin{bmatrix} w_1^{>} \\ w_2^{>} \\ \vdots \\ w_n^{>} \end{bmatrix} = \begin{bmatrix} w_1^{>} w_i \\ w_2^{>} w_i \\ \vdots \\ w_n^{>} w_i \end{bmatrix}$$

$$) \quad y_i > 0$$

Image manipulation and obstacles

Obstacle: if w_i is similar to other coe cients, changing z to z' also changes cause changes to other attributes

Overcome: Orthogonality Regularization

$$J(w) = MSE(f_w(z); y) + \quad \| w^\top w \quad I \|$$

Table of Contents

Image manipulation

Orthogonality Regularization e ectiveness

Orthogonality Regularization e ectiveness

Table of Contents

Contribution summary

- Supervised method to explicitly map the latent space with a meaningfully pre-de ned semantic space with advantages compared to existing methods:

Contribution summary

- Supervised method to explicitly map the latent space with a meaningfully pre-de ned semantic space with advantages compared to existing methods:
  - Ultilize pre-trained GAN model, easy to implement

Contribution summary

- Supervised method to explicitly map the latent space with a meaningfully pre-de ned semantic space with advantages compared to existing methods:
  - Ultilize pre-trained GAN model, easy to implement
  - Allow to change the intensity of images' characteristic

Contribution summary

- Supervised method to explicitly map the latent space with a meaningfully pre-defined semantic space with advantages compared to existing methods:
  - Ultilize pre-trained GAN model, easy to implement
  - Allow to change the intensity of images' characteristic
- More robust image manipulation with orthogonality regularization

Sun*

Thank you for listening!