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

Sun*

Table of Contents

1. Understanding GAN's latent space
2. Existing methods. Pros and Cons
   - Linear interpolation
   - Vector Arithmetic
   - InfoGAN
3. Our Method
   - Idea
   - Image manipulation and obstacles
4. Experiment Results
   - Image manipulation
   - Orthogonality Regularization effectiveness
5. Conclusion

Sun*

Table of Contents

1 Understanding GAN's latent space

2 Existing methods. Pros and Cons
- Linear interpolation
- Vector Arithmetic
- InfoGAN

3 Our Method
- Idea
- Image manipulation and obstacles

4 Experiment Results
- Image manipulation
- Orthogonality Regularization effectiveness

5 Conclusion

Sun*

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?

Sun*

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?

Sun*

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

Sun*

Table of Contents

1 Understanding GAN's latent space

2 Existing methods. Pros and Cons
- Linear interpolation
- Vector Arithmetic
- InfoGAN

3 Our Method
- Idea
- Image manipulation and obstacles

4 Experiment Results
- Image manipulation
- Orthogonality Regularization effectiveness

5 Conclusion

Sun*

Linear interpolation

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

- Pros: Easy implementation

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

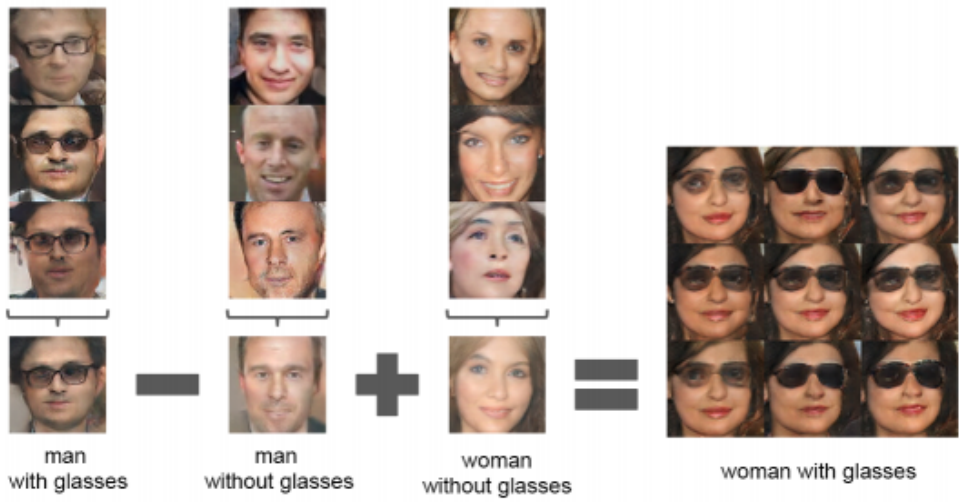
Interpolate between 2 points in latent space (corresponding to 2 pictures) to observe the change in interpolated images.

- 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

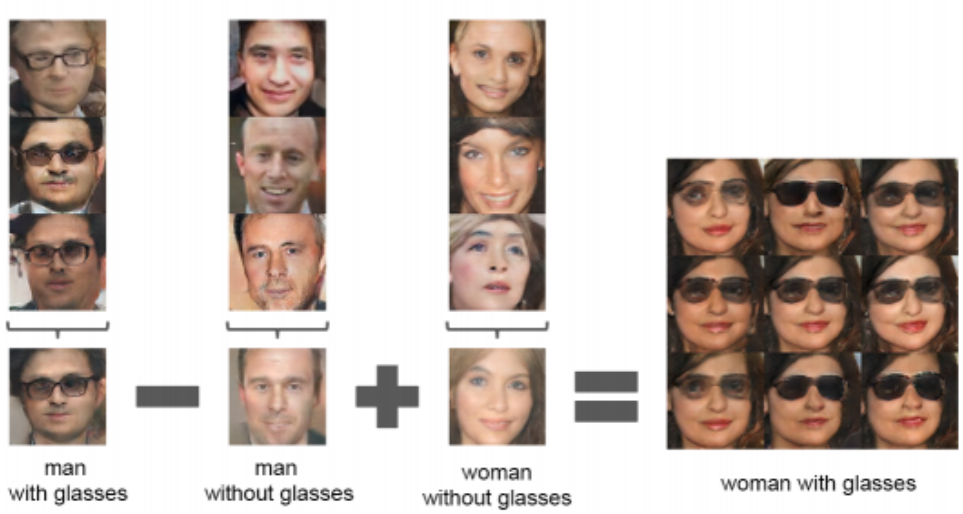
- Pros:



Vector Arithmetic

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

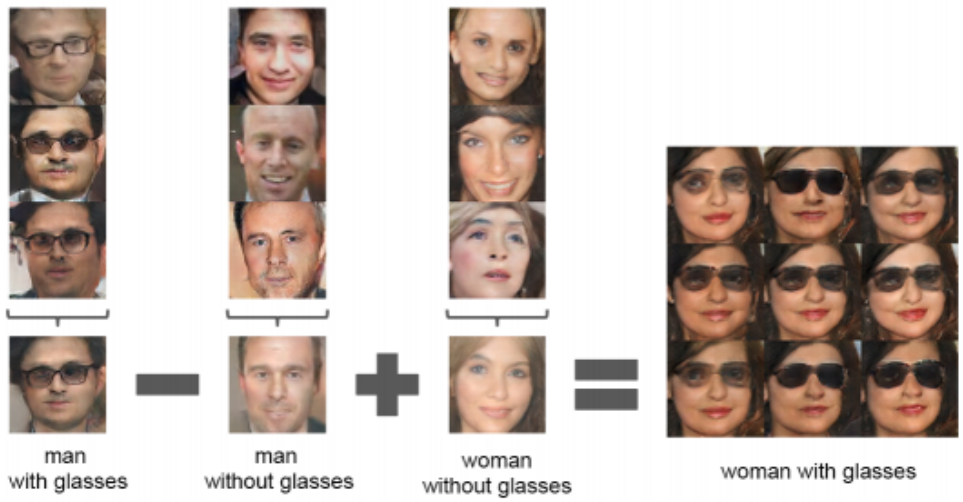
- Pros:
  - Easy implementation



Vector Arithmetic

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

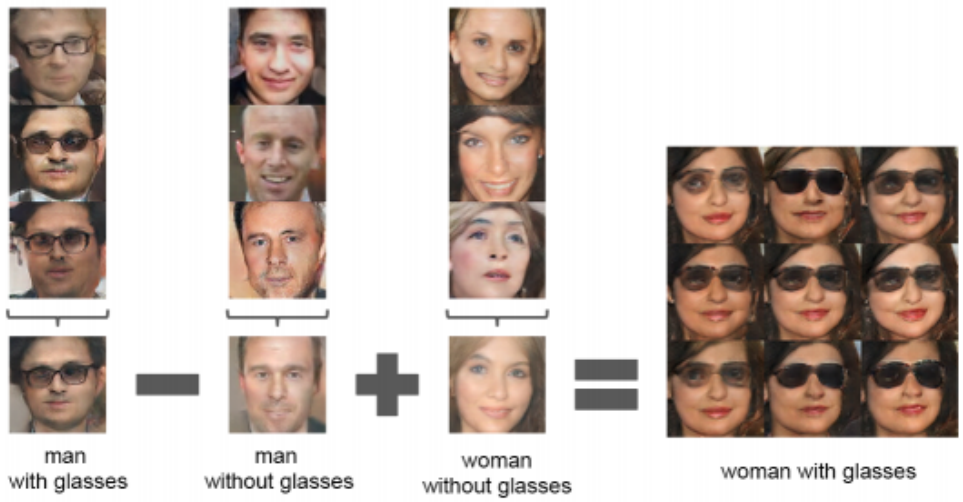
- Pros:
  - Easy implementation
  - Produce desired generated images



Vector Arithmetic

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

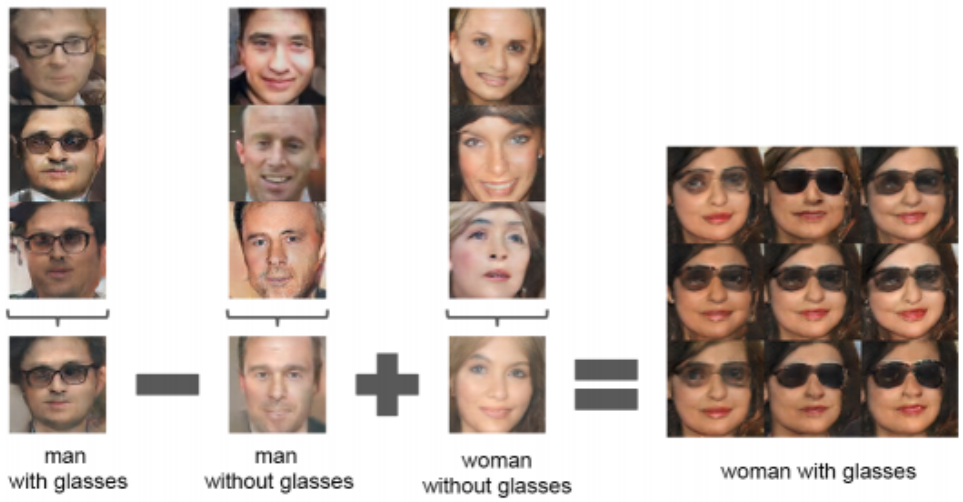
- Pros:
  - Easy implementation
  - Produce desired generated images
- Cons:



Vector Arithmetic

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

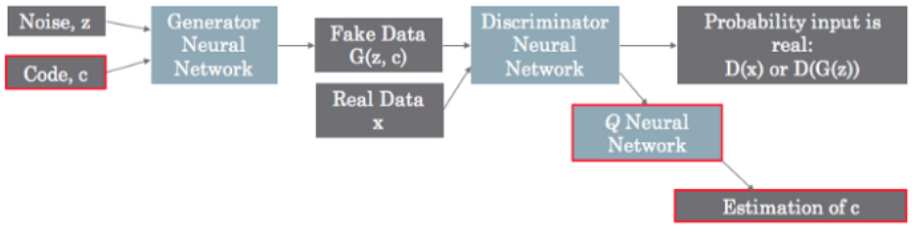
- Pros:
  - Easy implementation
  - Produce desired generated images
- Cons:
  - Hand-select generated images with desired characteristics used for manipulation

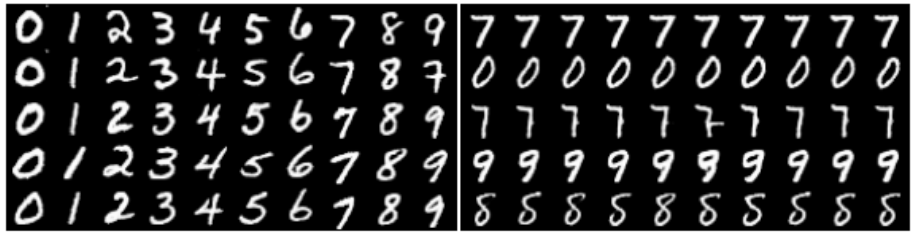


InfoGAN

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

- Pros:

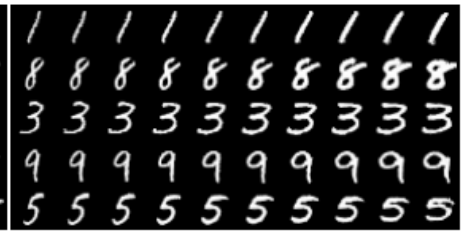




(a) Varying c_1 on InfoGAN (Digit type)

(b) Varying c_1 on regular GAN (No clear meaning)

(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

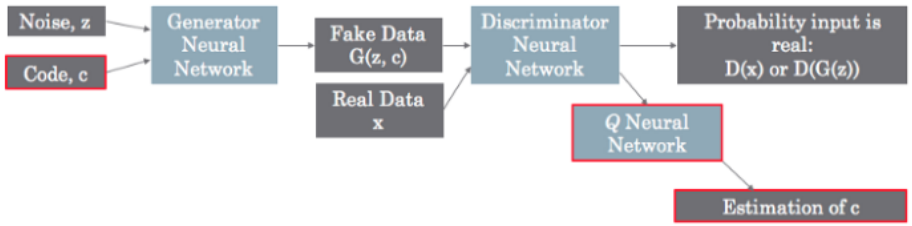


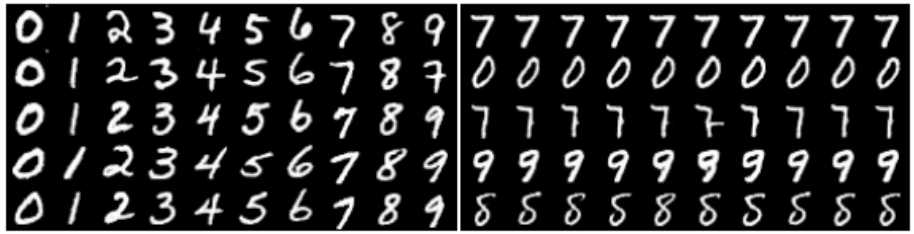
(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

InfoGAN

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

- Pros:
  - Unsupervised learning (saving human effort)

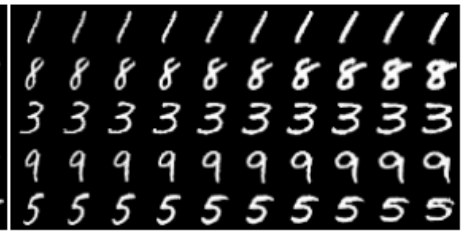




(a) Varying c_1 on InfoGAN (Digit type)

(b) Varying c_1 on regular GAN (No clear meaning)

(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

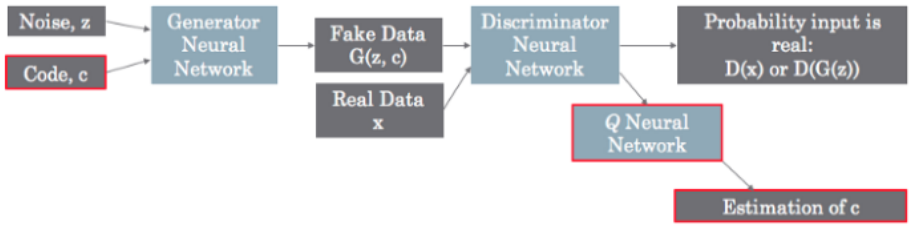


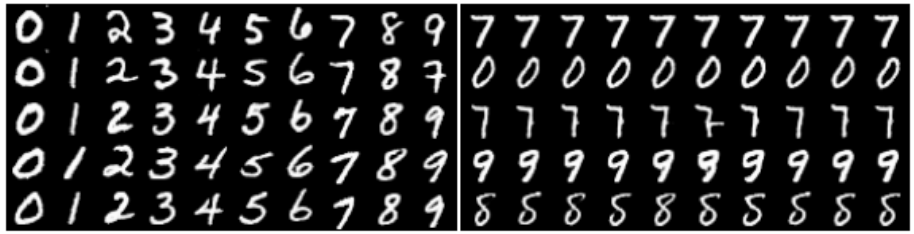
(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

InfoGAN

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

- Pros:
  - Unsupervised learning (saving human effort)
- Cons:

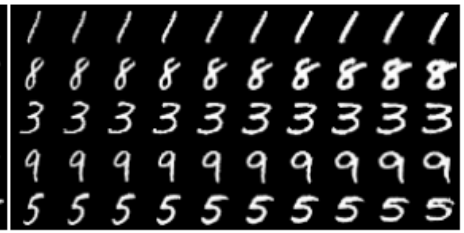




(a) Varying c_1 on InfoGAN (Digit type)

(b) Varying c_1 on regular GAN (No clear meaning)

(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

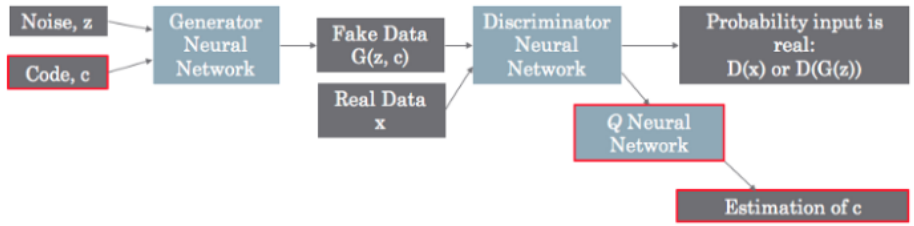


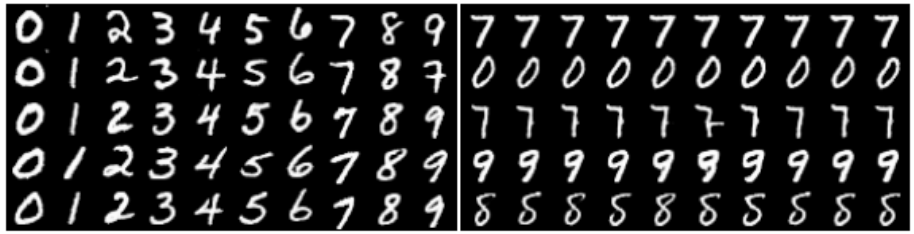
(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

InfoGAN

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

- Pros:
  - Unsupervised learning (saving human effort)
- Cons:
  - Require training a new GAN model (difficult)

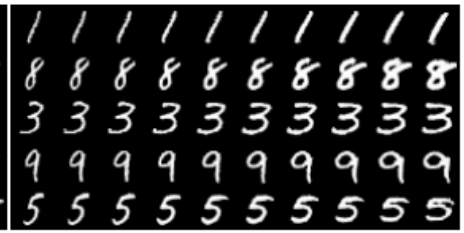




(a) Varying c_1 on InfoGAN (Digit type)

(b) Varying c_1 on regular GAN (No clear meaning)

(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

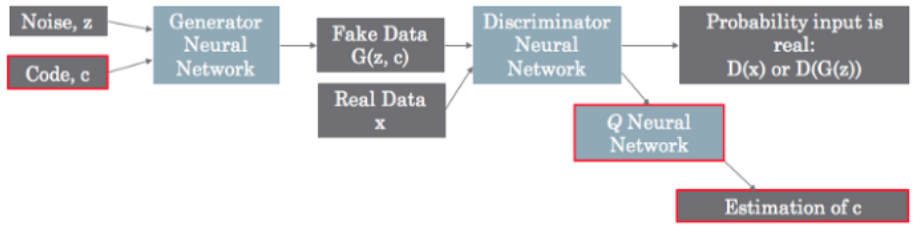


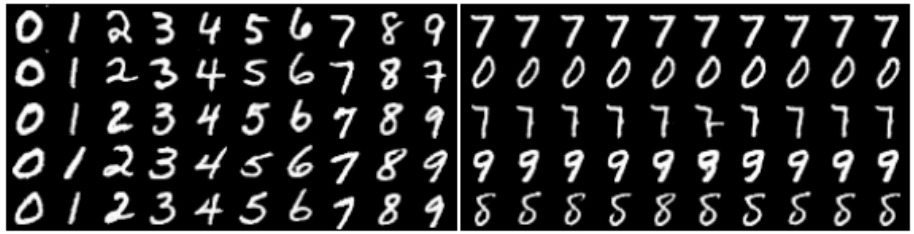
(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

InfoGAN

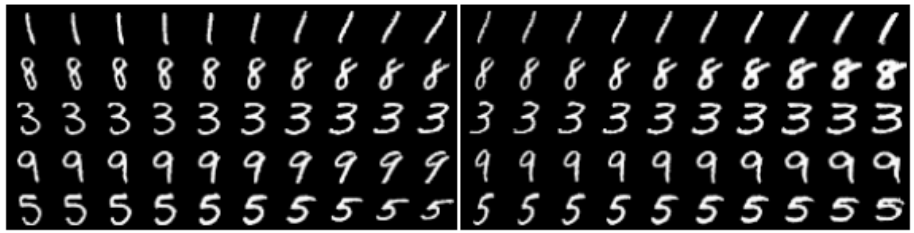
Used another easily, semantically understandable hidden variable in a lower dimension to encode the latent variables

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





(a) Varying c_1 on InfoGAN (Digit type) (b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation) (d) Varying c_3 from -2 to 2 on InfoGAN (Width)

Table of Contents

1 Understanding GAN's latent space

2 Existing methods. Pros and Cons
- Linear interpolation
- Vector Arithmetic
- InfoGAN

3 Our Method
- Idea
- Image manipulation and obstacles

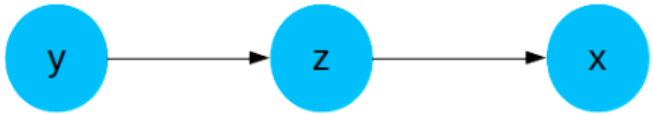
4 Experiment Results
- Image manipulation
- Orthogonality Regularization effectiveness

5 Conclusion

Sun*

Idea

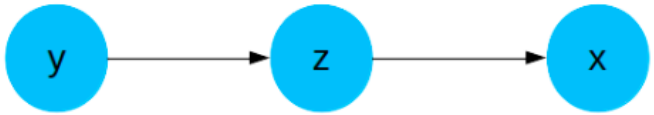
Create a linear mapping from GAN's latent space to the pre-defined, semantically meaningful characteristic space.

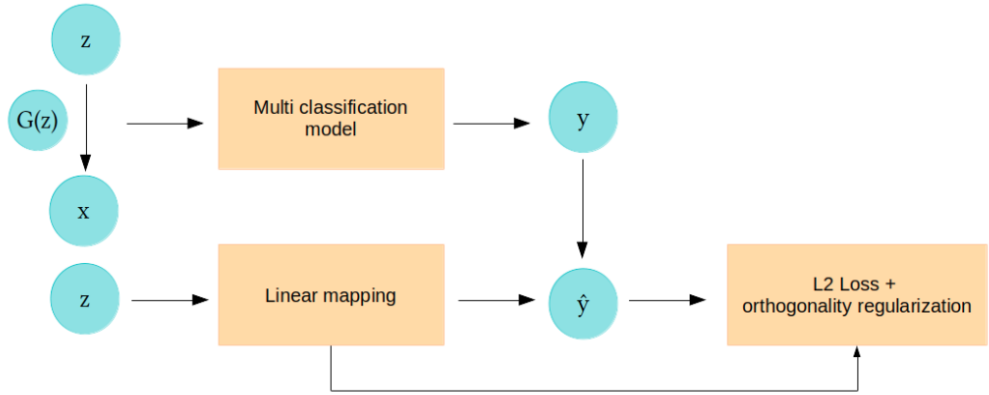


$$y = zW^\top + b, \qquad (1)$$

Idea

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





$$y = zW^\top + b, \qquad (1)$$

Image manipulation and obstacles

Image manipulation

$$z' = z + \alpha w_i$$
$$y' = z' W^\top + b$$

$$\Leftrightarrow y + \Delta y = (z + \alpha w_i) W^\top + b$$
$$\Leftrightarrow y + \Delta y = (z W^\top + b) + \alpha w_i W^\top$$
$$\Leftrightarrow \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_n \end{bmatrix} = \alpha w_i \begin{bmatrix} \Delta w_1^\top \\ \Delta w_2^\top \\ \vdots \\ \Delta w_n^\top \end{bmatrix} = \begin{bmatrix} \Delta \alpha w_1^\top w_i \\ \Delta \alpha w_2^\top w_i \\ \vdots \\ \Delta \alpha w_n^\top w_i \end{bmatrix}$$
$$\Rightarrow \Delta y_i > 0$$

Image manipulation and obstacles

Obstacle: if w_i is similar to other coefficients, changing z to z' also changes cause changes to other attributes
Overcome: Orthogonality Regularization

$$J(w) = \text{MSE}(f_w(z), y) + \lambda \|w^\top w - I\|$$

Sun*

Table of Contents

1 Understanding GAN's latent space

2 Existing methods. Pros and Cons
- Linear interpolation
- Vector Arithmetic
- InfoGAN

3 Our Method
- Idea
- Image manipulation and obstacles

4 Experiment Results
- Image manipulation
- Orthogonality Regularization effectiveness

5 Conclusion

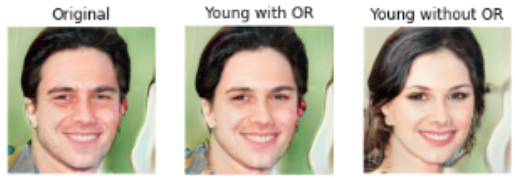
Sun*

Image manipulation

A)

B)

C)



Orthogonality Regularization effectiveness

	Without Regularization	With Regularization
Young	1.000000	1.000000
Attractive	0.859559	0.313507
HeavyMakeup	0.795307	-0.002151
Lipstick	0.791769	0.282624
NoBeard	0.599438	0.262185
BigLips	0.542247	-0.000376
ArchedEyebrows	0.513396	0.042719
OvalFace	0.453567	-0.017463
WavyHair	0.430066	0.036656
PointyNose	0.413676	0.022878
DoubleChin	-0.786877	0.019917
Male	-0.769961	0.039103
BagsUnderEyes	-0.754783	0.015170
GrayHair	-0.749704	0.014643
BigNose	-0.734693	0.034192
WearingTie	-0.669467	0.038437
Bald	-0.585021	0.042678
RecedingHairline	-0.523914	0.009939
Goatee	-0.495711	0.038485

Sun*

Orthogonality Regularization effectiveness

	Original	tfm. w/o reg.	abs_diff_no_reg	tfm_attr_reg	abs_diff_reg
Male	0.999618	0.000217	0.999401	0.997077	0.002541
Makeup	0.001739	0.999788	0.998050	0.009618	0.007879
Lipstick	0.001942	0.999757	0.997815	0.014711	0.012769
Earrings	0.030412	0.814261	0.783849	0.088462	0.058051
BagsUnderEyes	0.782150	0.027579	0.754571	0.671357	0.110794
WavyHair	0.118599	0.870955	0.752356	0.190311	0.071713
5oClockShadow	0.688798	0.000454	0.688344	0.672848	0.015950
BigNose	0.750373	0.073389	0.676984	0.643183	0.107190
OvalFace	0.337116	0.944806	0.607690	0.531252	0.194136
DoubleChin	0.542601	0.001489	0.541111	0.3252063	0.290538
BushyEyebrows	0.960830	0.421305	0.539525	0.972407	0.011577
Young	0.612252	0.991196	0.378944	0.896836	0.284584
ArchedEyebrows	0.89847	0.463946	0.374099	0.065678	0.024169
NarrowEyes	0.391831	0.067585	0.324246	0.296349	0.095482

Sun*

Table of Contents

1 Understanding GAN's latent space

2 Existing methods. Pros and Cons
- Linear interpolation
- Vector Arithmetic
- InfoGAN

3 Our Method
- Idea
- Image manipulation and obstacles

4 Experiment Results
- Image manipulation
- Orthogonality Regularization effectiveness

5 Conclusion

Sun*

Contribution summary

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

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

Sun*

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

Sun*

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!