GeneGAN

Learning Object Transfiguration and Attribute Subspace from Unpaired Data

Taihong Xiao

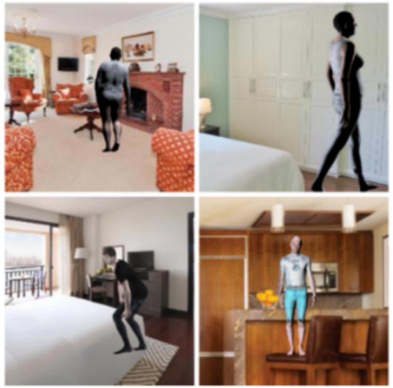
Nov. 1, 2017

Conditional Image Generation

- Applications: Image Editing, Training Data Synthesis
- Photo-realistic modeling and rendering are difficult.

Image ⇒ Model ⇒ Transform ⇒ Render ⇒ Generated Image

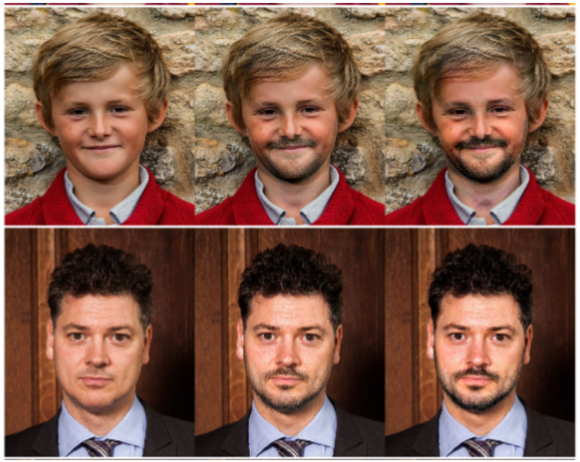
3DMM-family methods
http://cn.arxiv.org/pdf/1612.04904.pdf



SURREAL dataset (CVPR'17)

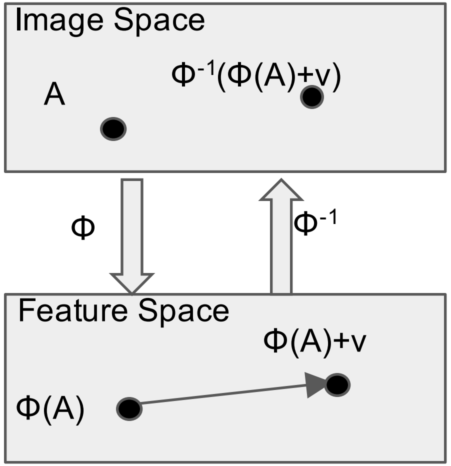
Feature Space Transformation

Image → Project → Transform → Project back → Generated Image



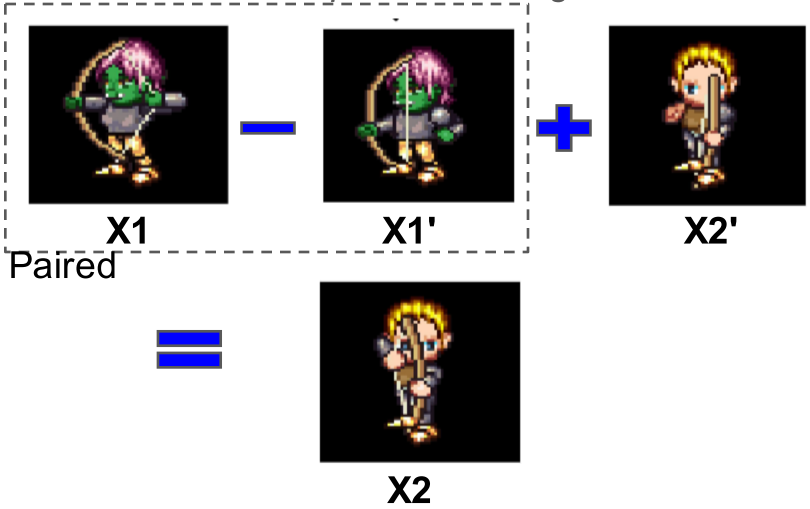
$\Phi^{-1}(\Phi(A))$ $\Phi^{-1}(\Phi(A)+½\,v)$ $\Phi^{-1}(\Phi(A)+$
$v)$ Deep Feature Interpolation (2016)

*Transformation vector **v** as difference between feature cluster centers. Diversity limited by the number of clusters.*

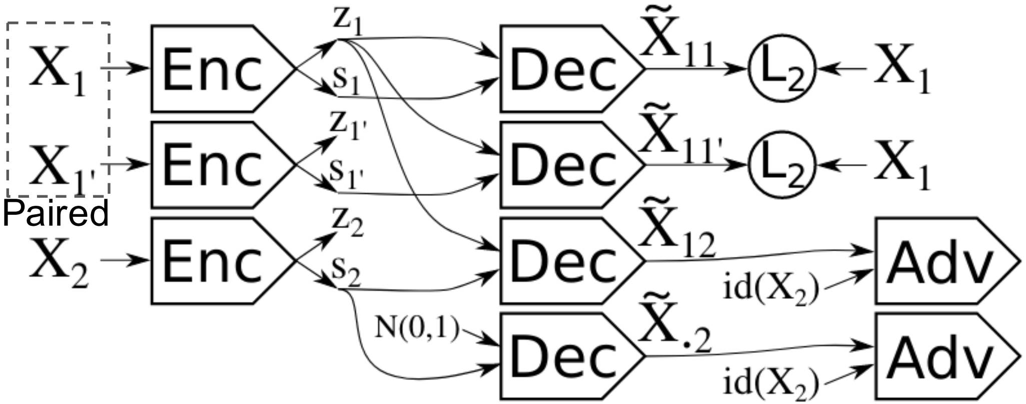


Generation by Exemplars with Paired Training Data

- Using a pair of image for specifying the transformation
  - Increase diversity.
  - *But paired training data are **hard** to collect.*



Deep Visual Analogy-Making, NIPS'15



Disentangling Factors of Variation, NIPS'16

X1 and X1' are required to have the same label, i.e., s1 == s1'.

Feature Space Interpolation Methods

	Generation by Exemplar	Unpaired Training data	Exploits Cyclic Loss
Deep Feature Interpolation	✗	✓	✗
InfoGAN	✗	✓	✗
Visual Analogy-Making	✓	✗	✗
Disentangling Factors of Variation	✓	✗	✓
CycleGAN	✗	✓	✓
GeneGAN	✓	✓	✓

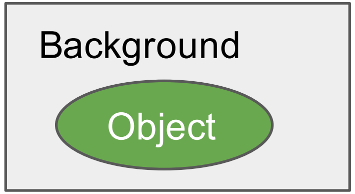
GeneGAN Training Data

- A positive set and a negative set
  - need not be paired

	Glasses	Hair	Lighting	Smiling
Positive	Eyeglass/sun glasses	Bangs	Side/Up/Down	Smiling
Negative	No glasses	Bald/Receding Hairline	Frontal lighting	Not smiling

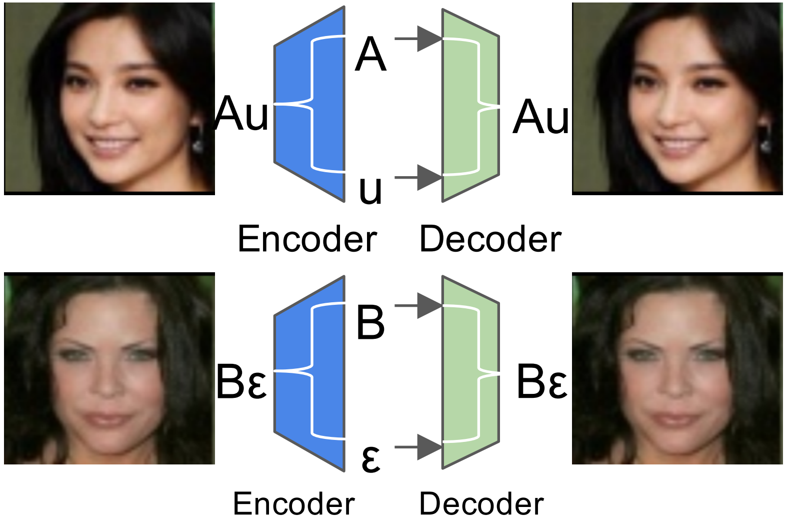
GeneGAN components: Encoder and Decoder

- Encoder: disentangle the object (smiling) from the background (face). Object can be abstract.
- Decoder: inverse of Encoder

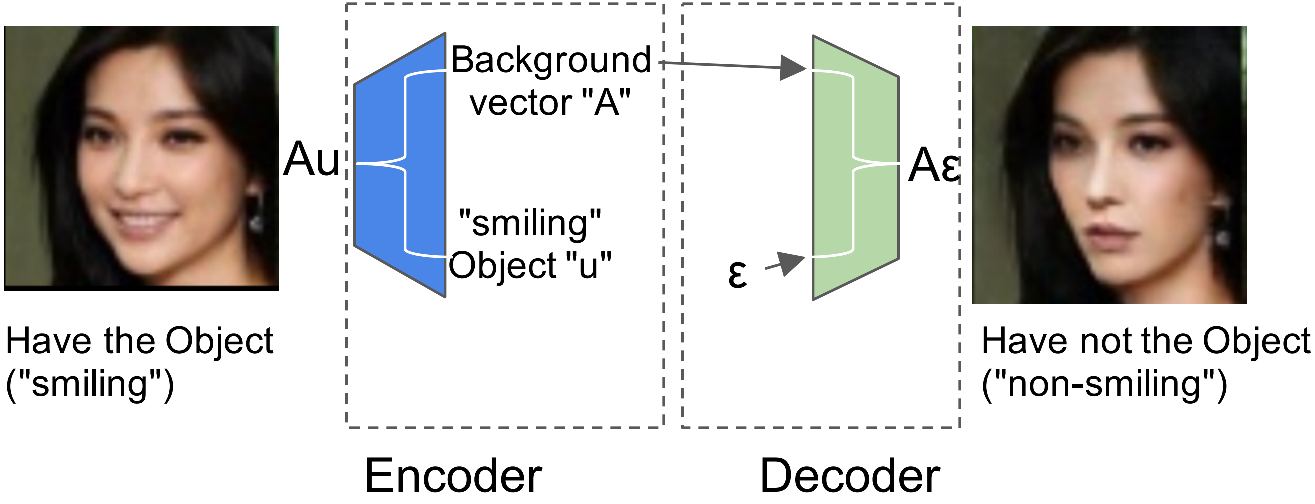


Have the Object (positive)

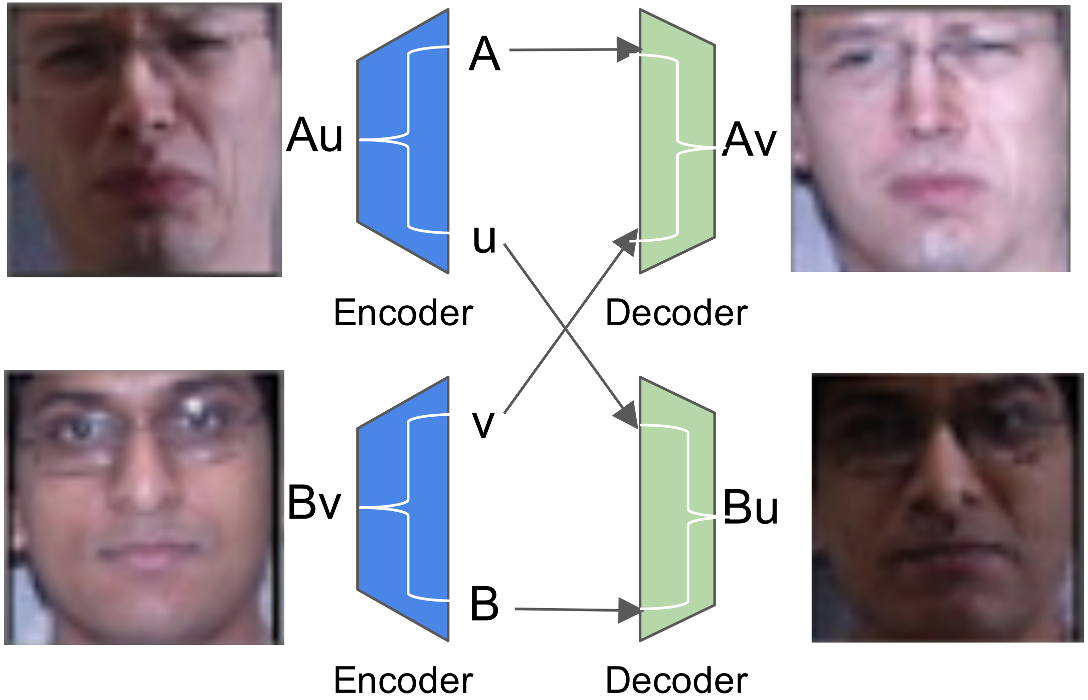
Have not the Object (negative)



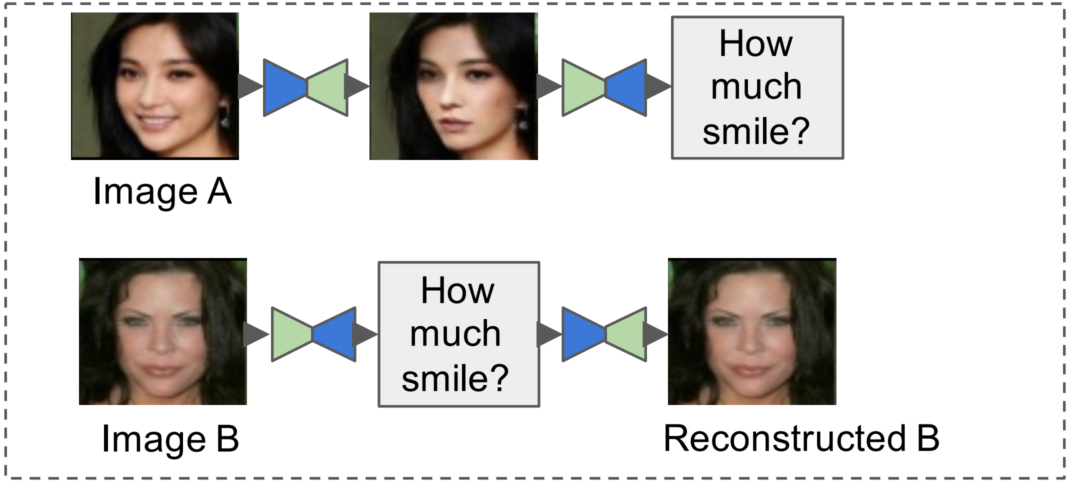
GeneGAN Usage: Object Removal



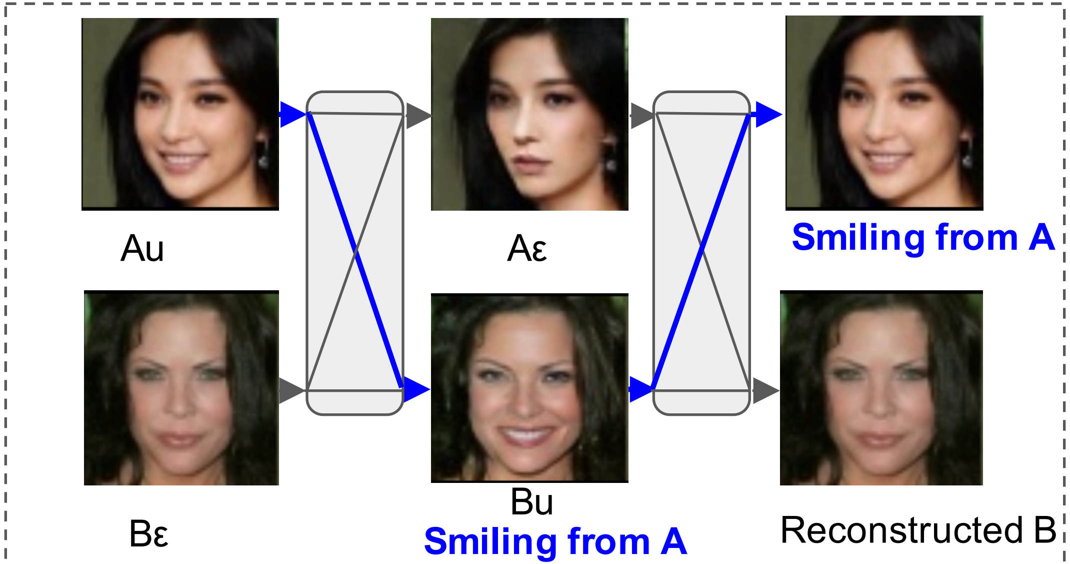
GeneGAN Usage: Swapping Objects



Underdetermined
CycleGAN pattern

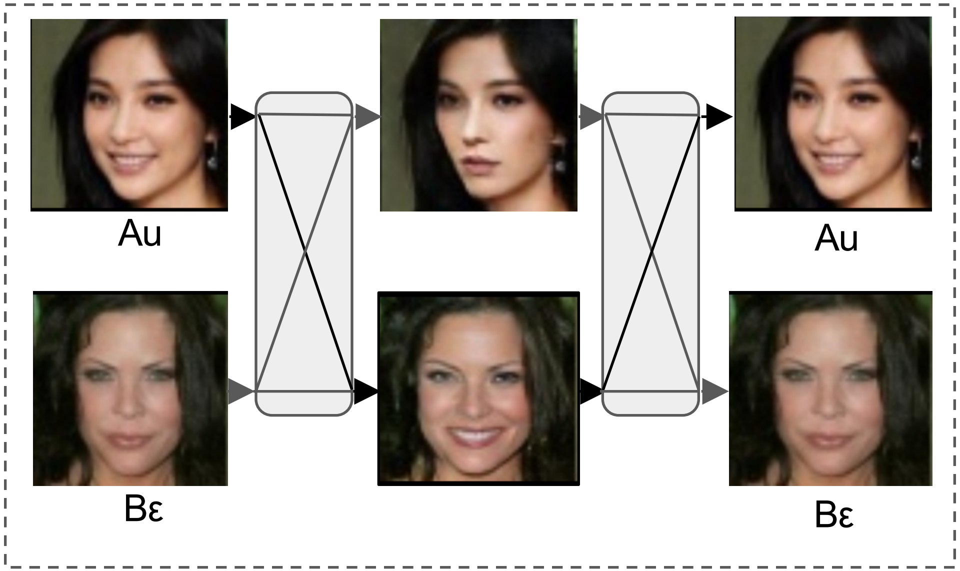


Information Preserving
GeneGAN pattern

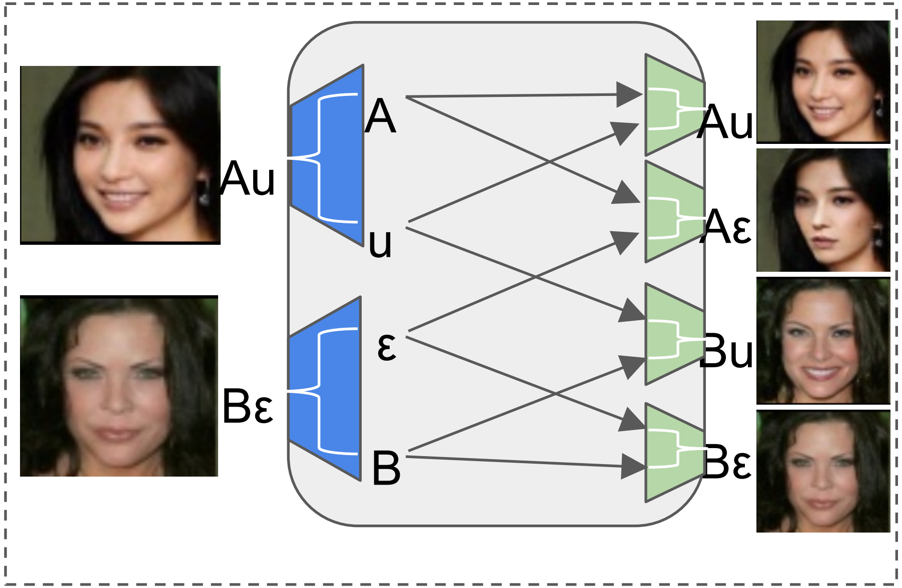


Shorten the Cycle to help Training

- Lift the grandchildren to be children

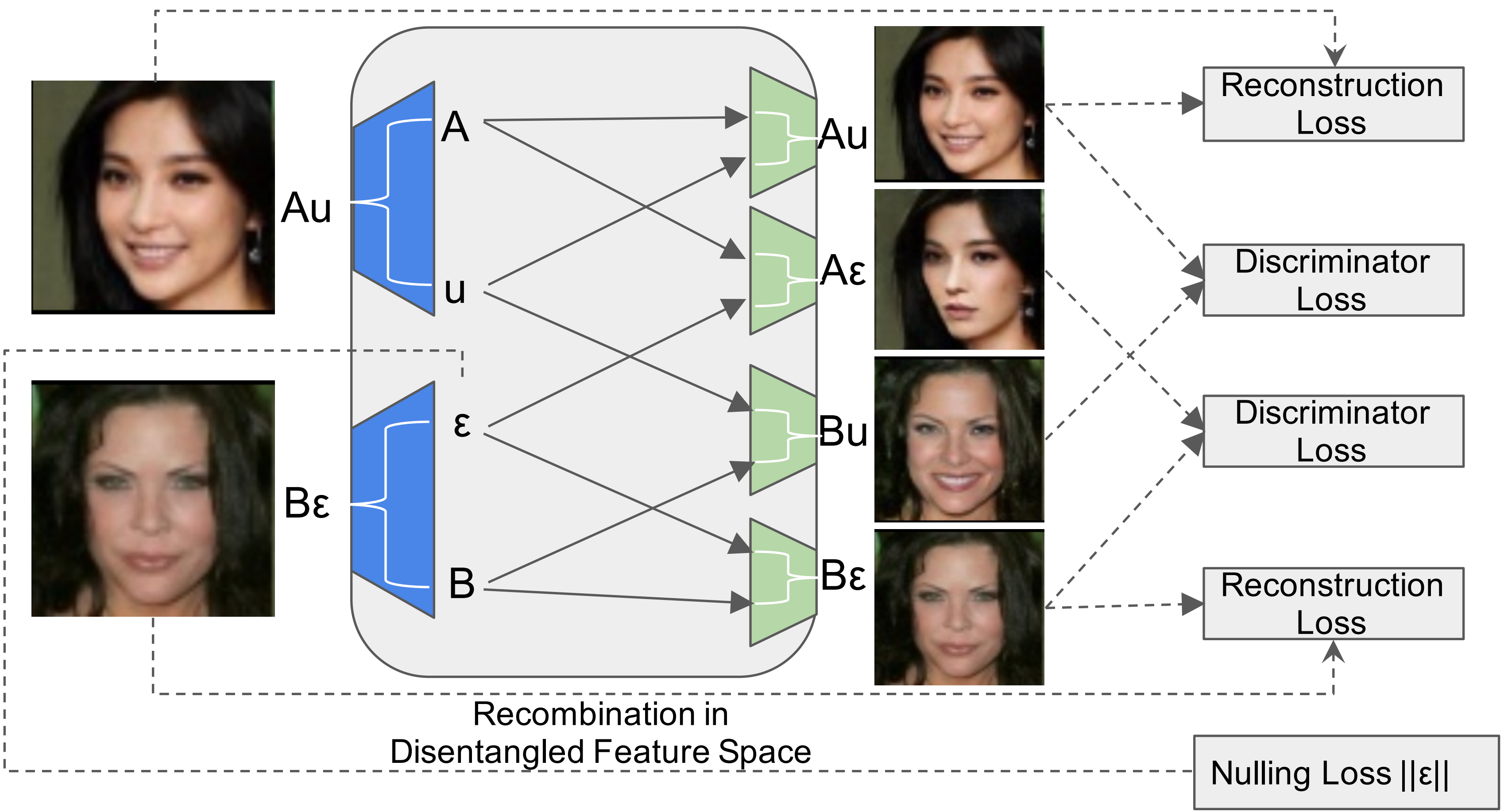


Two generations



Only one "generation", less distortions.

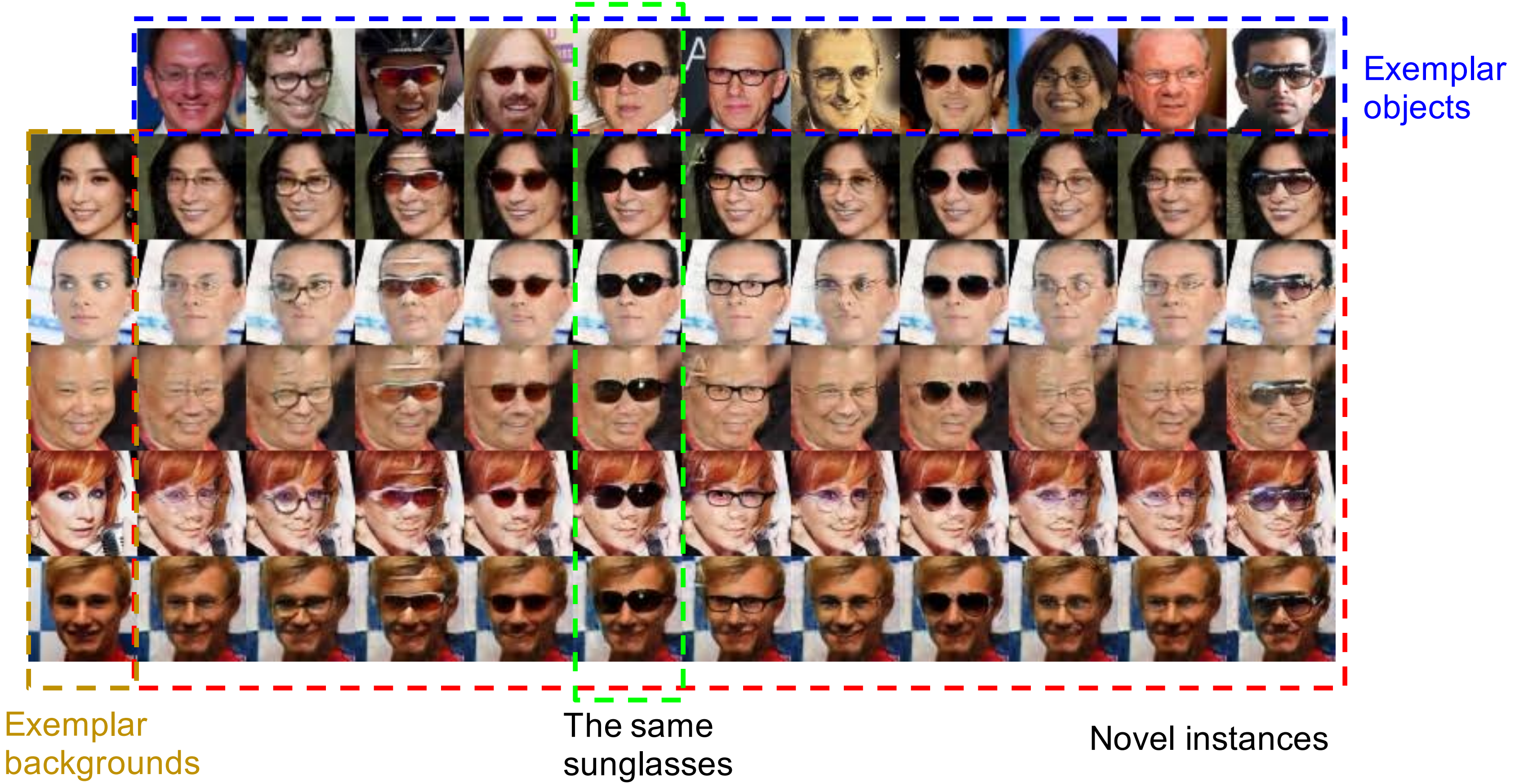
GeneGAN Training Diagram



Mechanism

- Constraints
  - Discriminator loss used in Adversarial Training
    - The background output of encoder will not contain smiling information, as "Bε" is not smiling
    - "u" contains the smiling information. As "Bu" is smiling.
  - Nulling loss
    - the object output of encoder will not contain background information, as "ε" can replace it without problem.
  - Reconstruction loss
    - Decoder and Encoder are inverse to each other
    - "A" contains background information, as Decoder can recreate "Au" from "A" and "u"

Experiments: Diversity from Exemplars



Swapping Attributes: Diversity of Smiles

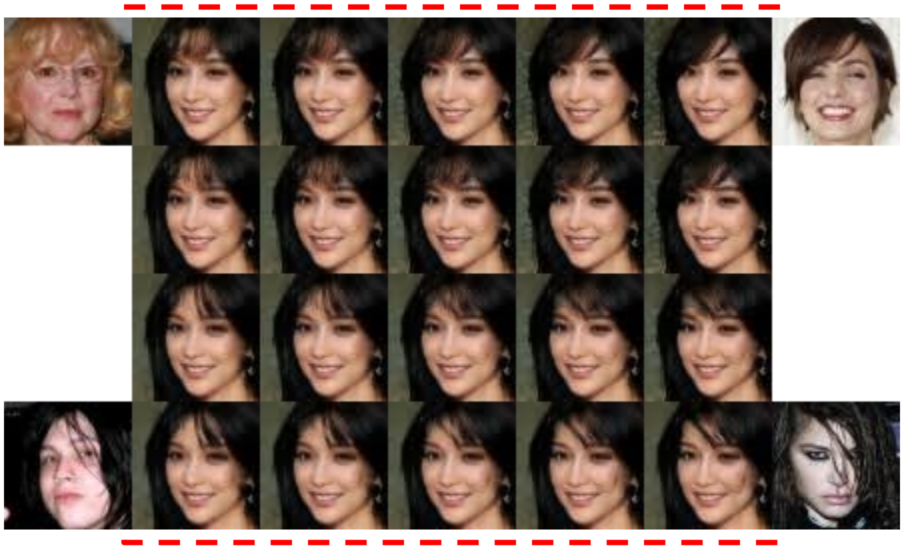
Can tell a smile by the mouth, and sometimes by eyes.

Object Subspace

- Multidimensional representation of hair

Interpolation in Object Subspace

Check the directions of the hairs.



Bi-linearly interpolated

ε instance

Conclusion & Future Work

- Disentangle the factors in feature space
  - Feature space = object space + background space
- Only require unpaired training data
  - Two unpaired image: positive and negative
- Usage cases
  - For single input, can output disentangled object code and background code.
  - For two inputs that both contain objects, can swap the objects in them. The objects can be null.
  - Can interpolate the objects in feature space.
- Futre work
  - Investigate whether more complex crossbreeding patterns between more parents would allow further improvements

References

1. Yoshua Bengio, Grégoire Mesnil, Yann Dauphin, and Salah Rifai. Better mixing via deep representations. In Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013, pages 552–560, 2013. URL http://jmlr.org/proceedings/papers/v28/bengio13.html.
2. Scott E. Reed, Yi Zhang, Yuting Zhang, and Honglak Lee. Deep visual analogy- making. In Advances in Neural Information Processing Systems 28: Annual Con- ference on Neural Information Processing Systems 2015, December 7-12, 2015, Mon- treal, Quebec, Canada, pages 1252–1260, 2015. URL http://papers.nips. cc/paper/5845-deep-visual-analogy-making.
3. Paul Upchurch, Jacob R. Gardner, Kavita Bala, Robert Pless, Noah Snavely, and Kil- ian Q. Weinberger. Deep feature interpolation for image content changes. CoRR, abs/1611.05507, 2016. URL http://arxiv.org/abs/1611.05507.
4. Michaël Mathieu, Junbo Jake Zhao, Pablo Sprechmann, Aditya Ramesh, Yann LeCun: Disentangling factors of variation in deep representation using adversarial training. NIPS 2016: 5041-5049
5. Xi Chen, Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, Pieter Abbeel: InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. NIPS 2016: 2172-2180
6. Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, Cordelia Schmid: Learning from Synthetic Humans. CoRR abs/1701.01370 (2017)
7. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to- image translation using cycle-consistent adversarial networks. CoRR, abs/1703.10593, 2017. URL http://arxiv.org/abs/1703.10593.

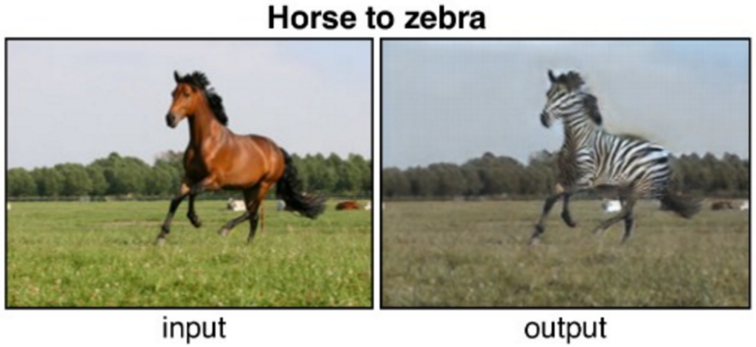
More in the paper.

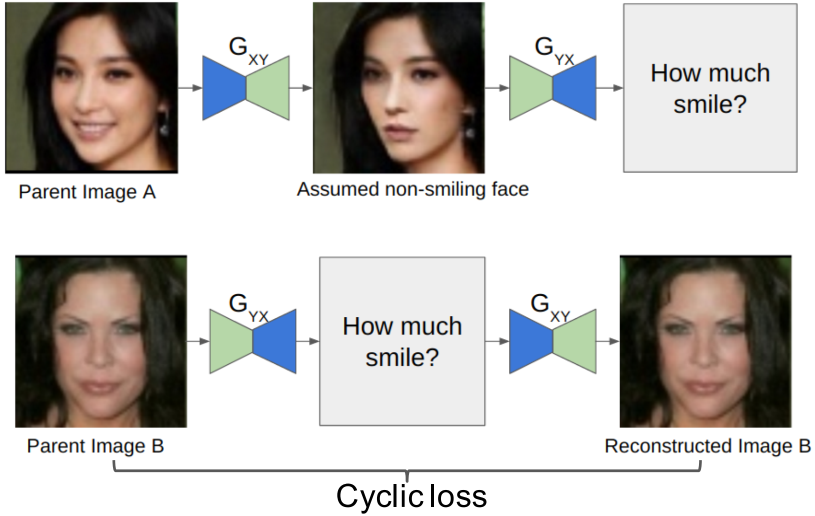
Backup after this slide

Github: https://github.com/Prinsphield/GeneGAN

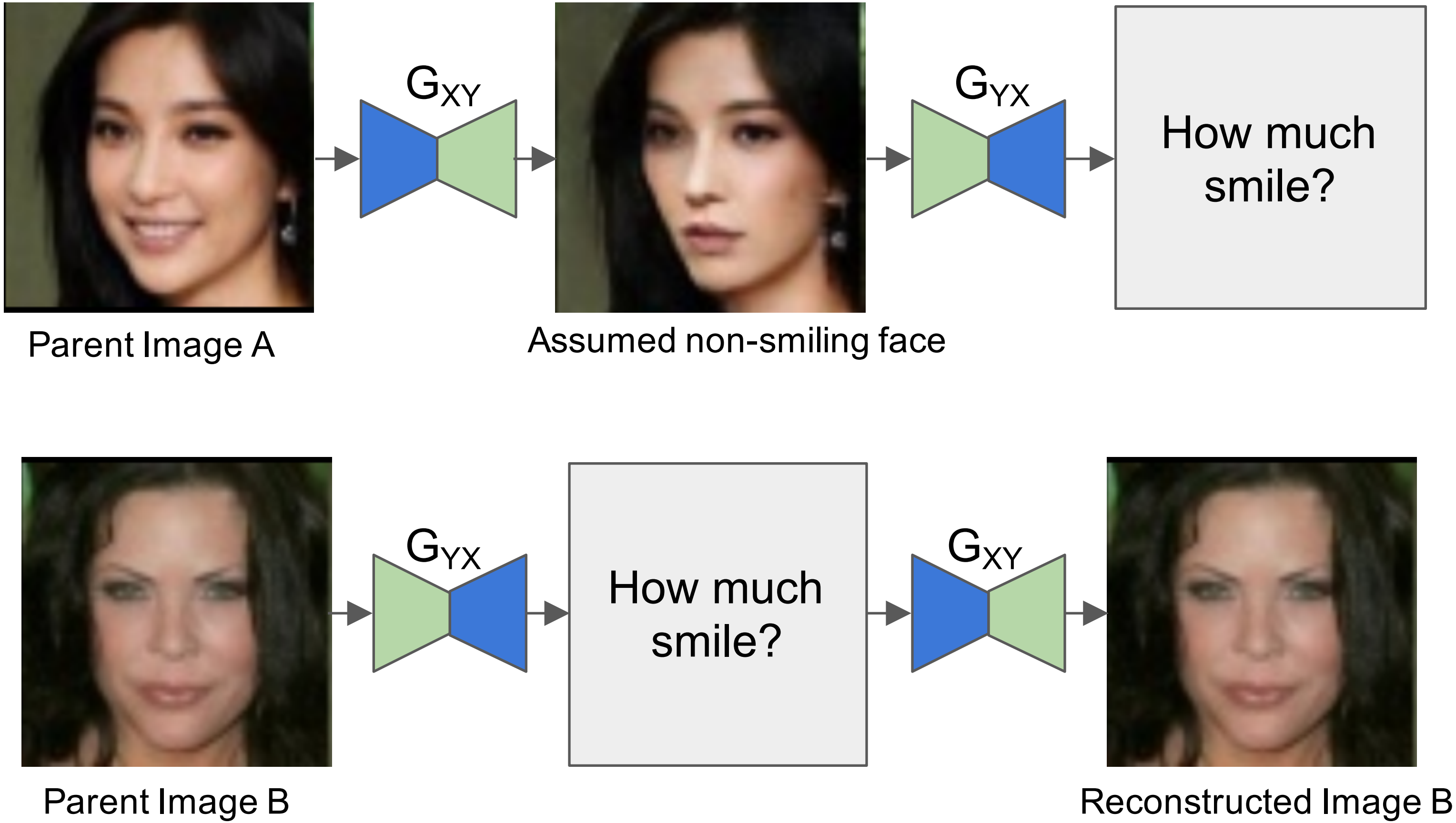
CycleGAN/DiscoGAN and Object Transfiguration

- Pros
  - Learn from Unpaired Data
  - Exploits Cyclic loss to stabilize training
- Cons
  - Backgrounds change when transforming objects
  - Under-determination problem
    - non-smiling is well defined. But smiling's have different levels and styles.

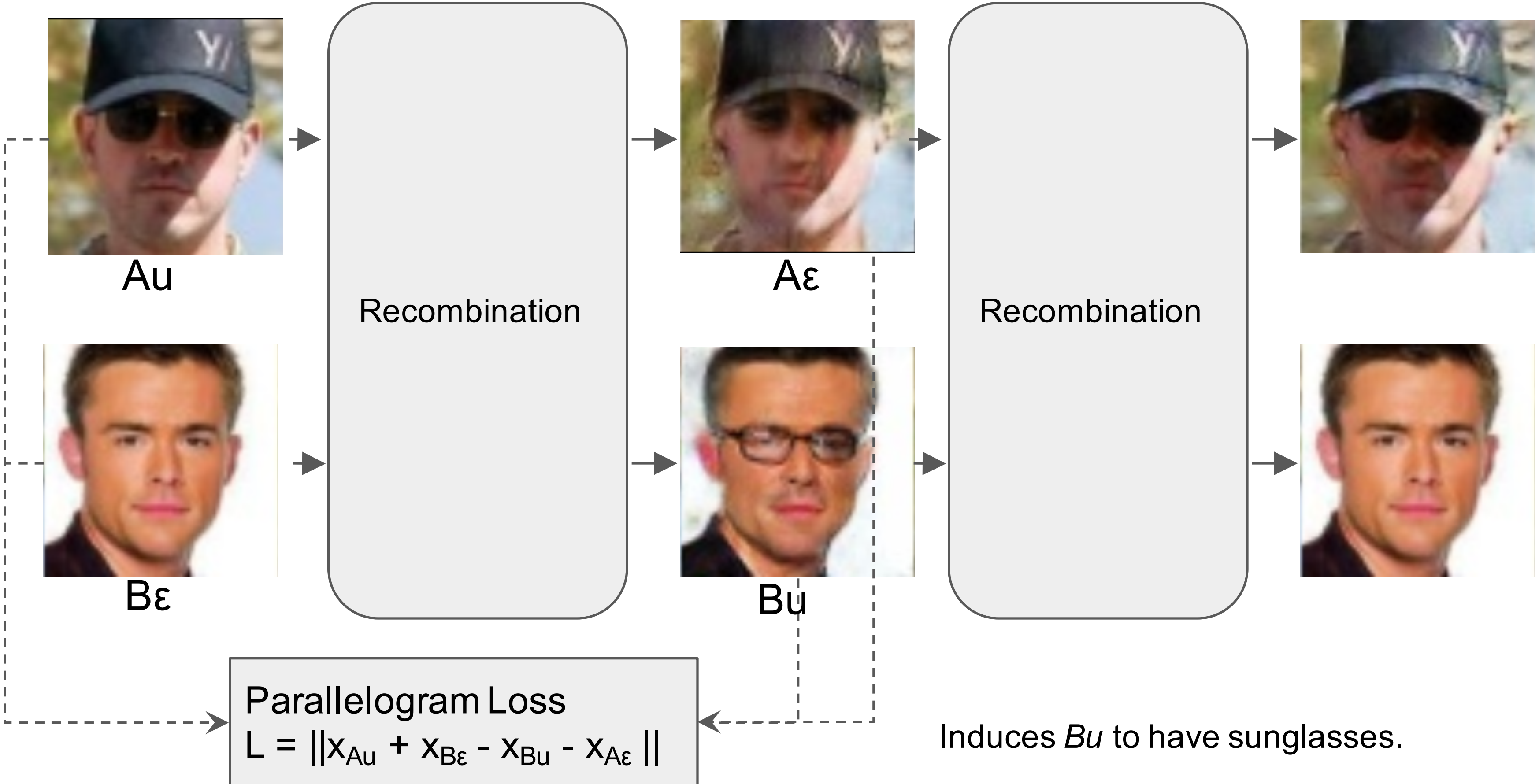




Underdetermination Problem



Parallelogram loss



Induces *Bu* to have sunglasses.