CSC 2547: Machine Learning for Vision as Inverse Graphics

Anthony Bonner

www.cs.toronto.edu/~bonner

Paper Presentations

- Each week will focus on one or two topics, as listed on the course web page (soon).
- You can vote for your choice of topic/week (soon).
- I will assign you to a week (soon).
- Papers on each topic will be listed on the course web page.
- If you have a particular paper you would like to add to the list, please let me know.

Paper Presentations

- Goal: high quality, accessible tutorials.
- 7 weeks and 40 students = 6 students per week and 20 minutes per student (including questions).
- 2-week planning cycle:
  - 2 weeks before your presentation, meet me after class to discuss and assign papers.
  - The following week, meet the TA for a practice presentation (required).
  - Present in class under strict time constraints.

Team Presentatations

- Papers may be presented in teams of two or more with longer presentations (20 minutes per team member).
- Unless a paper is particularly difficult or long, a team will be expected to cover more than one paper (one paper per team member).
- A team may cover one of the listed papers and one or more of its references (but see me first).

Tentative Topics

- Discriminative and generative approaches
- Capsule networks
- Point Nets and 3D point clouds
- Group symmetries and equivariance
- Visual attention and transformers
- Inferring 3D structure
- Inferring part-whole relationships
- Self-supervised and contrastive learning
- Adversarial learning

Project Ideas

- Improve upon the work in a paper
  - Even a small improvement is OK
- For example,
  - Make a generative model conditional
  - Disentangle (some) latent variables
  - Adapt a method to new circumstances
    - Different kinds of data
    - Missing or noisy data
  - Make a supervised method semi-supervised

Project Ideas

- Examples (continued)
  - Modify the cost function
    - Introduce learnable parameters into a cost function
    - Use an adversarial cost
    - Try a variation on KL divergence
  - Modify the latent priors
    - Make the prior learnable
    - Do not assume Gaussianity
  - Modify the variational assumptions
    - Do not assume complete independence
    - Do not assume Gaussianity

Project Ideas

- Implement and compare different methods for the same problem (e.g., different methods for inferring 3D structure)
  - Clearly and succinctly describe each method
  - Clearly articulate their differences
  - Describe their strengths and weaknesses
  - Ideally, include experiments highlighting the differences between the methods on realistic problems.

Project Considerations

- Is your idea sensible?
- Can you download all the necessary data?
- Do you have the computational resources (GPUs)?
- Do you have time to complete it?
- Start by duplicating the results in the paper (if the paper gives enough details).

Project Dates

- Proposal due February 23
  - about 2 pages
  - include preliminary literature search
- Project presentations: March 30 and April 6
  - about 5 minutes per student (like “spotlight presentations” at a conference)
- Project due: April 18 (tentative)
  - project report (4-8 pages) and code

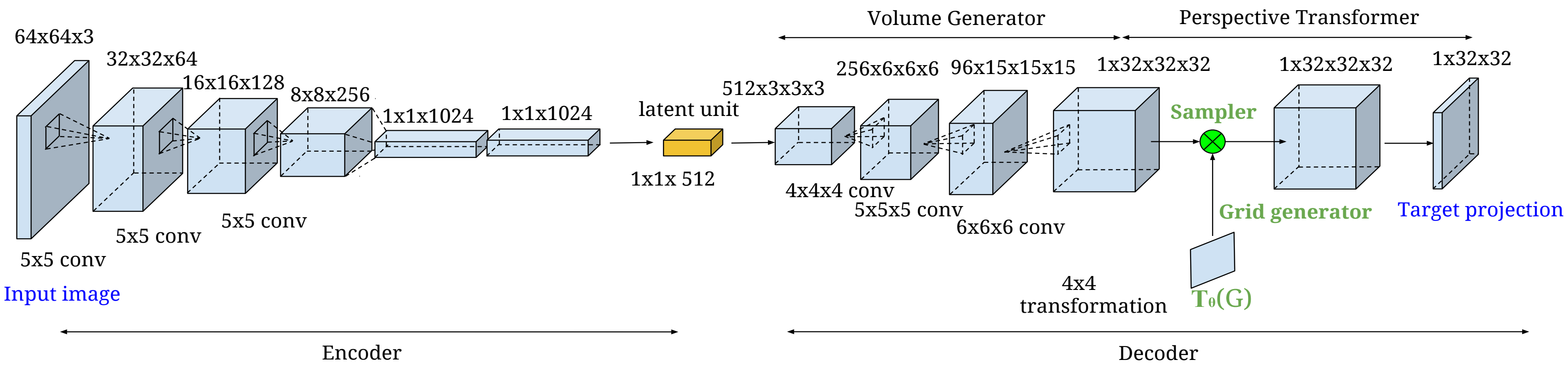
Generative Approaches

- Given a scene, s, a graphics program, G, produces an image, G(s).
- Given an image, x, find s such that G(s) ≈ x
- More generally, find P(s|x),.
- P(s|x) is high when G(s) is close to x.

Variational Approximations

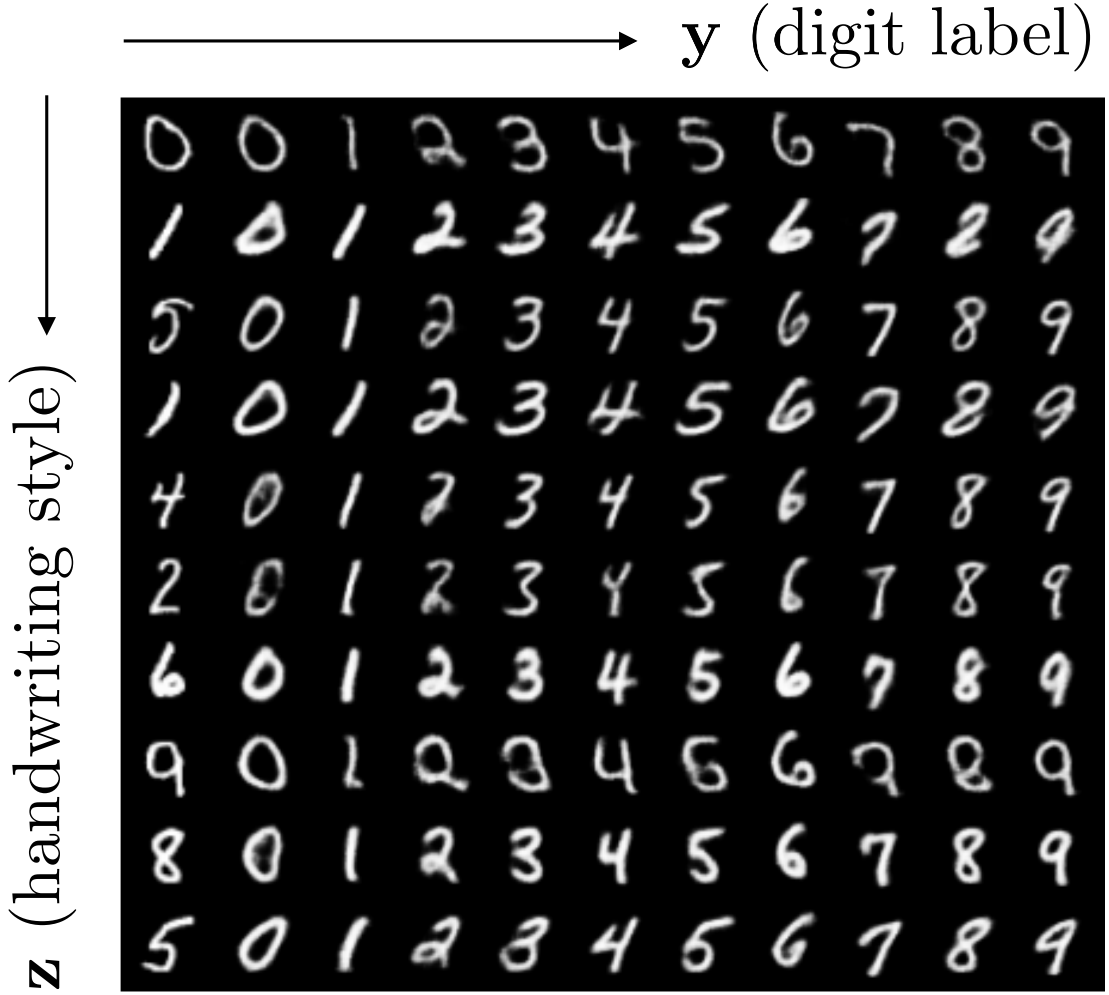
- Finding P(s|x) is intractable in general.
- Use variational approximations.
- Variational auto-encoders work very well.
- G can be a neural net that we learn (unsupervised).
- Computationally intensive.

Variational Autoencoders



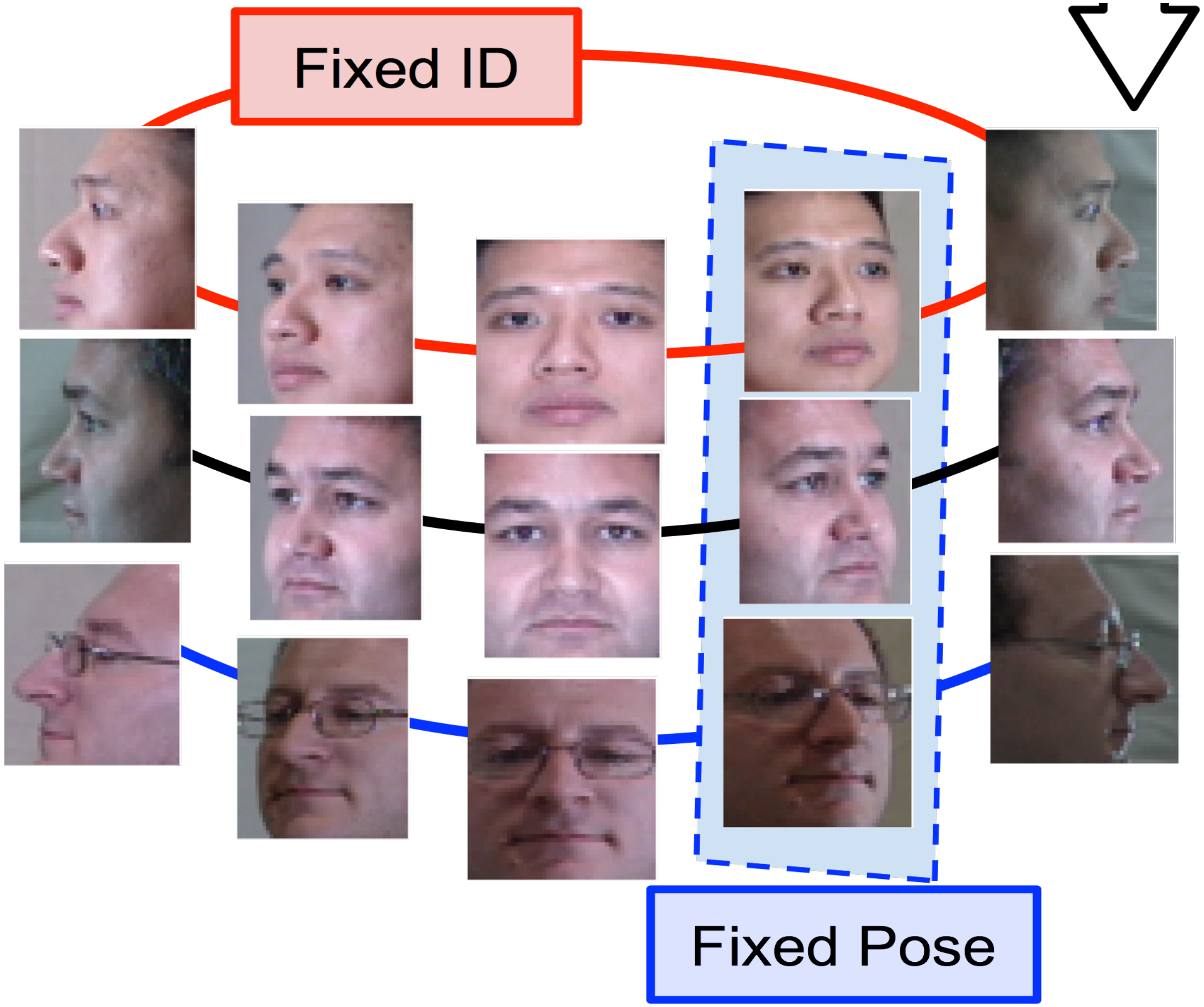
From Yan et al, *Perspective Transformer Nets*, arXiv 2017

Disentangled Representations



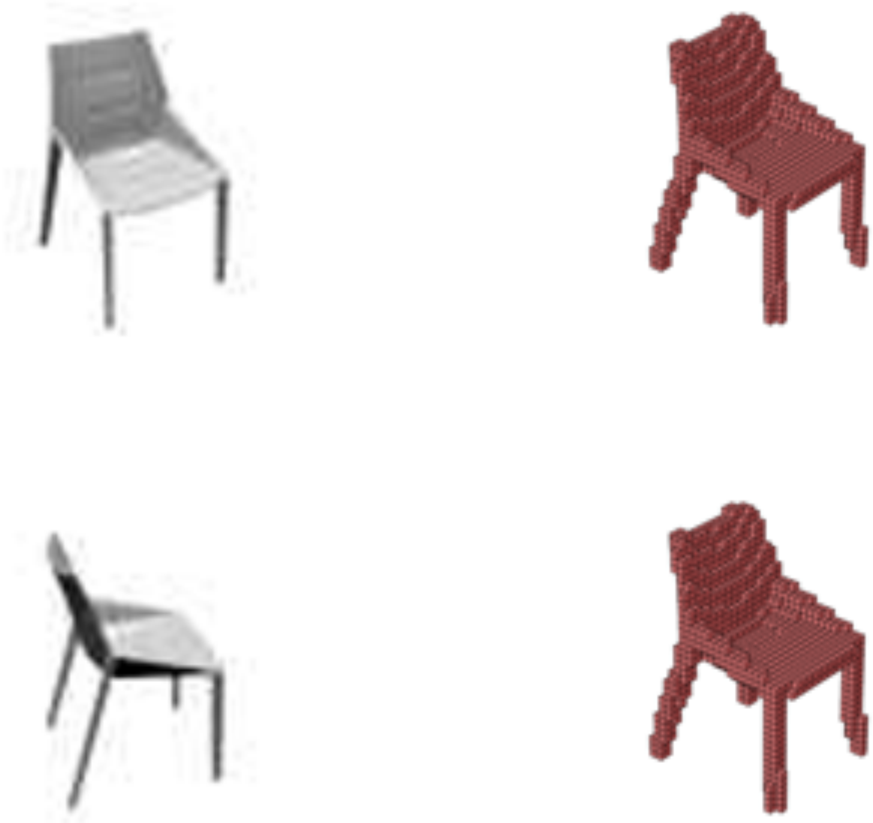
From Siddharth et al, *Semi-supervised Deep Generative Models*, NIPS 2017

Disentangled Representations



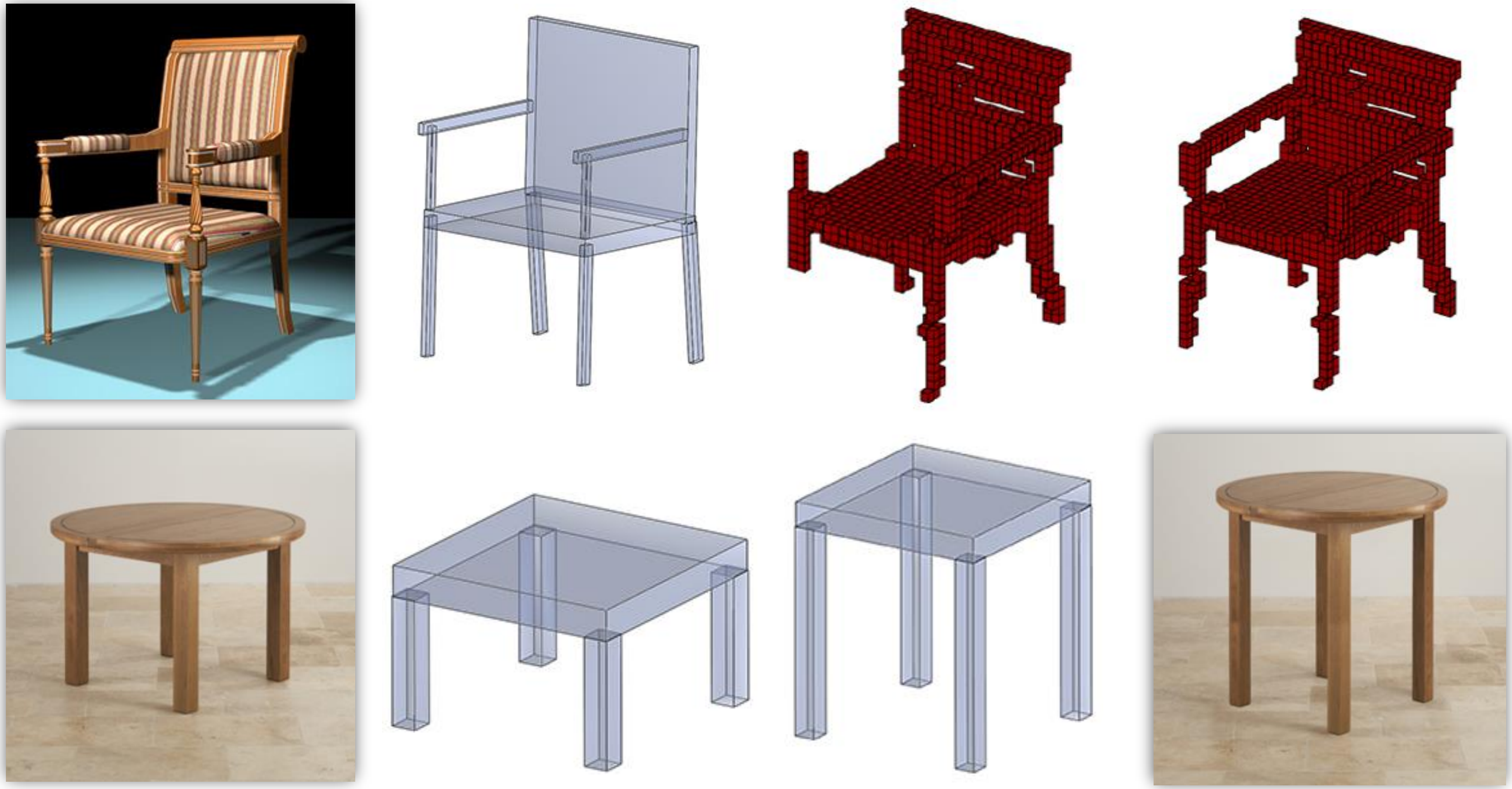
From Reed et al, *Learning to Disentangle Factors of Variation*, ICML 2014

Learning 3D Shape



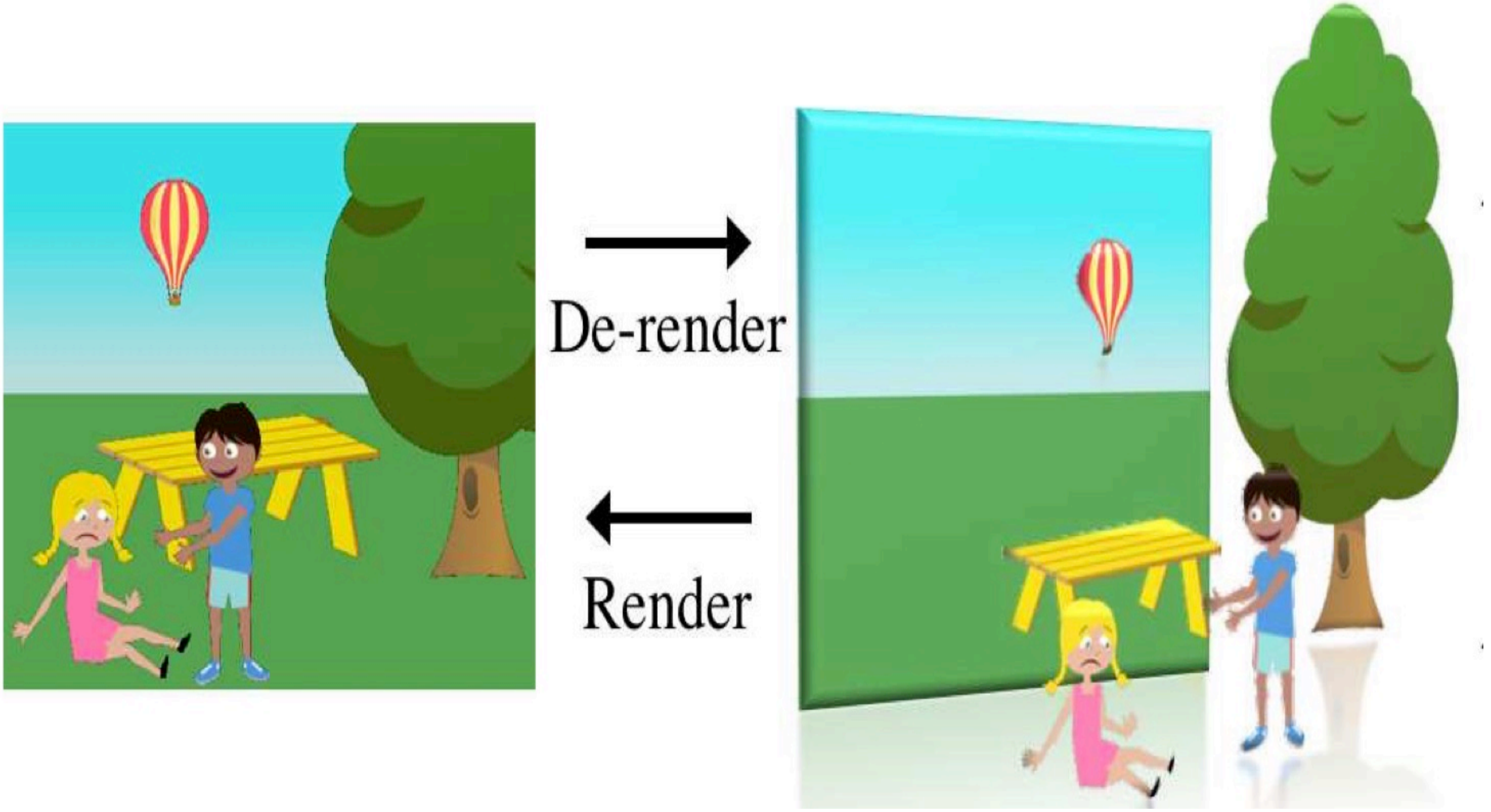
From Yan et al, *Perspective Transformer Nets*, arXiv 2017

Learning 3D Structure



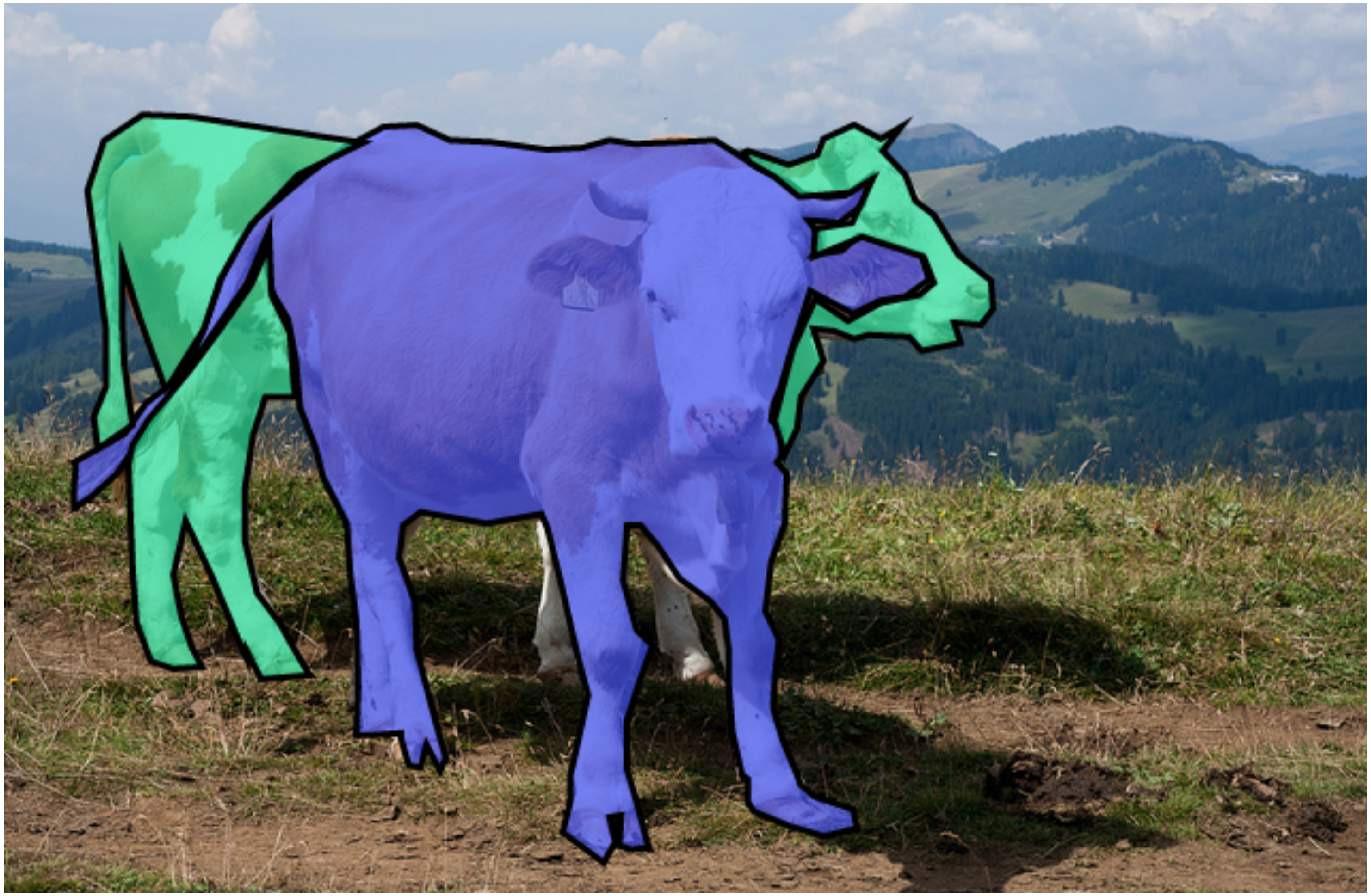
From Niu et al, *Im2Struct: recovering 3D Shape Structure*, CVPR 2018

Scene Understanding



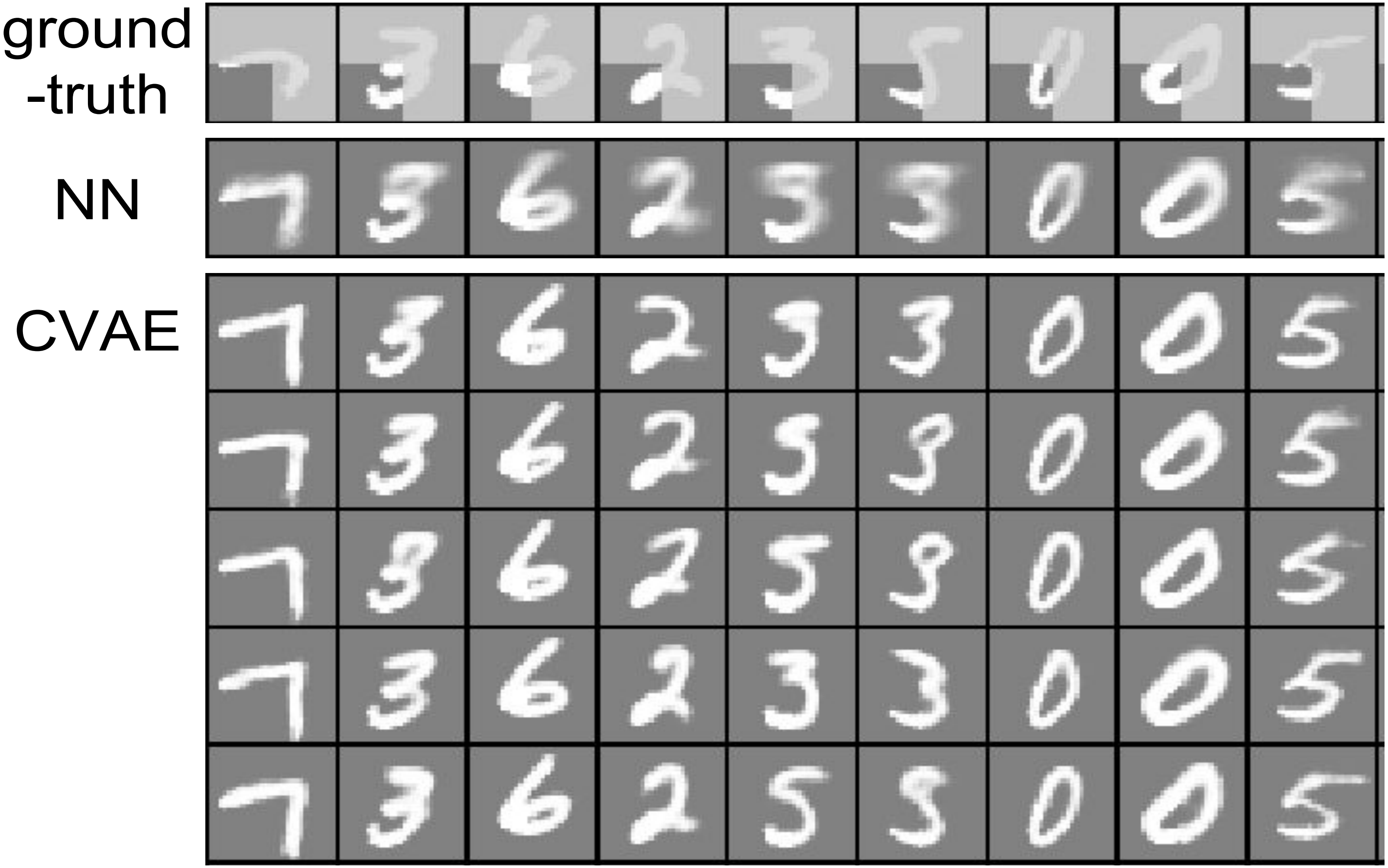
From Wu et al*, Neural Scene De-rendering*, CVPR 2017

Scene Understanding



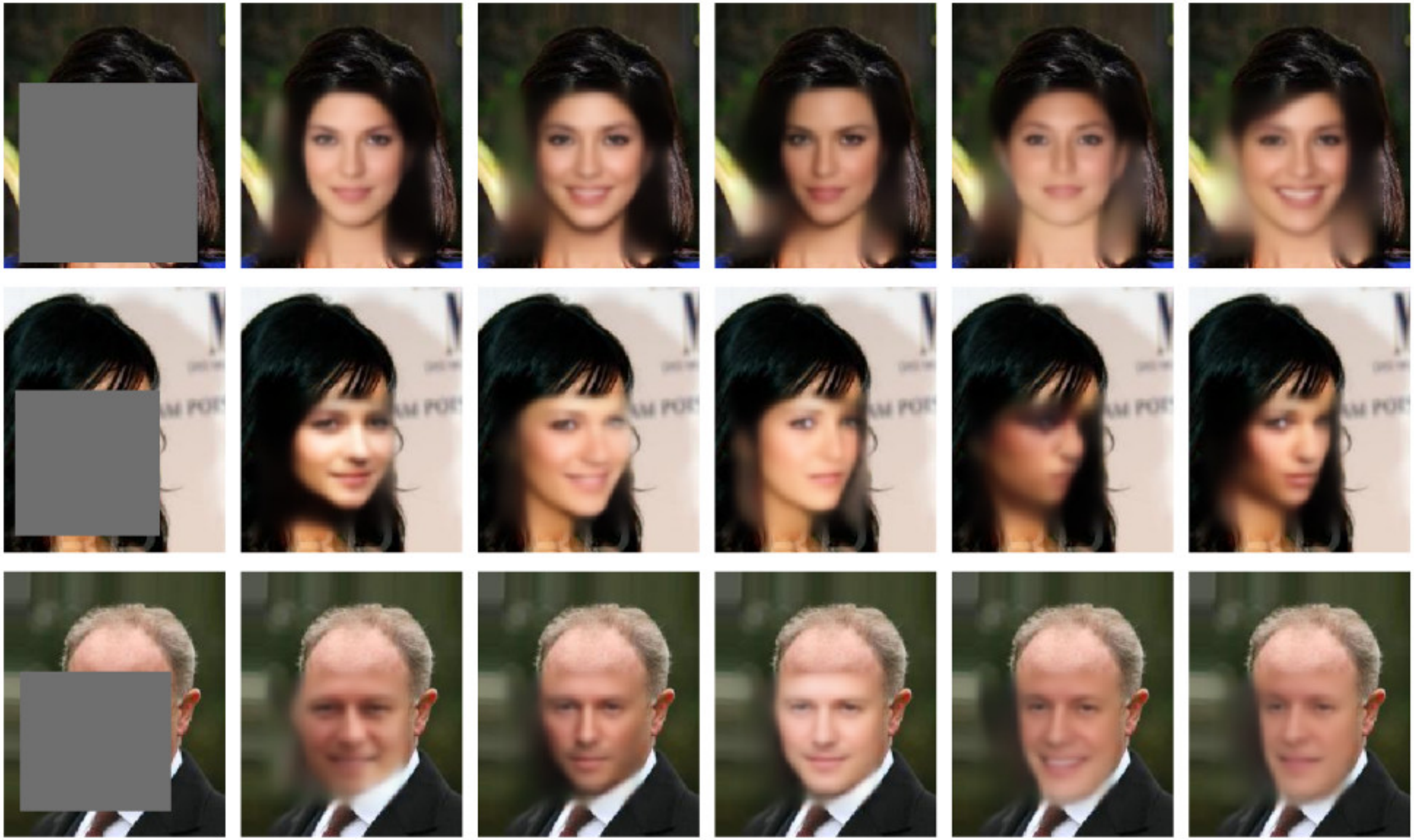
From Huang et al, *Occlusion Aware Generative Models*, ICLR 2016

Conditional Image Generation



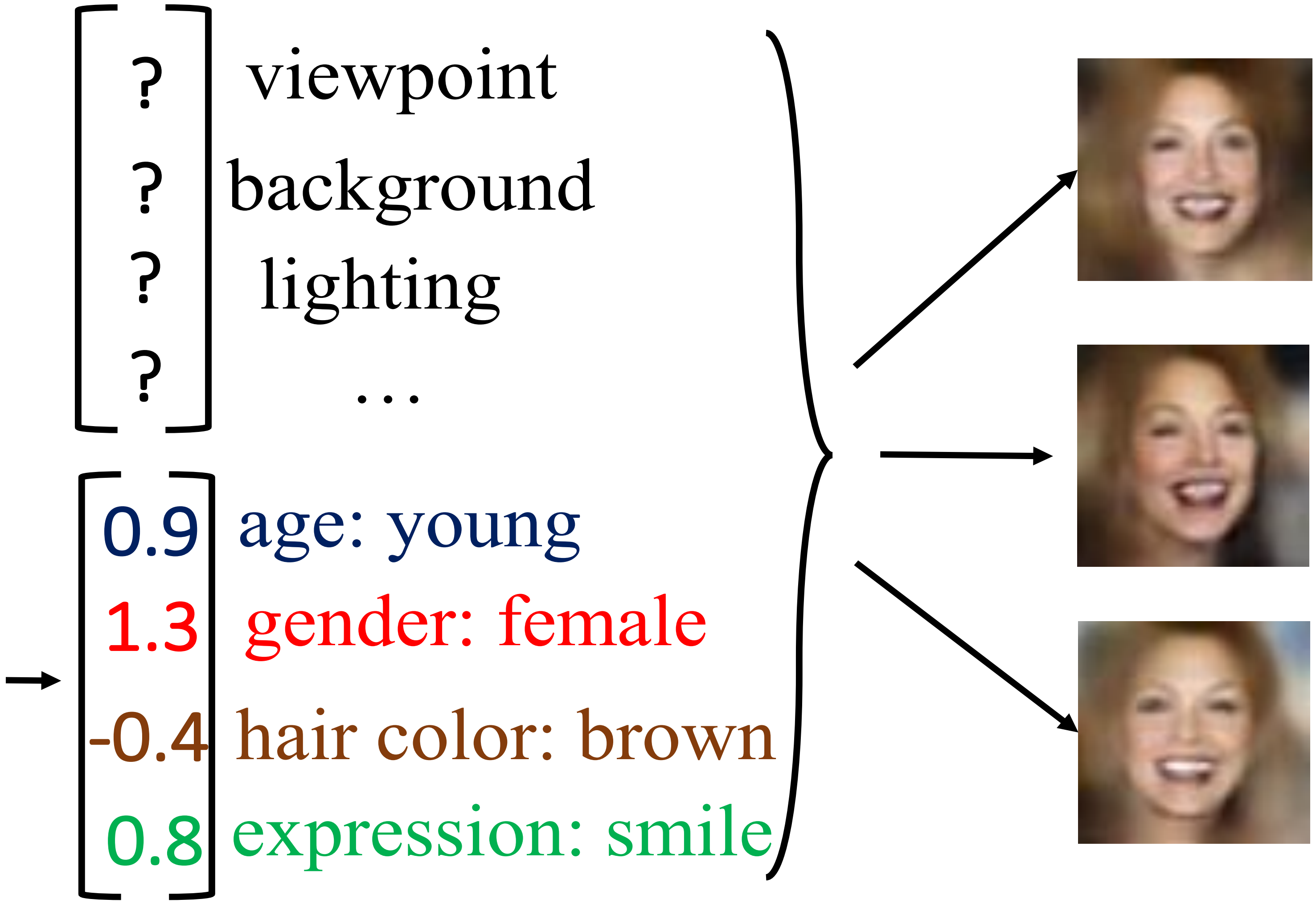
From Sohn et al*, Deep Conditional Generative Models*, NIPS 2015

Conditional Image Generation



From Ivanov et al, *Variational Autoencoder with Arbitrary Conditioning*, ICLR 2019

Attribute Conditioned Image Generation



? viewpoint

? background

? lighting

? …

0.9 age: young

1.3 gender: female

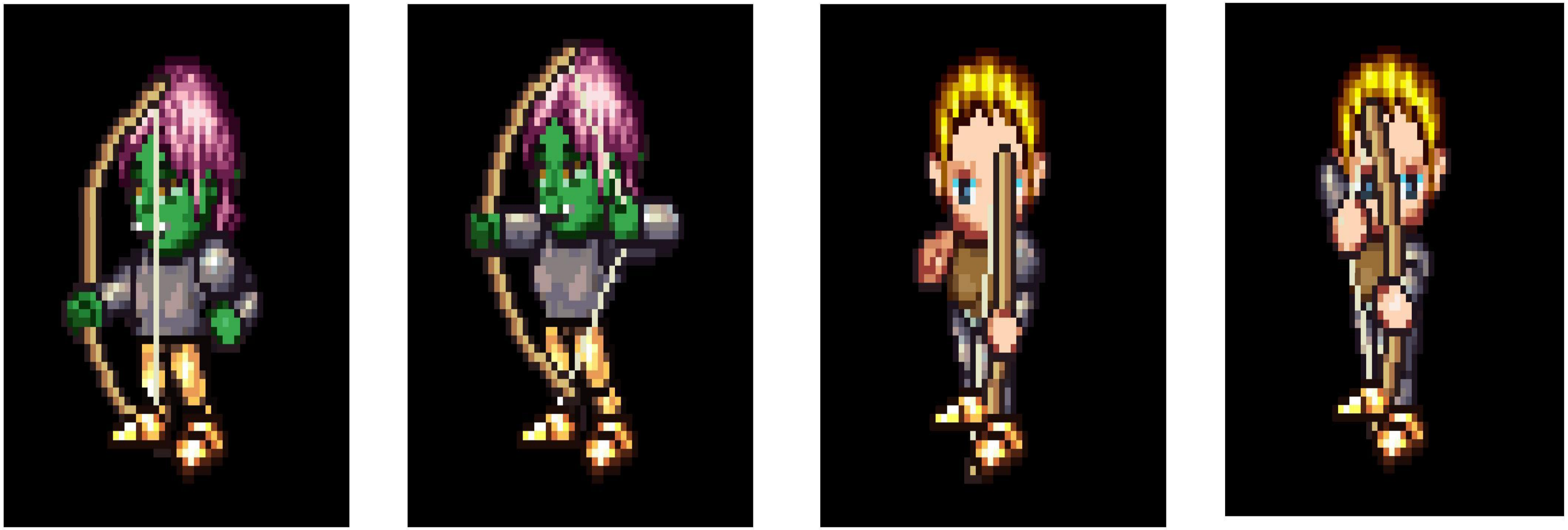
-0.4 hair color: brown

0.8 expression: smile

From Yan et al, *Attribute2Image: Conditional Image Generation*, arXiv 2016

Making Visual Analogies

- Given images A, B, C, generate image D so that D is to C as B is to A.



From Reed et al, *Deep Visual Analogy-Making*, NIPS 2015