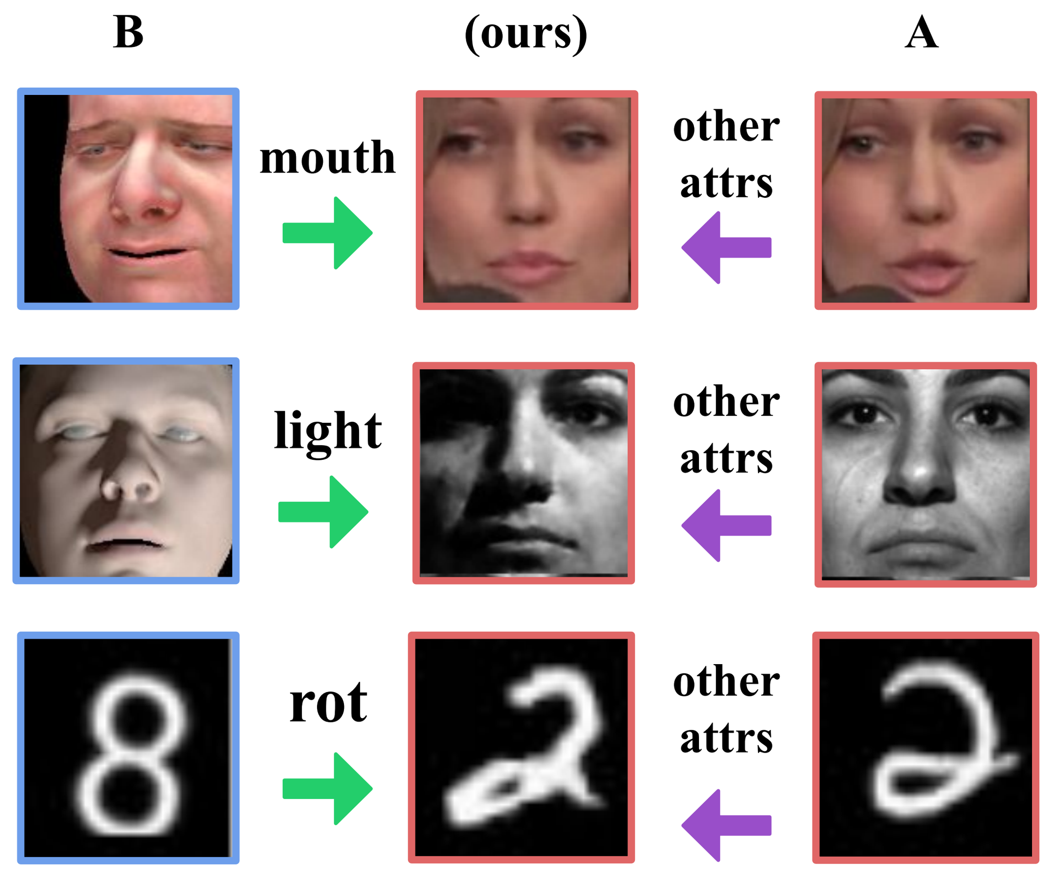
PuppetGAN: Cross-Domain Image Manipulation by Demonstration

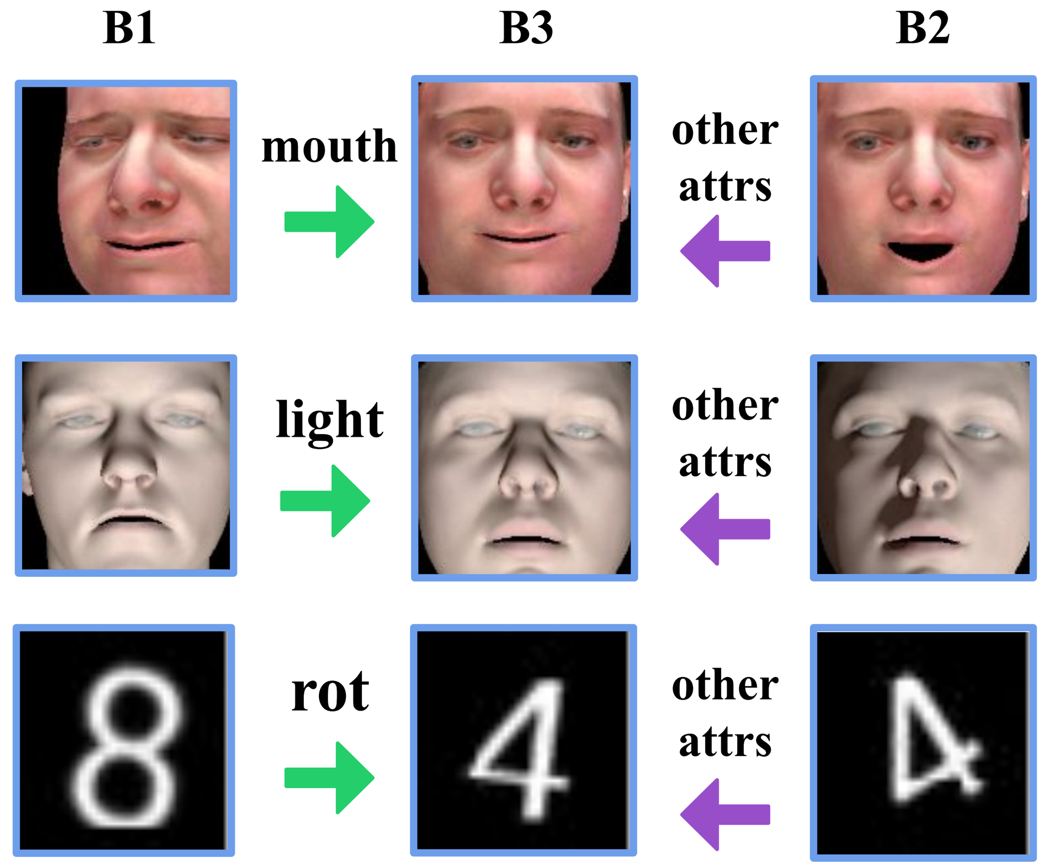
Ben Usman [1,2], Nick Dufour [1], Kate Saenko [2], Chris Bregler [1]
[1] Google AI [2] Boston University

Task

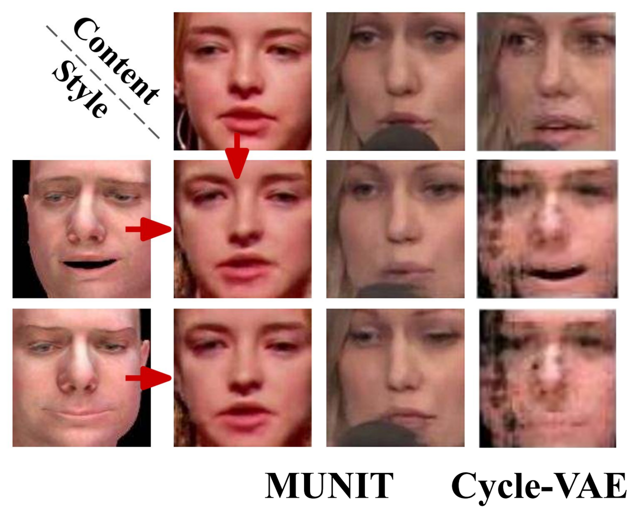
Our model can manipulate a **single** specific **attribute** of a **real** image **A** using a **synthetic** reference **B**.



It is trained exclusively on ***synthetic demonstrations*** and unlabeled real images.



Related Work



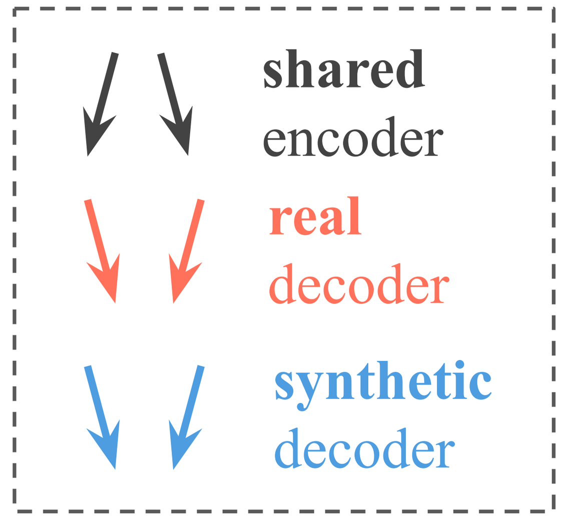
Unsupervised Cross-Domain Adaptation produces entangled representations (e.g. CycleGAN).

Unsupervised Cross-Domain Disentanglement might disentangle wrong attributes (e.g. MUNIT).

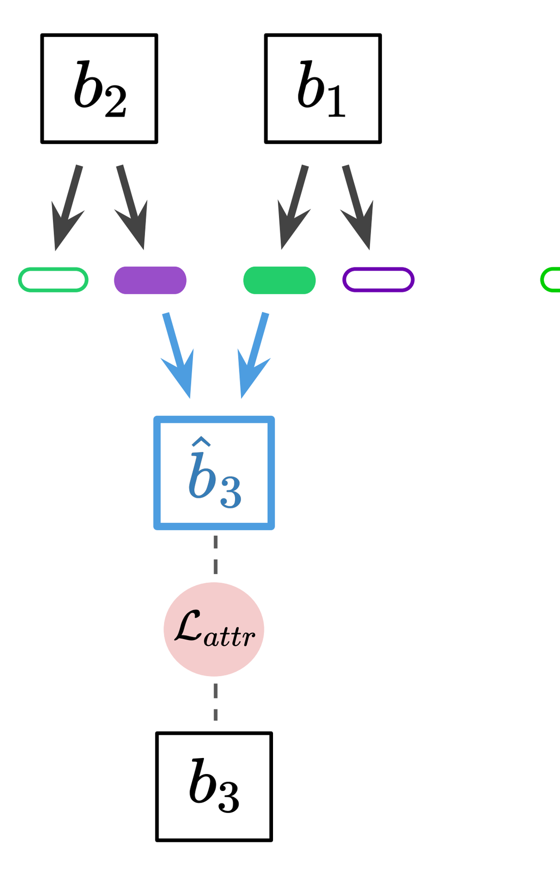
Supervised Single-Domain Disentanglement fails to generalize to a different domain (e.g. InfoGAN, Cycle-Consistent VAE).

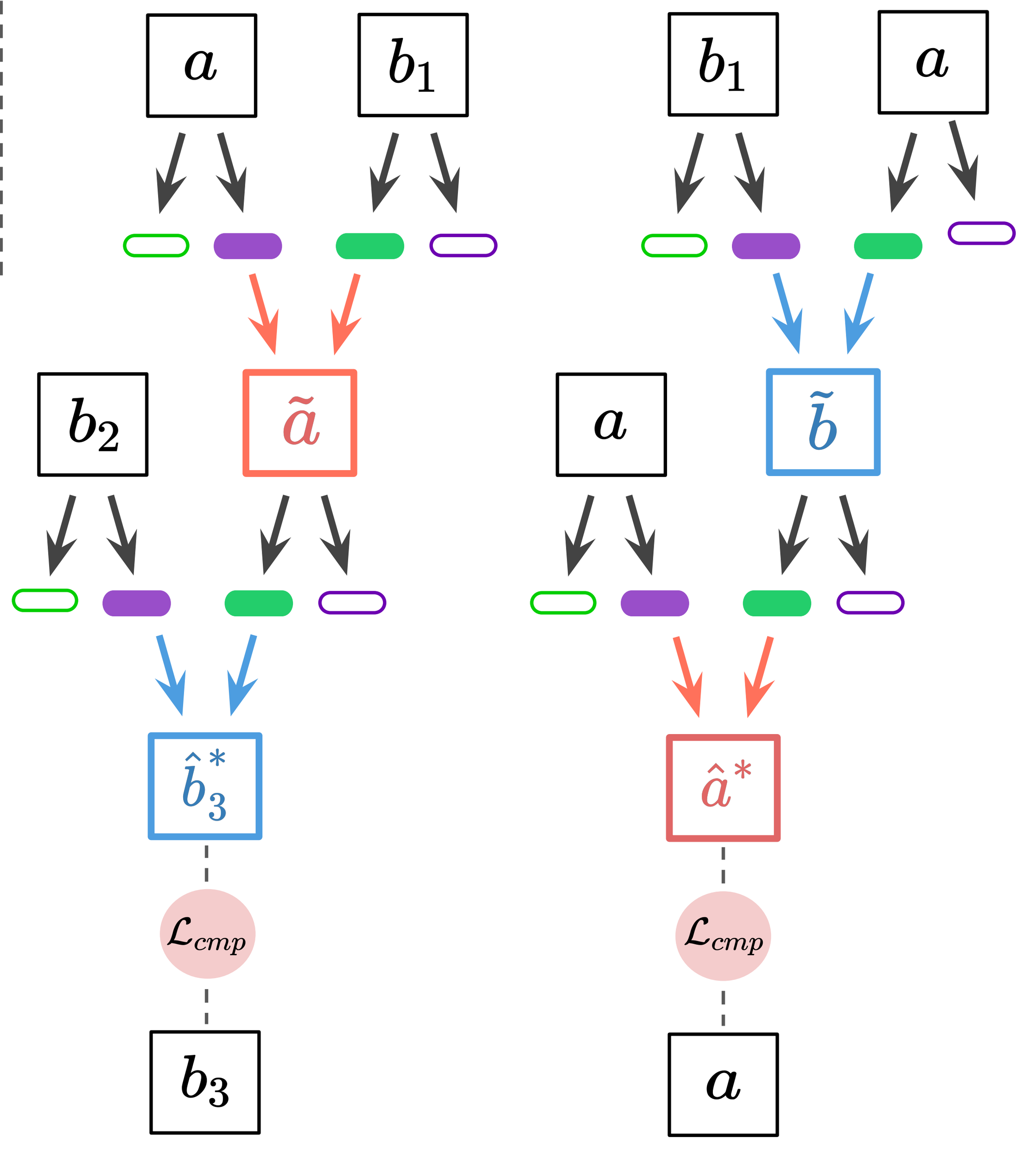
Existing Supervised Cross-Domain Disentanglement Methods yield degenerate solution that ignore parts of the learned embeddings (e.g. DiDA). The PuppetGAN model is more resilient against such degenerate solution.

Method



The proposed **compositional consistency loss** ensures that **both** embedding components are **used** during image generation.





(a) supervised disentanglement (b) compositional consistency

Results

Other techniques we used:

- reconstruction and cycle losses
- adversarial domain alignment
- regularization with instance noise

Other findings reported in the paper:

- disentanglement quality metrics
- failure case analysis
- input outlier robustness
- comparison to other models

ai.bu.edu/puppetgan

Model	Disentanglement quality (MNIST ⇄ Rendered Digits)							
	Size				Rotation			
	Acc ↑	r^{syn}_{attr} ↑	r^{syn}_{rest} ↓	V_{rest} ↓	Acc ↑	r^{syn}_{attr} ↑	r^{syn}_{rest} ↓	V_{rest} ↓
PuppetGAN	**0.73**	**0.85**	**0.02**	**0.02**	**0.97**	**0.40**	**0.11**	**0.01**
CycleGAN [28]	0.10	0.28	**0.06**	0.28	0.11	**0.54**	0.37	0.33
DiDA [2]	**0.71**	0.18	0.09	**0.02**	**0.86**	0.04	0.35	**0.02**
MUNIT [10]	**0.96**	0.06	0.09	**0.01**	**1.00**	0.00	0.15	**0.01**
Cycle-VAE [8]	0.17	**0.92**	0.16	**0.01**	0.29	**0.45**	**0.10**	**0.01**
PuppetGAN†	**0.64**	0.28	0.07	**0.01**	0.10	0.06	**0.04**	**0.01**

† larger discrepancy in attribute distributions between A and B ⇒ lower disentanglement quality