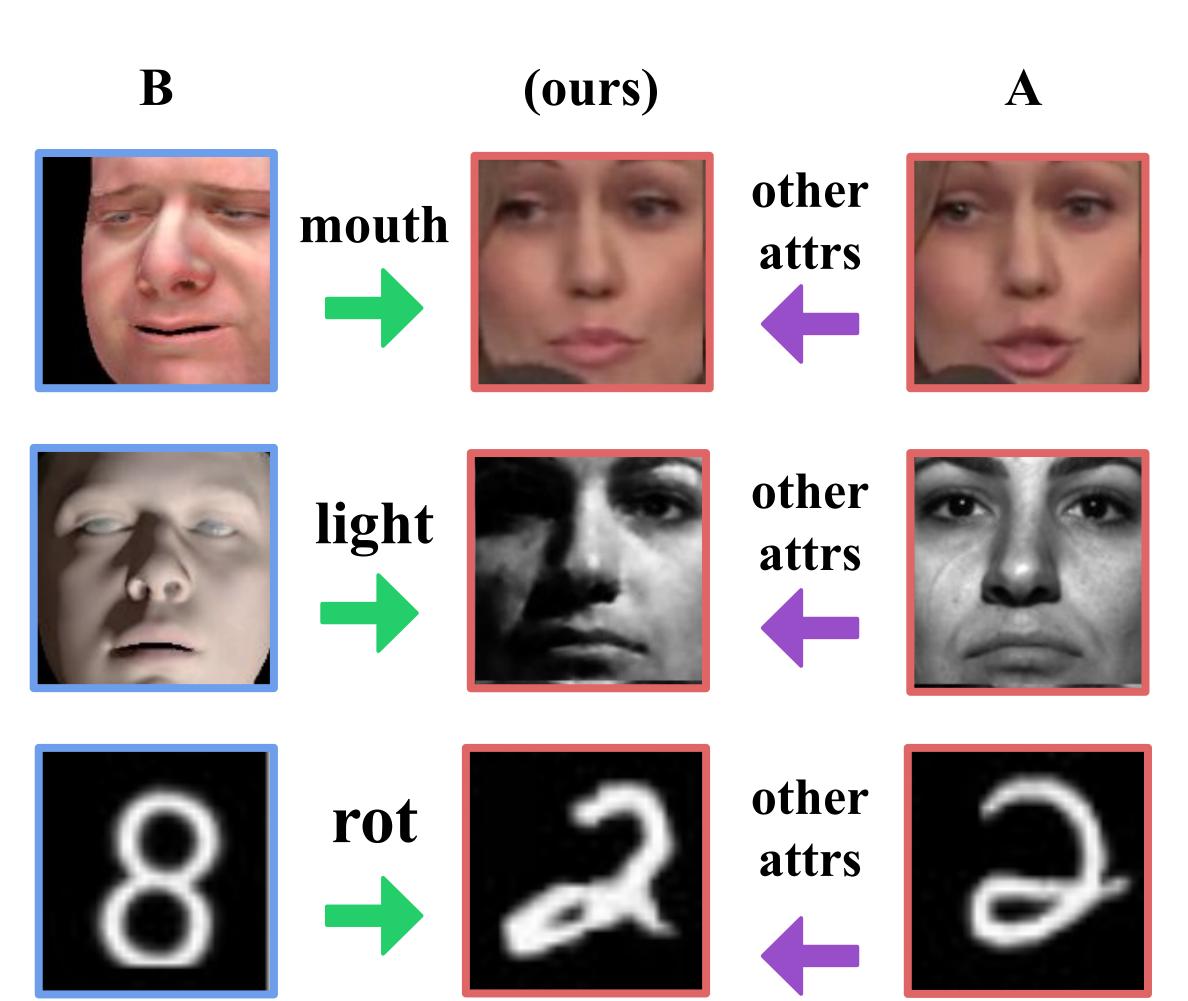
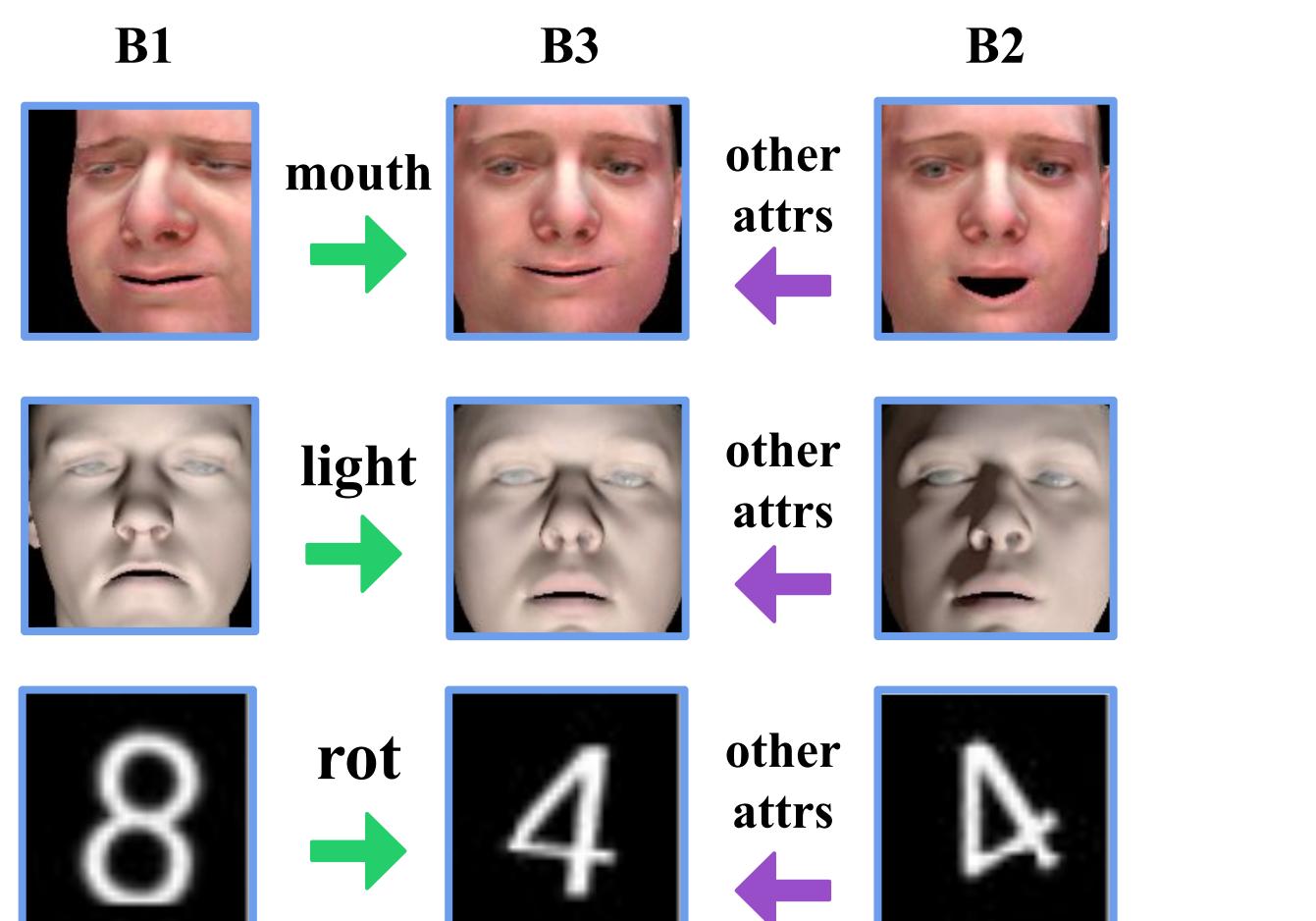
# Google Al

### Task

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

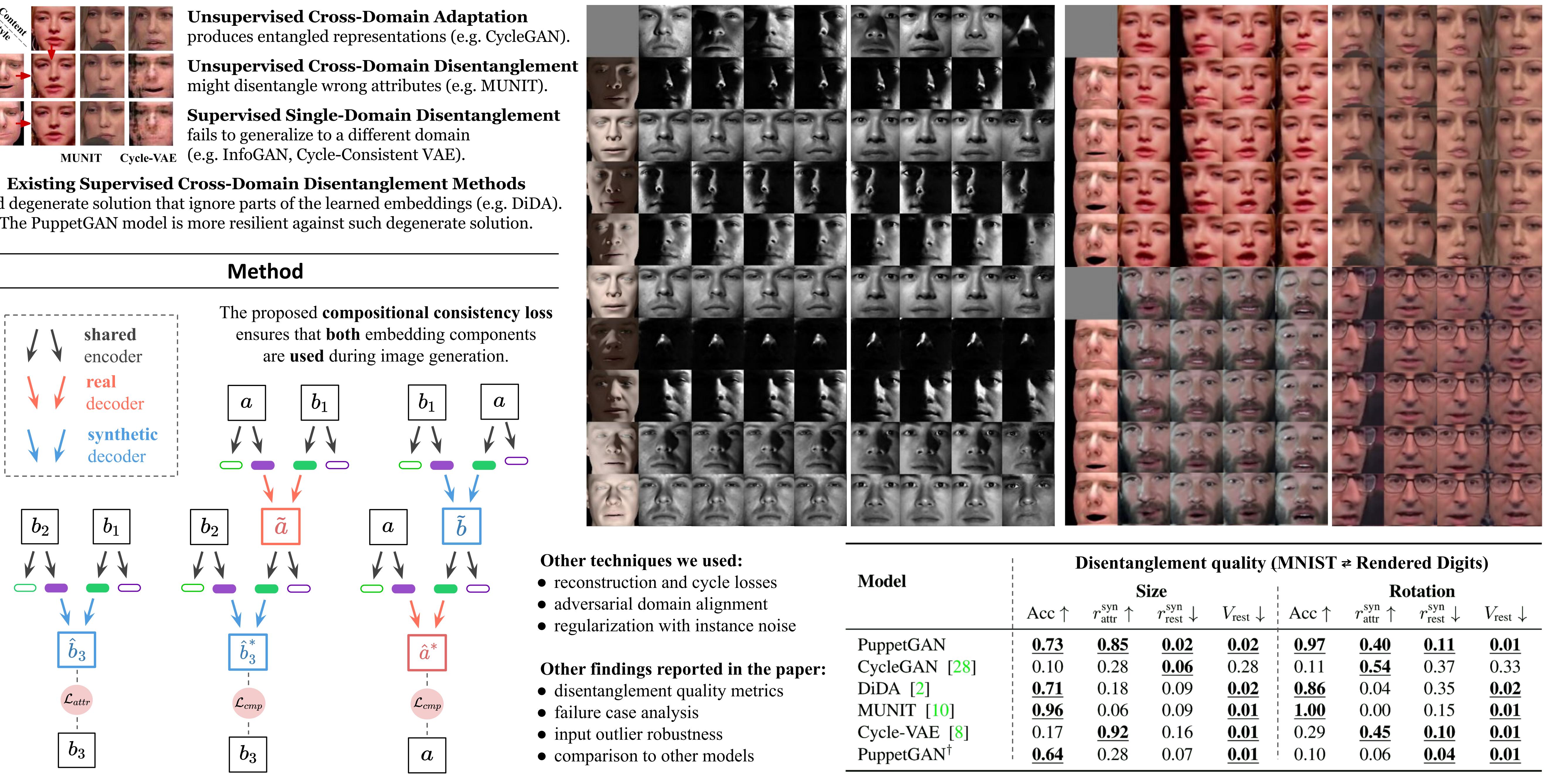


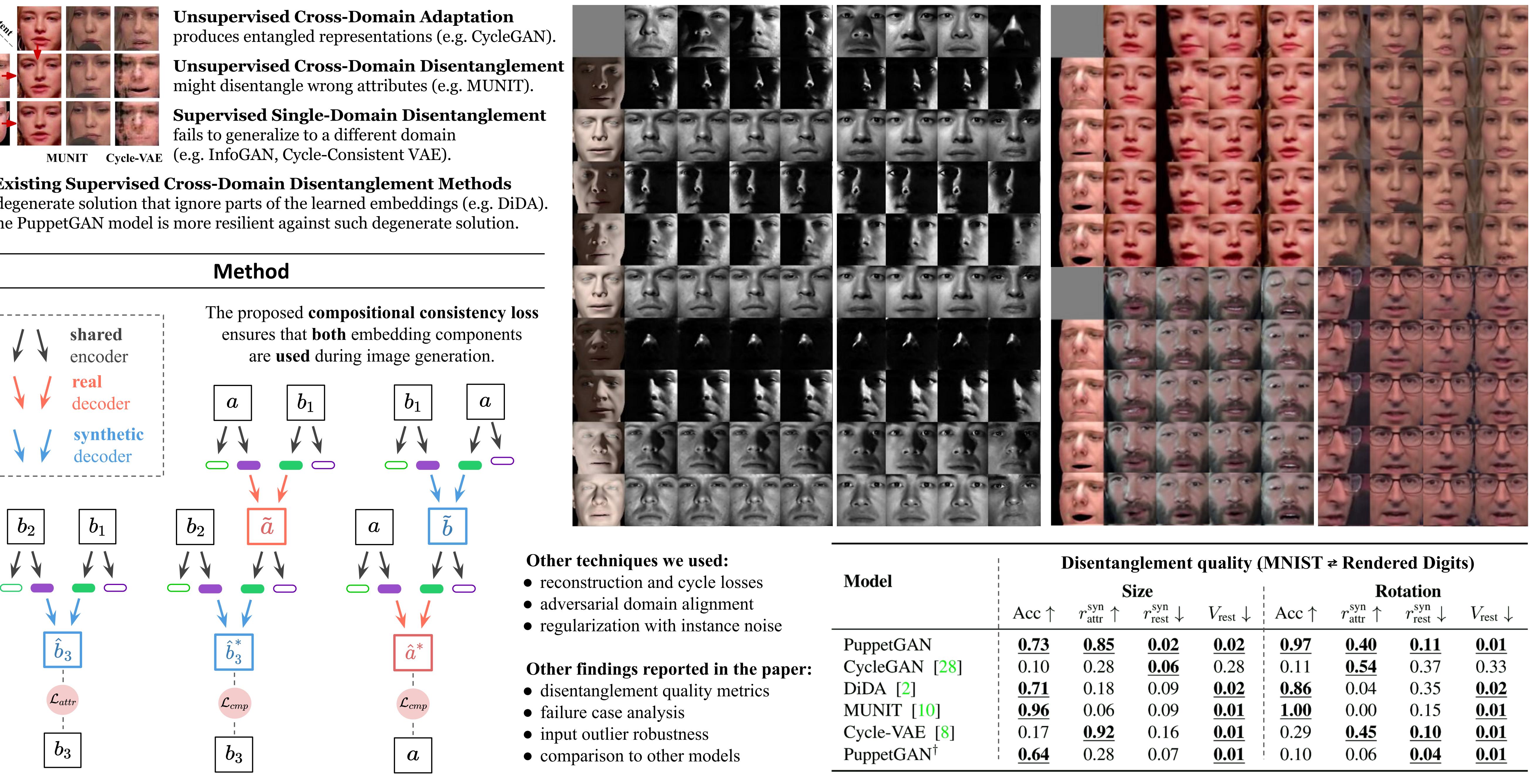
It is trained exclusively on synthetic demonstrations and unlabeled real images.





**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.





(a) supervised disentanglement

## PuppetGAN: Cross-Domain Image Manipulation by Demonstration

<sup>1</sup>Google AI

## **Related Work**

(b) compositional consistency

Ben Usman<sup>1,2</sup>, Nick Dufour<sup>1</sup>, Kate Saenko<sup>2</sup>, Chris Bregler<sup>1</sup> <sup>2</sup> Boston University

Results

er	techniques	we used:	
----	------------	----------	--

ai.bu.edu/puppetgan

Nodel	Size			Rotation				
	Acc $\uparrow$	$r_{ m attr}^{ m syn} \uparrow$	$r_{\mathrm{rest}}^{\mathrm{syn}}\downarrow$	$V_{\text{rest}}\downarrow$	Acc $\uparrow$	$r_{ m attr}^{ m syn} \uparrow$	$r_{\mathrm{rest}}^{\mathrm{syn}}\downarrow$	$V_{\text{rest}}\downarrow$
PuppetGAN	<u>0.73</u>	<u>0.85</u>	<u>0.02</u>	<u>0.02</u>	<u>0.97</u>	<u>0.40</u>	<u>0.11</u>	<u>0.01</u>
CycleGAN [28]	0.10	0.28	<u>0.06</u>	0.28	0.11	<u>0.54</u>	0.37	0.33
DiDA [2]	<u>0.71</u>	0.18	0.09	<u>0.02</u>	<u>0.86</u>	0.04	0.35	<u>0.02</u>
MUNIT [10]	<u>0.96</u>	0.06	0.09	<u>0.01</u>	<u>1.00</u>	0.00	0.15	<u>0.01</u>
Cycle-VAE [8]	0.17	<u>0.92</u>	0.16	<u>0.01</u>	0.29	<u>0.45</u>	<u>0.10</u>	<u>0.01</u>
PuppetGAN <sup>†</sup>	<u>0.64</u>	0.28	0.07	<u>0.01</u>	0.10	0.06	<u>0.04</u>	<u>0.01</u>

*†* larger discrepancy in attribute distributions between A and  $B \Rightarrow$  lower disentanglement quality

BOSTON JNIVERSITY