

# XAGen: 3D Expressive Human Avatars Generation

*Zhongcong Xu, Jianfeng Zhang, Jun Hao Liew, Jiashi Feng, and Mike Zheng Shou*



Project page: <https://showlab.github.io/xagen>

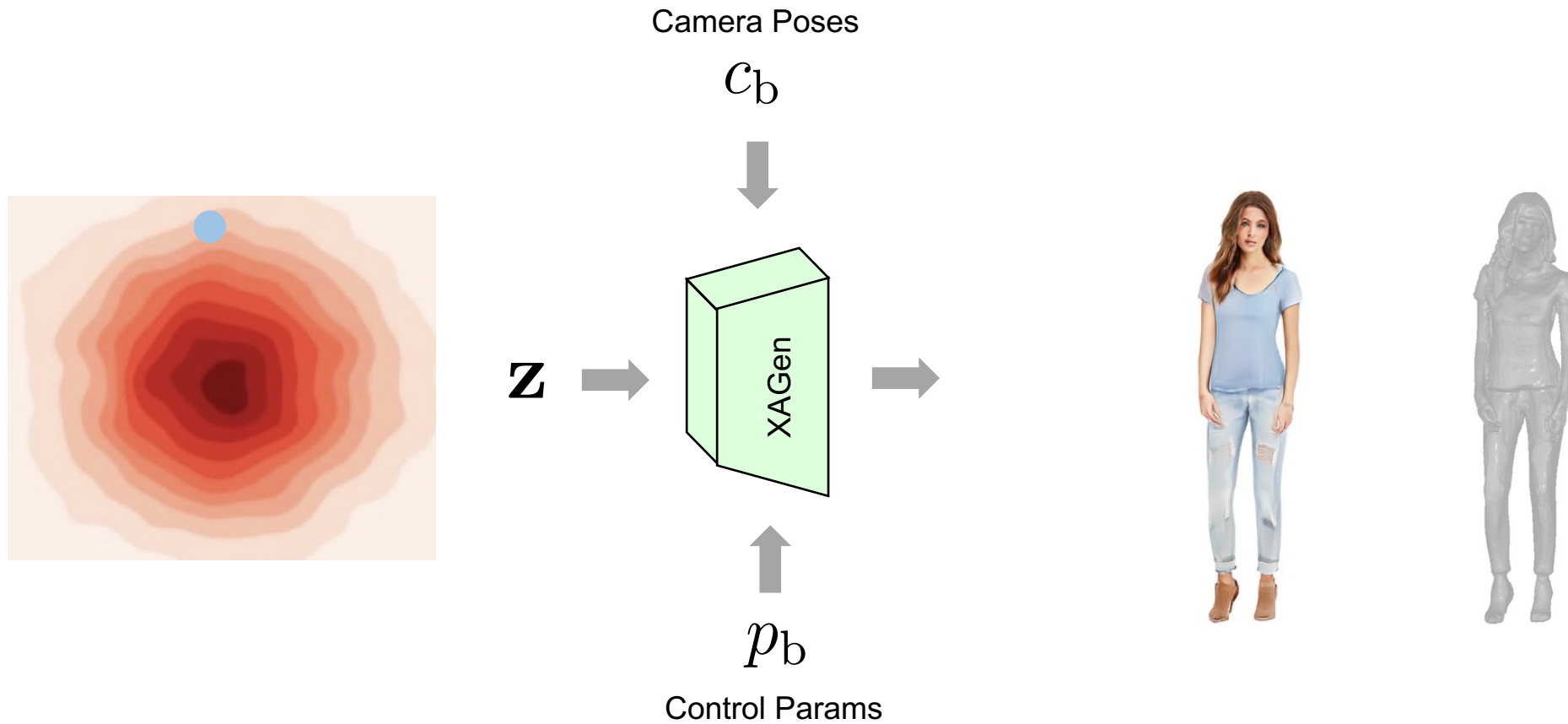
# Background



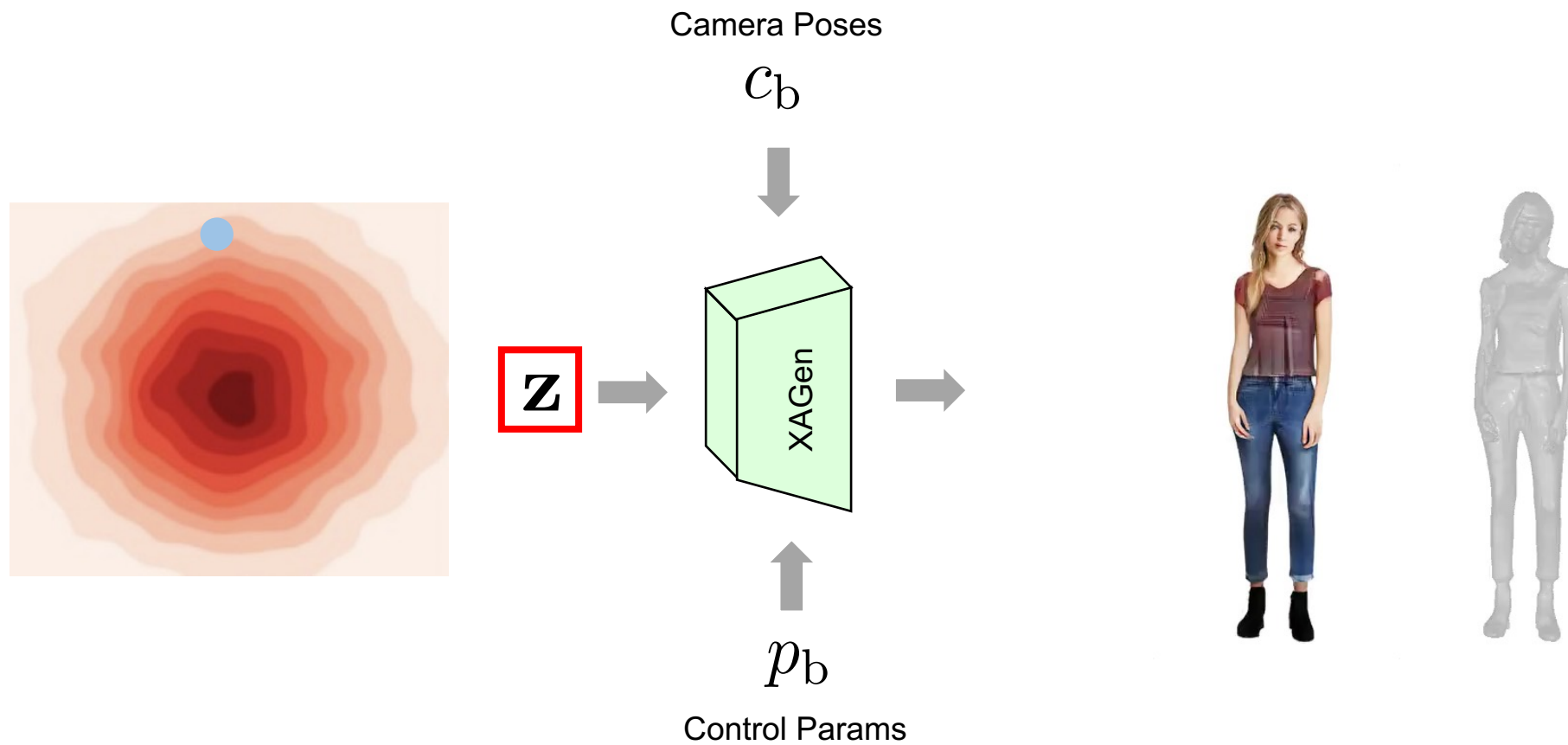
More domain specific priors. More control ability.

From general object to human face/body.

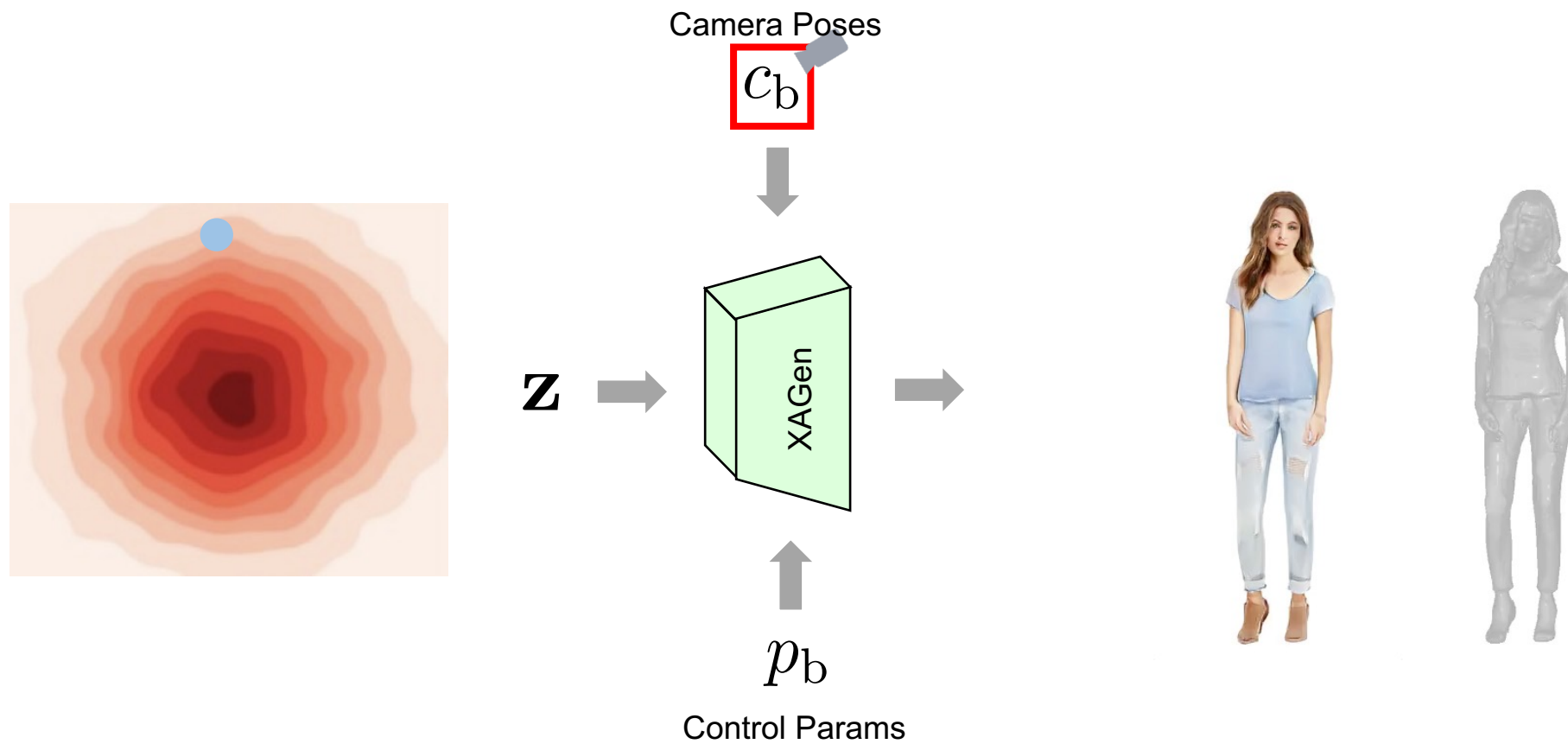
# Target: 3D avatar generation with expressive controllability



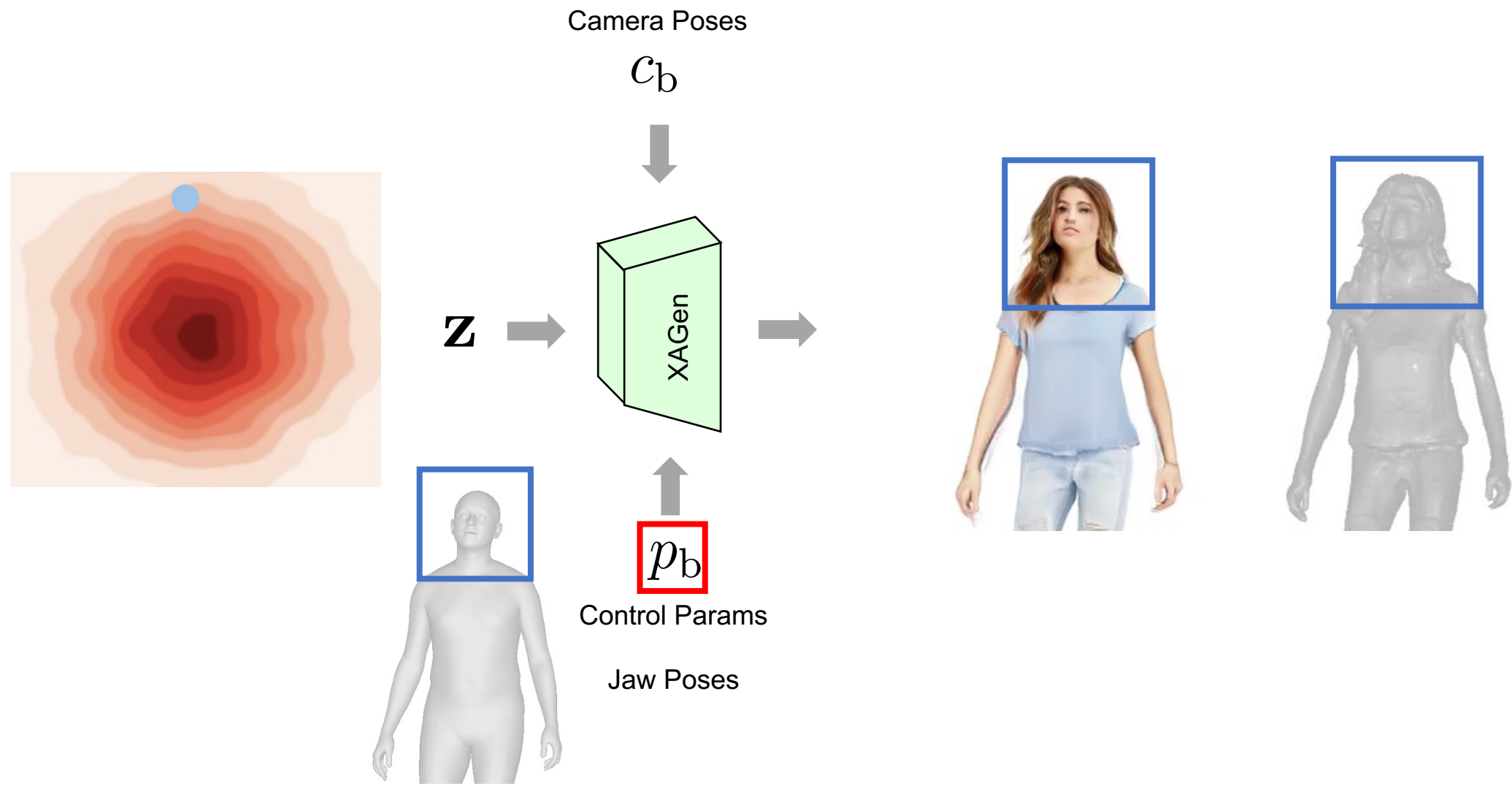
# Various latent code



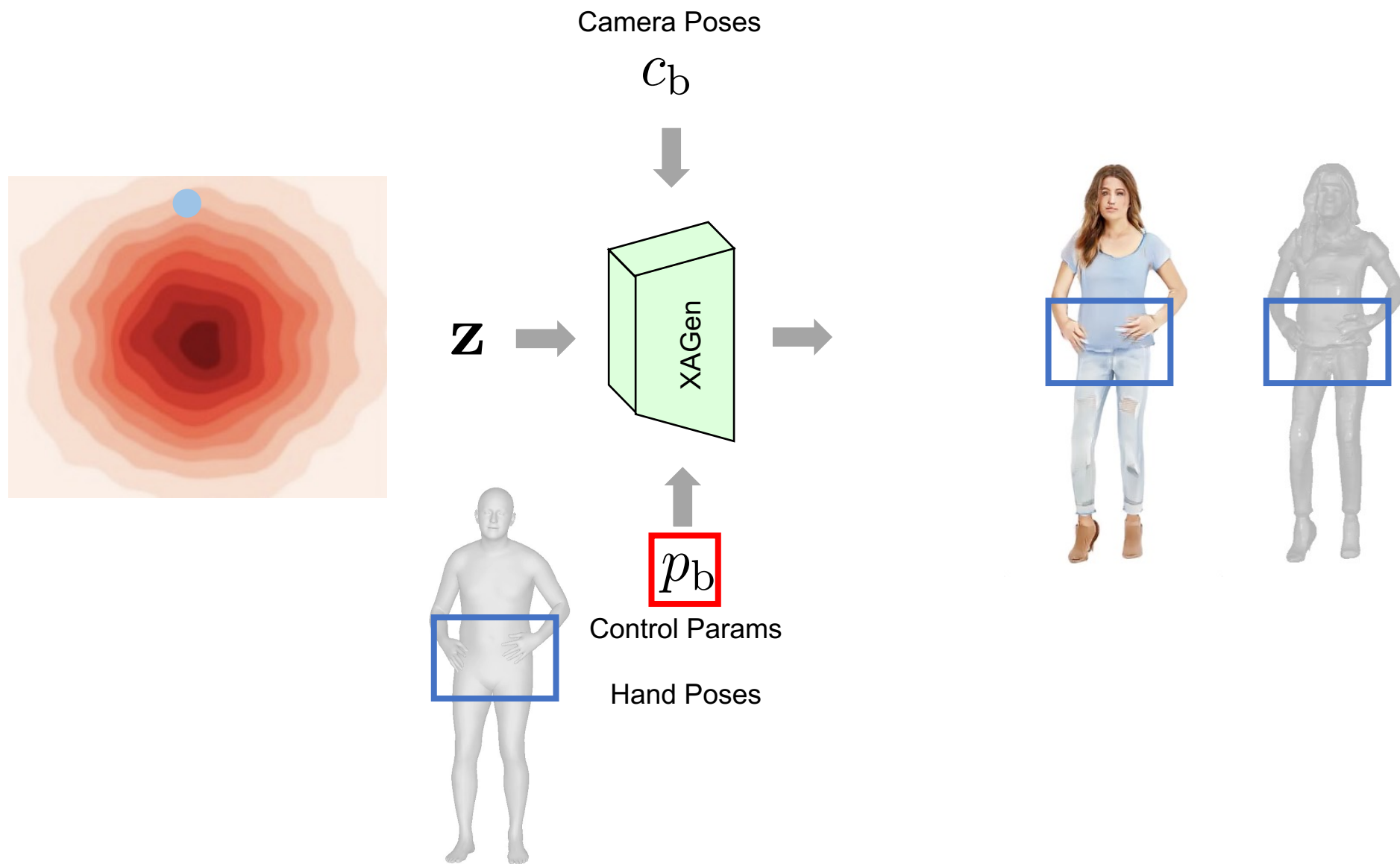
# Various camera poses



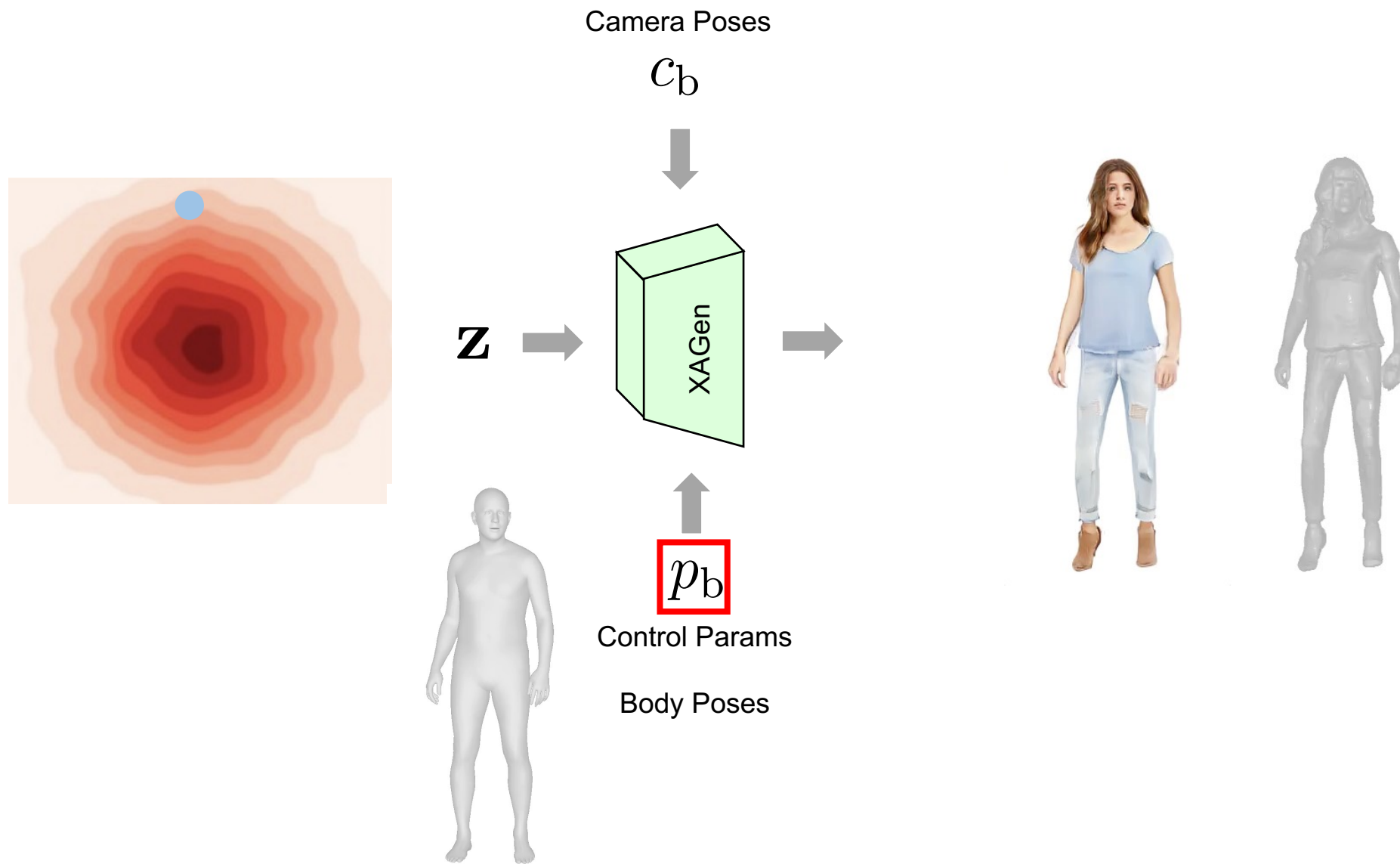
# Various jaw poses



# Various hand poses

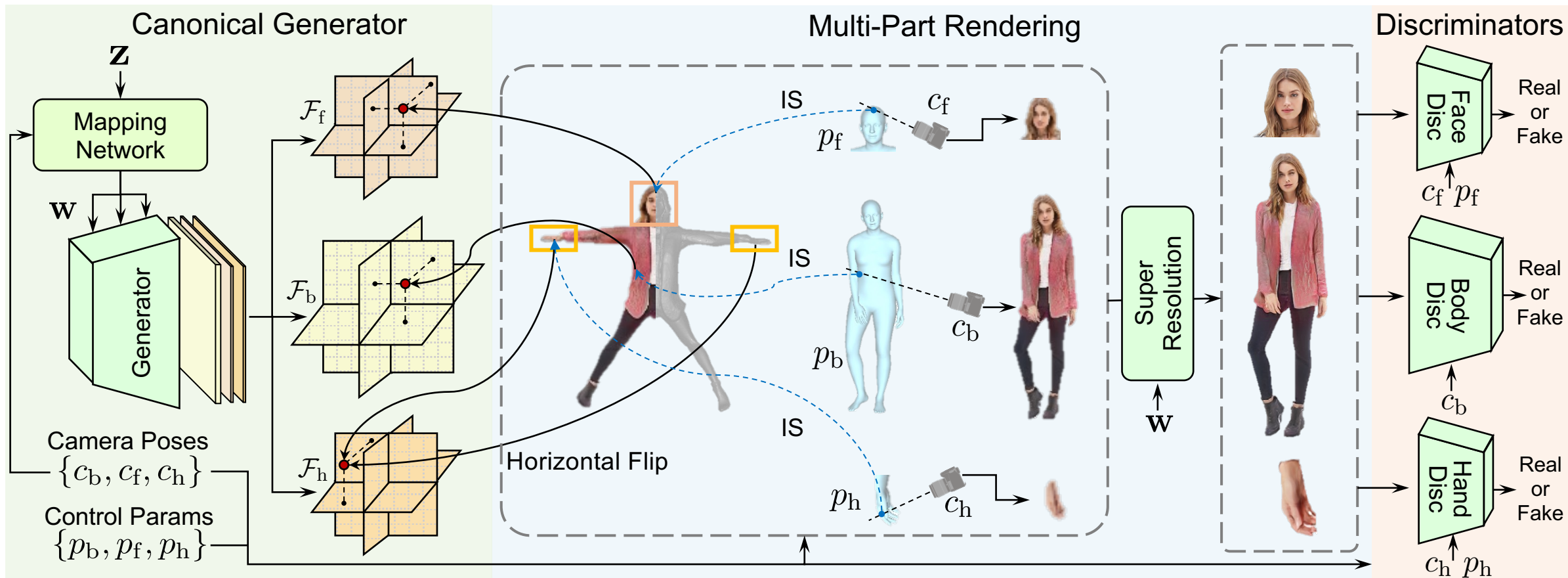


# Various body poses

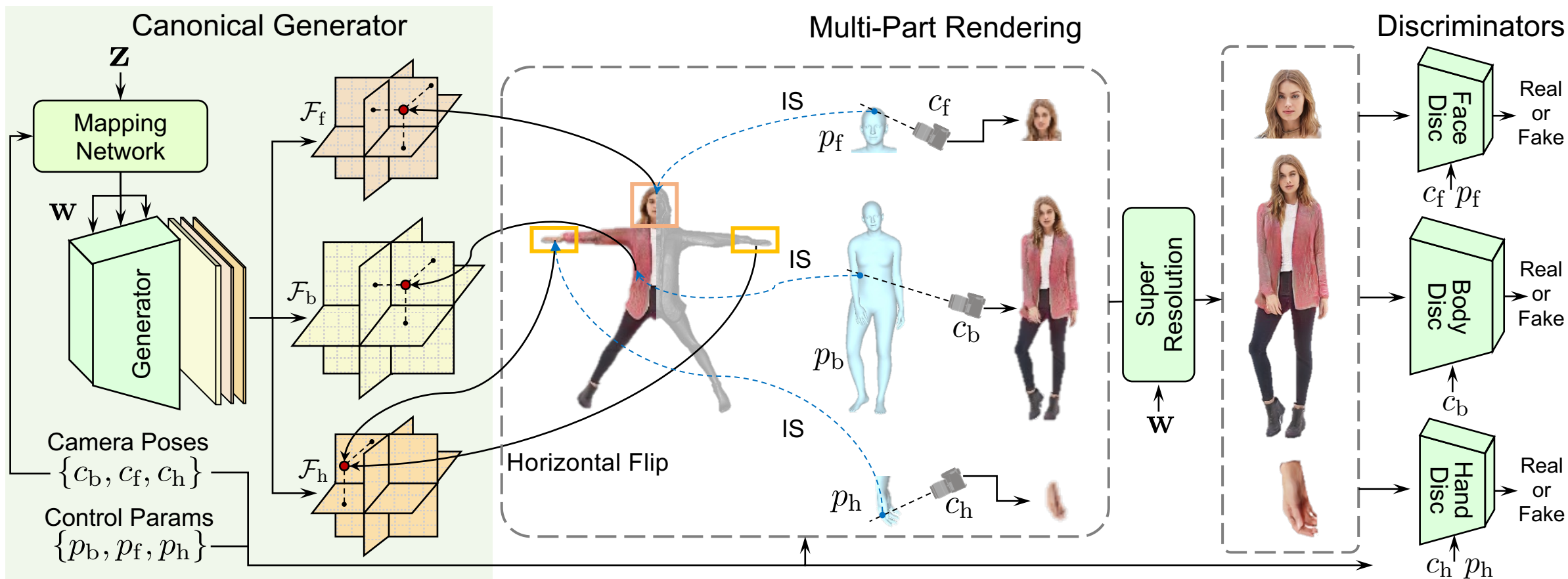




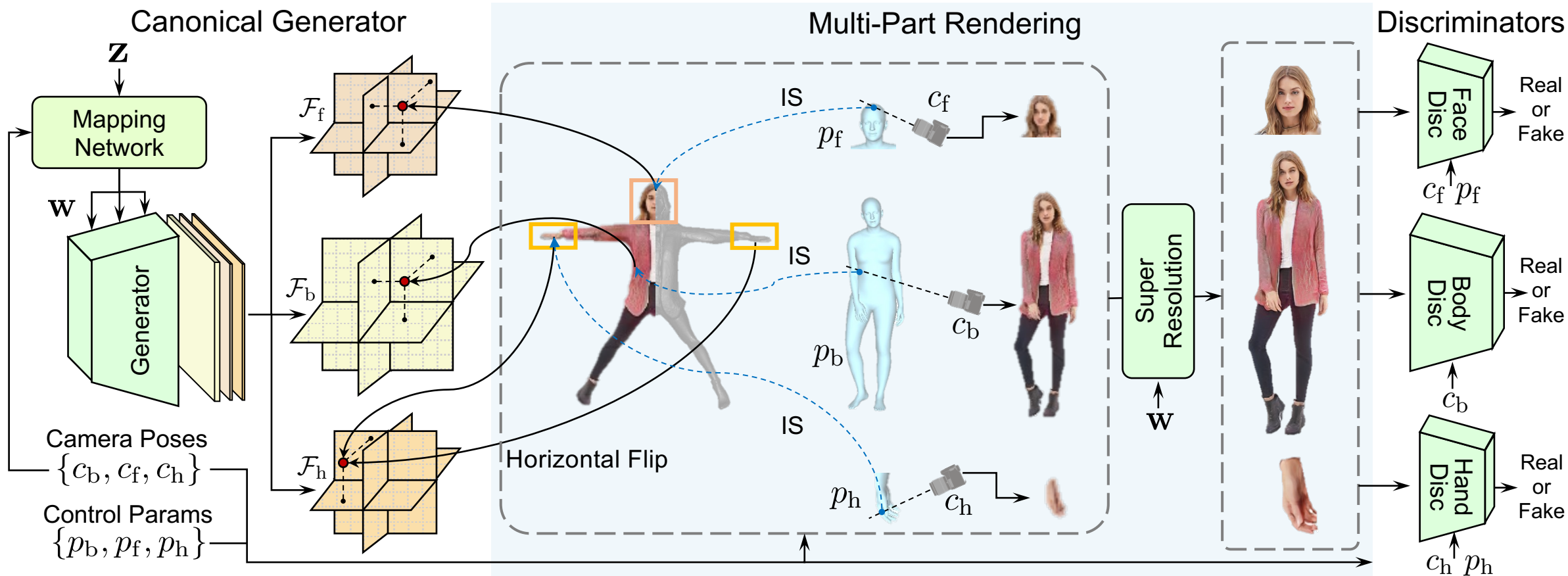
# Pipeline



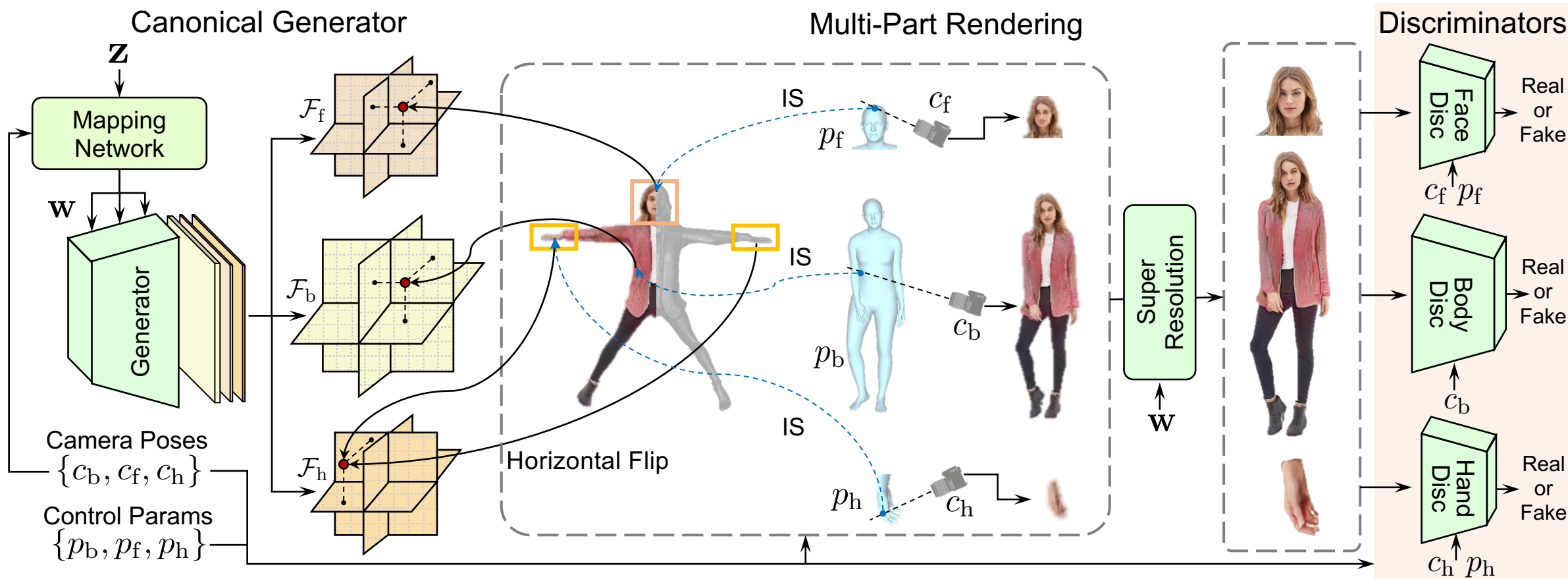
# Pipeline: a) Canonical Generator



# Pipeline: b) Multi-Part Rendering



# Pipeline: c) Discriminators



# Experiments

Table 1: Quantitative comparisons with baselines in terms of appearance and overall control ability, with best results in **bold**. F.Ctl. indicates whether the approach generates fully controllable human body or not. \*We implement AvatarGen by conditioning it on SMPL-X.

	DeepFashion [35]							MPV [14]					
	F.Ctl.	FID↓	FID <sub>f</sub> ↓	FID <sub>h</sub> ↓	PCK↑	PCK <sub>f</sub> ↑	PCK <sub>h</sub> ↑	FID↓	FID <sub>f</sub> ↓	FID <sub>h</sub> ↓	PCK↑	PCK <sub>f</sub> ↑	PCK <sub>h</sub> ↑
ENARF [42]	✗	68.62	52.17	46.86	3.54	3.79	1.34	65.97	47.71	37.08	3.06	3.55	0.67
EVA3D [26]	✗	15.91	14.63	48.10	56.36	75.43	23.14	14.98	27.48	32.54	33.00	42.47	19.24
AG3D [16]	✗	10.93	14.79	-	-	-	-	-	-	-	-	-	-
AvatarGen [65]*	✓	9.53	13.96	27.68	60.12	73.38	46.50	10.06	13.08	19.75	38.32	45.26	30.75
XAGen (Ours)	✓	<b>8.55</b>	<b>10.69</b>	<b>24.26</b>	<b>66.04</b>	<b>87.06</b>	<b>47.56</b>	<b>7.94</b>	<b>12.07</b>	<b>17.35</b>	<b>48.84</b>	<b>63.77</b>	<b>32.01</b>
<hr/>													
	UBC [64]							SHHQ [18]					
	F.Ctl.	FID↓	FID <sub>f</sub> ↓	FID <sub>h</sub> ↓	PCK↑	PCK <sub>f</sub> ↑	PCK <sub>h</sub> ↑	FID↓	FID <sub>f</sub> ↓	FID <sub>h</sub> ↓	PCK↑	PCK <sub>f</sub> ↑	PCK <sub>h</sub> ↑
ENARF [42]	✗	36.39	34.27	32.72	6.90	7.44	6.37	79.29	50.19	46.97	4.43	4.62	2.71
EVA3D [26]	✗	12.61	36.87	45.66	36.31	55.31	8.38	11.99	20.04	39.83	31.24	37.60	18.38
AG3D [16]	✗	11.04	15.83	-	-	-	-	-	-	-	-	-	-
AvatarGen [65]*	✓	9.75	13.23	18.09	65.31	77.09	55.09	10.52	12.57	28.21	59.18	78.71	36.29
XAGen (Ours)	✓	<b>8.80</b>	<b>9.82</b>	<b>16.72</b>	<b>69.18</b>	<b>84.18</b>	<b>55.17</b>	<b>5.88</b>	<b>10.06</b>	<b>19.23</b>	<b>65.14</b>	<b>91.44</b>	<b>38.53</b>

# Experiments

UBC



DeepFashion



ENARF

EVA3D

AvatarGen

Ours

# Experiments

SHHQ



MPV



ENARF

EVA3D

AvatarGen

Ours

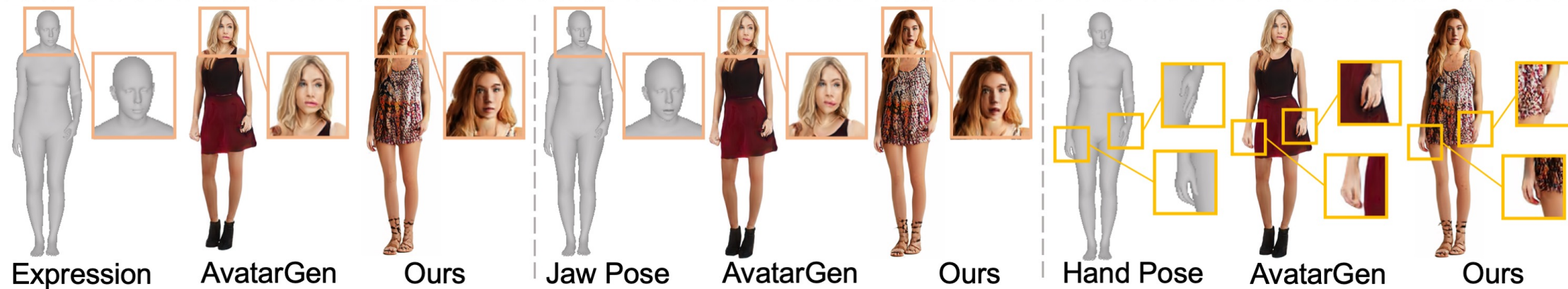
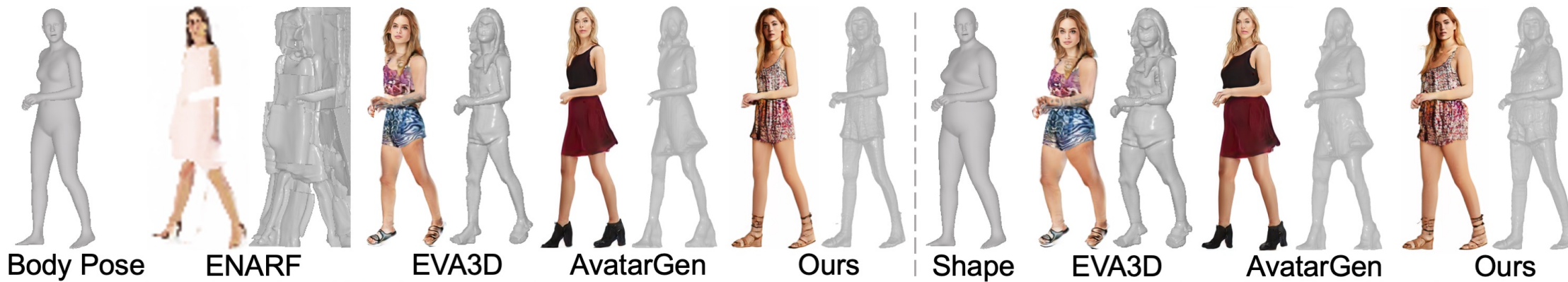
# Experiments

Table 2: Quantitative comparisons with baselines in terms of disentangled control ability measured by MSE. We report  $\text{Jaw} \times 10^{-4}$  and others  $\times 10^{-2}$  for simplicity, with best results in **bold**. \*We implement AvatarGen by conditioning it on SMPL-X.

	DeepFashion [36]					MPV [14]				
	Exp↓	Shape↓	Jaw↓	Body↓	Hand↓	Exp↓	Shape↓	Jaw↓	Body↓	Hand↓
ENARF [43]	13.47	6.30	5.79	3.14	9.87	11.21	4.91	8.36	2.75	12.90
EVA3D [26]	6.03	2.87	5.11	1.78	3.68	9.97	4.14	13.83	1.80	4.65
AvatarGen [69]*	4.92	3.06	5.05	<b>1.23</b>	3.17	8.98	<b>3.88</b>	15.22	1.11	3.47
XAGen (Ours)	<b>4.46</b>	<b>2.77</b>	<b>3.67</b>	1.26	<b>2.95</b>	<b>6.31</b>	<b>3.88</b>	<b>7.43</b>	<b>0.94</b>	<b>2.23</b>
	UBC [68]					SHHQ [18]				
	Exp↓	Shape↓	Jaw↓	Body↓	Hand↓	Exp↓	Shape↓	Jaw↓	Body↓	Hand↓
ENARF [43]	10.70	6.11	<b>3.62</b>	1.07	8.19	14.51	6.43	8.16	3.27	9.83
EVA3D [26]	7.00	2.98	5.36	1.00	2.78	7.43	4.15	9.26	1.93	5.15
AvatarGen [69]*	9.59	4.50	9.34	1.22	3.01	9.01	3.99	8.87	1.52	4.99
XAGen (Ours)	<b>5.35</b>	<b>2.57</b>	4.76	<b>0.73</b>	<b>1.63</b>	<b>5.56</b>	<b>3.66</b>	<b>6.57</b>	<b>1.24</b>	<b>3.30</b>



# Experiments



# Experiments



# Experiments



# Experiments



# Experiments

*brown hair woman, red T-shirt, blue jeans*



*blonde hair woman, pink T-shirt, black trousers*

Thank you!

