A Appendices

In this supplementary, we first give the detailed network structures and then provide more supported results.

A.1 Details about network structures

The structure of G on 3D chair. The decoder structure of G on the 3D chair is shown in Figure 1 (b). Different from the structure for MultiPIE shown in Figure 1 (a), the Z sampled from the posterior $E(Z|X_a, Y_a)$ or prior $N(0, I)$ is given to the network from the side branch by AdaIN. The features in the main branch are affected by two outputs from the side branches through the AdaIN and DFNM, respectively. Their results are then concatenated along the channel dimension, and given to the next layer.

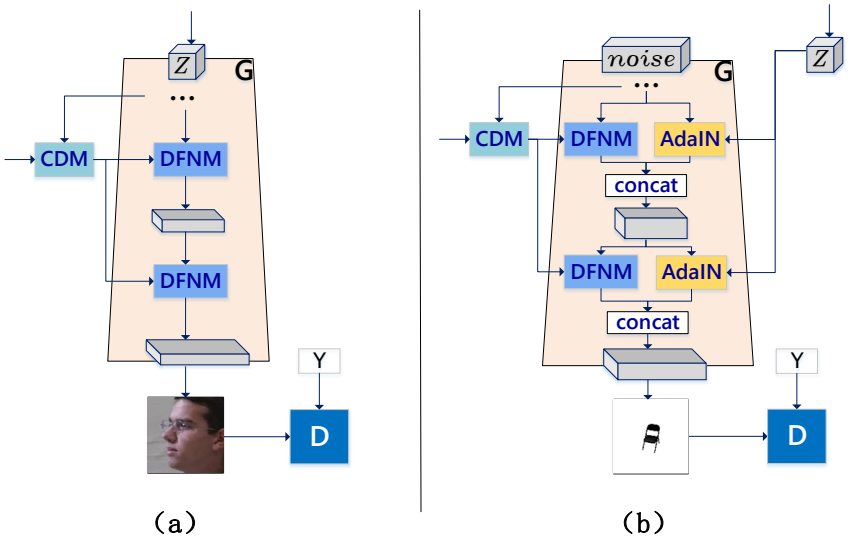


Figure 1: The two different proposed structures for G. (a) The structure on MultiPIE. (b) The structure on 3D chair.

Detailed structures of different modules. Table 1, 2, 3, 4, 5, 6 are the specific structures of the network of E, G, D, CDM, Ψ, and DAC, respectively, $\rightarrow$ means directly given. Note that the decoder G in Table 2 is used on MuliPIE.

Encoder E
input $X \in \mathbb{R}^{128\times128\times3}$
conv (F:32, K:7, S:1), IN, lrelu
conv (F:64, K:4, S:2), IN, lrelu
conv (F:128, K:4, S:2), IN, lrelu
conv (F:128, K:4, S:2), IN, lrelu
input $Y \rightarrow \Psi_E \rightarrow$ CDM
Residual block: conv (F:128, K:3, S:1), DFNM, lrelu
Residual block: conv (F:128, K:3, S:1), DFNM, lrelu
Residual block: conv (F:128, K:4, S:1), DFNM, lrelu
conv (F:128, K:4, S:2), IN, lrelu
conv (F:128, K:3, S:2), IN, lrelu
fc, 1024, lrelu
fc, 256 (μ), fc, 256 (Σ)

Table 1: The structure of the encoder E. Note that the proposed CDM is used at the beginning of the first residual block to inject the input condition into the main branch. Details about CDM and Ψ are given in the following tables.

Decoder G
input $Z \in \mathbb{R}^{256}$
conv (F:128, K:3, S:1), lrelu
conv (F:128, K:4, S:2), lrelu
conv (F:128, K:4, S:2), lrelu
input $Y' \rightarrow \Psi_G \rightarrow$ CDM
Residual block: conv (F:128, K:3, S:1), DFNM, lrelu
Residual block: conv (F:128, K:3, S:1), DFNM, lrelu
Residual block: conv (F:128, K:3, S:1), DFNM, lrelu
conv (F:128, K:4, S:1), LN, lrelu
conv (F:128, K:4, S:1), LN, lrelu
conv (F:64, K:4, S:1), LN, lrelu
conv (F:32, K:7, S:1), LN, lrelu
conv (F:3, K:1, S:1), tanh

Table 2: The structure of decoder G for MultiPIE dataset.

Discriminator	
input $X \in \mathbb{R}^{128\times128\times3}$	
conv(F:64, K:1, S:1), lrelu	
Residual block: conv (F:128, K:3, S:1), SN, lrelu	
downsample	
Residual block: conv (F:128, K:3, S:1), SN, lrelu	
downsample	
Residual block: conv (F:256, K:3, S:1), SN, lrelu	
downsample	
Minibatch state concat [1]	
conv (F:256, K:3, S:1), SN, lrelu	
conv (F:256, K:4, S:1), SN, lrelu	
input Y →fc, SN → Inner product	fc (1), SN
add	

Table 3: The structure of projection discriminator D. Note that "SN" indicates the spectral normalization.

CDM
input $F \in \mathbb{R}^{16\times16\times128}$
conv (F:9×25, K:3, S:1)
input $W_y \in \mathbb{R}^{1\times1\times25}$ →KGconv (F:9, K:1, S:1)→ dy input $W_x \in \mathbb{R}^{1\times1\times25}$ →KGconv (F:9, K:1, S:1)→ dx concat
deformable conv (F:128, K:3, S:1) input F
concat input F

Table 4: The network structure of CDM.

Ψ
input $\mathbf{Y}$
fc(128)
concat **noise** $\in \mathbb{R}^{128}$
pixel norm
fc(256)

Table 5: The network structure of Ψ.

DAC
input $\mathbf{Z} \in \mathbb{R}^{256}$
fc(256)
fc(13 on MultiPIE) or (62 on 3D chair)

Table 6: The network structure of DAC.

A.2 Additional results and analysis

View synthesis from the paired data training. In this section, we provide the synthesis results obtained from the model trained by the paired data, which means that the target view image is directly used to constrain the model.

Figure 2: Synthesis results on MultiPIE training by the paired data. The first row shows the view-translated images, and the second row are the target images.

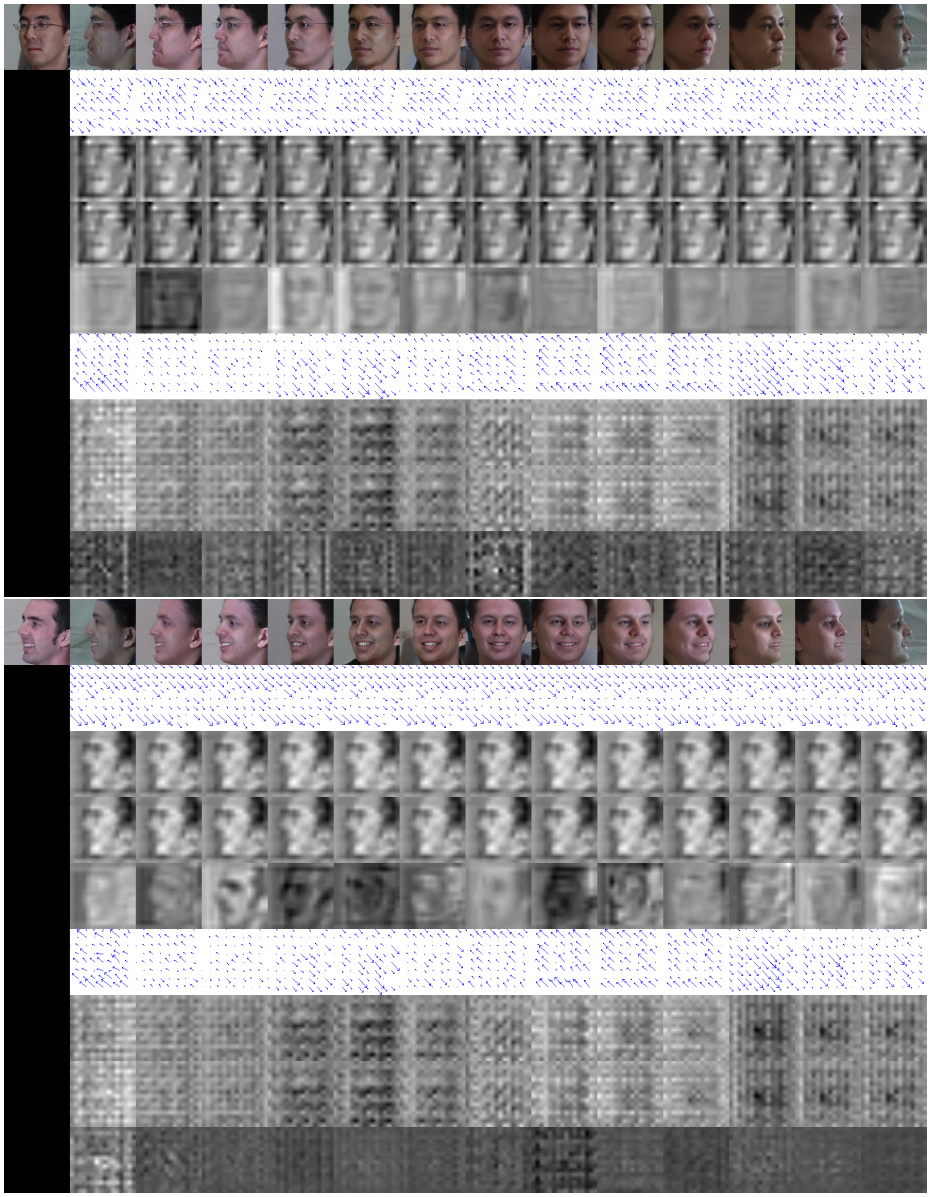
method	MultiPIE		
	L1	SSIM	FID
our-paired	13.34	0.63	23.52

Table 7: Quantitative result on the MultiPIE based on the paired training data.

Visualization and analysis of the conditional flows. As is shown in the Figure 3, we visualize the dx and dy used for deformation. The first row are the input image and the synthesis images under 13 different azimuths. The second row are the optical flows used by the CDM module in E. The third, fourth and fifth rows are dy, dx and their differences, respectively. The sixth, seventh, eighth and ninth are the optical flow, dy, dx and their differences in the CDM of G.

Since the flows are actually the intermediate features, they are hard to interpret. But it can be seen that when the input Y of the encoder E and decoder G are the same, that is, the source and target view labels are the same, the two flows are opposite. Note that the flows in E are similar since they are determined

mainly by the source input X and its label Y. The slightly differences are caused by the sampling noises introduced in Ψ.



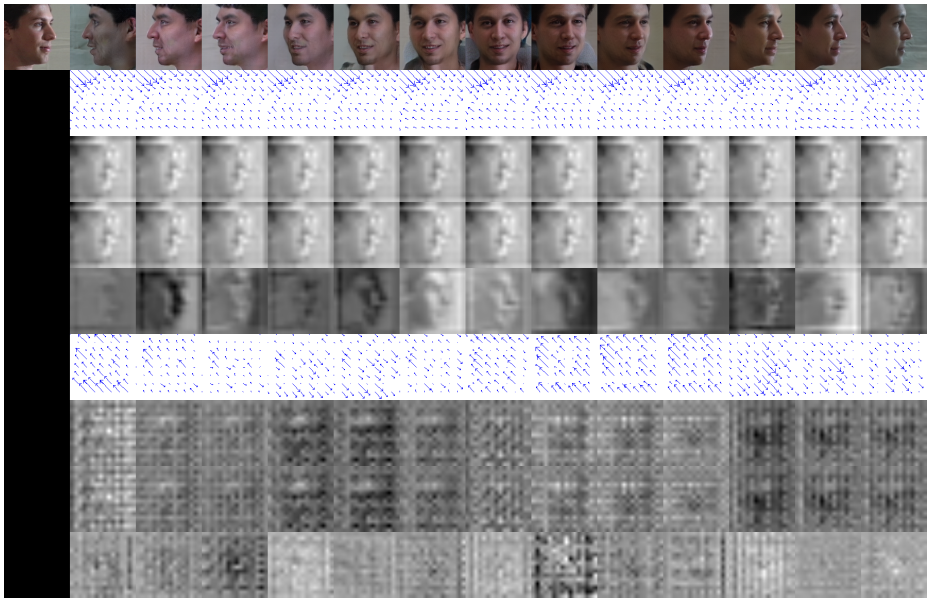


Figure 3: Visualization of the optical flow and its the two components dx and dy under different source and target view combinations. The first row shows the input source image and its different target view translation results. The second and sixth rows are the optical flows in the CDM of E and G. The third and fourth rows are dy and dx components in E, while fifth row explicitly shows their differences. dy, dx and their differences in G are shown in the seventh, eighth and ninth rows.

Additional experiment results on two datasets are provided in Figure 4, 5 and 6.

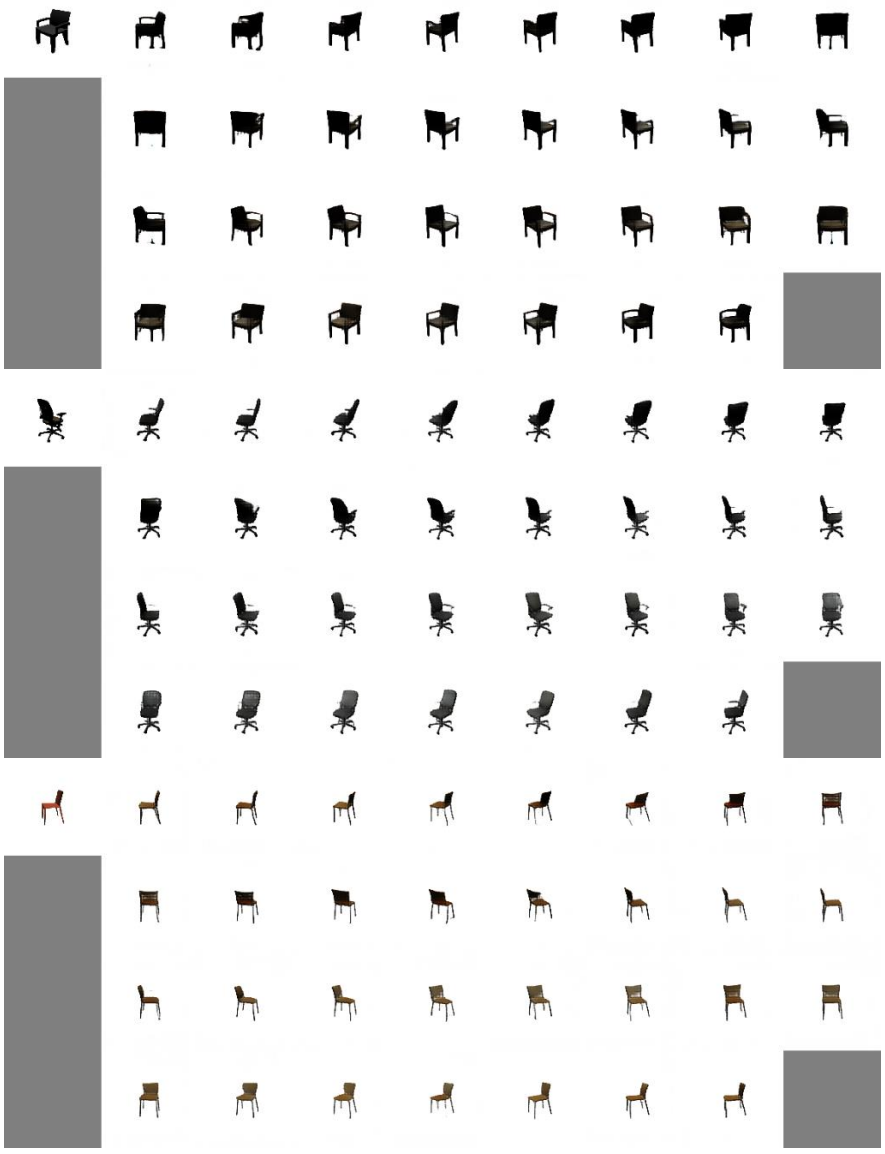


Figure 4: Synthesized images of different views on 3D chair dataset.The first one is input image, and the remaining are generated images under 31 different views.

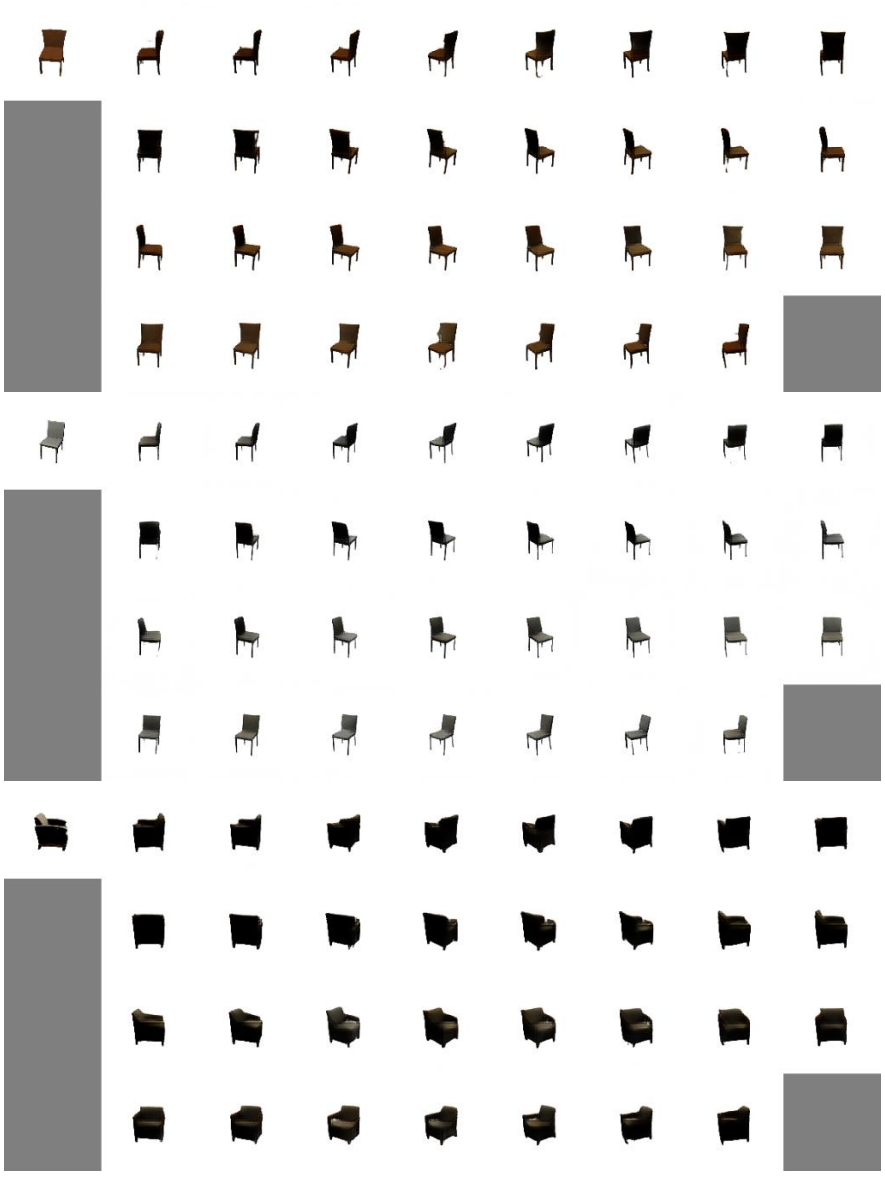


Figure 5: Synthesized images of different views on 3D chair dataset.The first one is input image, and the remaining are generated images under 31 different views.

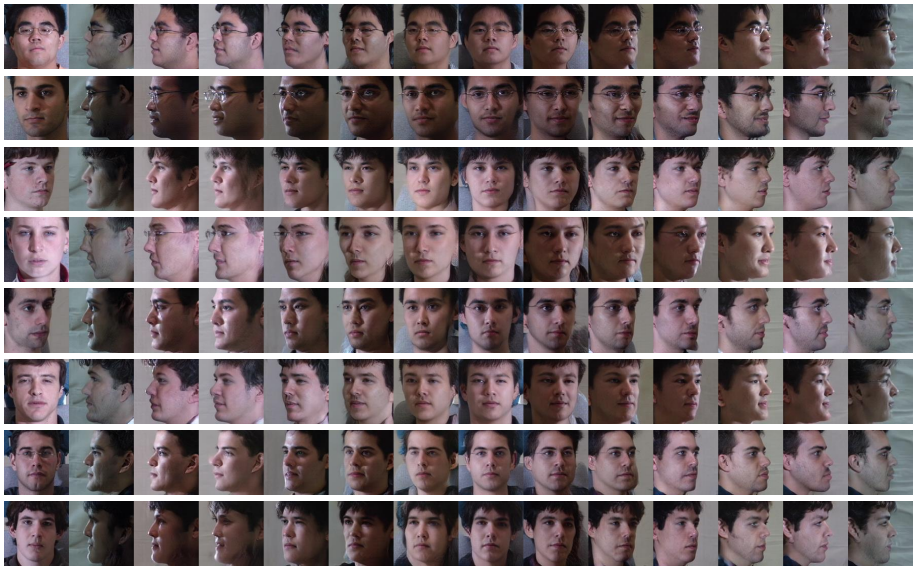


Figure 6: Synthesized images of different views on MultiPIE dataset. The first column is input image, and the remaining 13 columns are view-translated images under 13 different target views.

References

1. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)