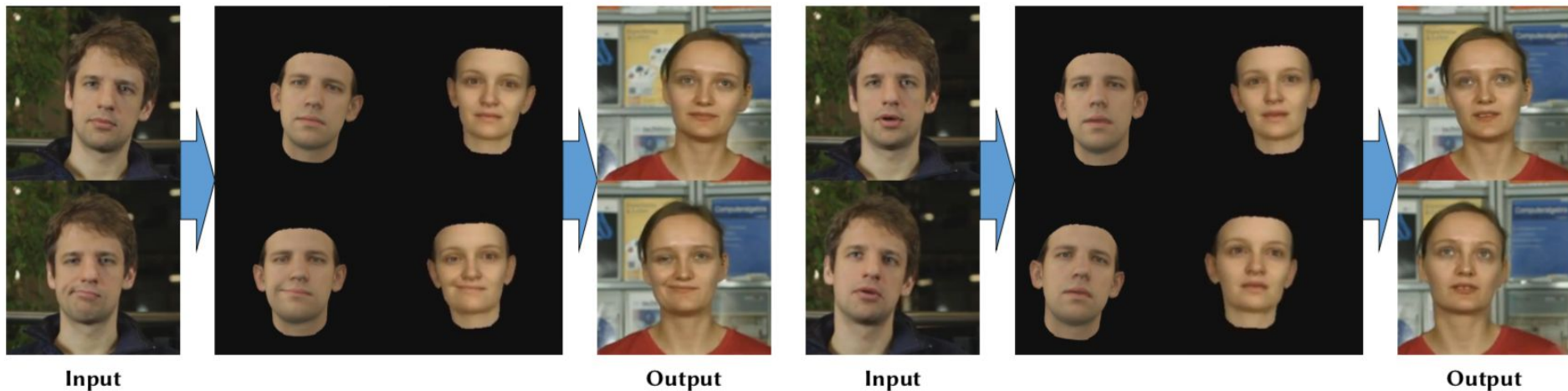
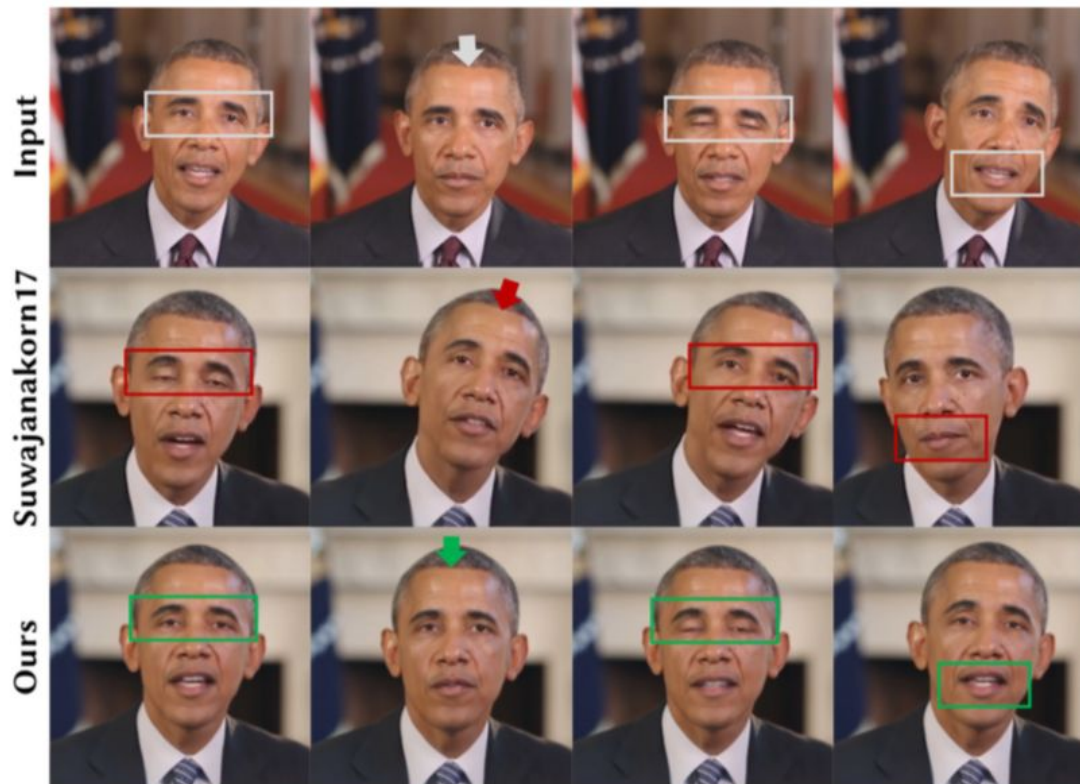


# Deep Video Portraits



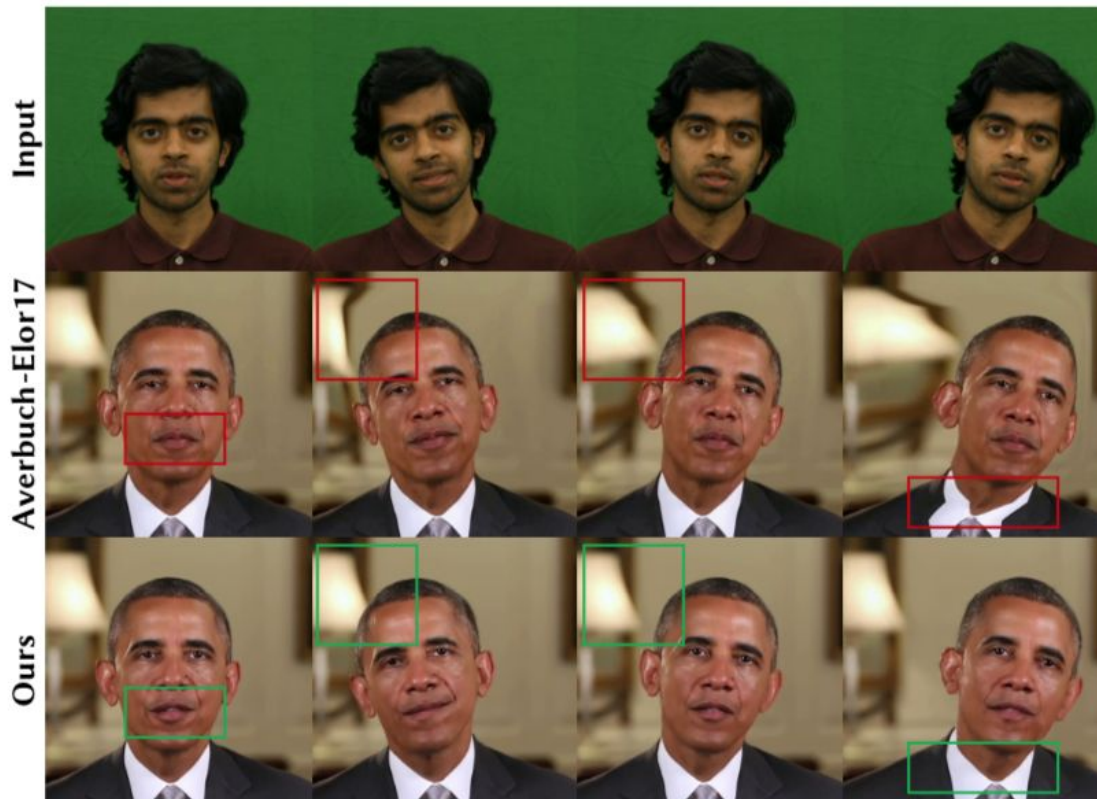
# Похожие задачи

- Видео дубляж



# Похожие задачи

- Video-based  
Facial Reenactment



# Идея метода

Заранее вычислять представление человека

- Идентичность (геометрия лица, отражение кожи)
- Поза головы (rotation and translation)
- Выражение лица
- Направление взгляда
- Освещение (3 точки освещения)

Заменяемые:

- Поза головы
- Выражение лица
- Направление взгляда

# Получение представления

$$\mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \mathbf{a}_{\text{geo}} + \sum_{k=1}^{N_{\alpha}} \alpha_k \mathbf{b}_k^{\text{geo}} + \sum_{k=1}^{N_{\delta}} \delta_k \mathbf{b}_k^{\text{exp}} \quad \mathbf{v} \in \mathbb{R}^{3N}$$

$$\mathbf{r}(\boldsymbol{\beta}) = \mathbf{a}_{\text{ref}} + \sum_{k=1}^{N_{\beta}} \beta_k \mathbf{b}_k^{\text{ref}} \quad \mathbf{r} \in \mathbb{R}^{3N}$$

# Получение изображений

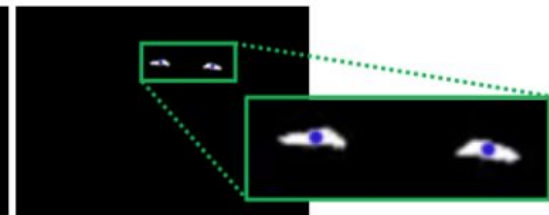
- Input:  $W \times H \times 9N_w$
- Output:  $W \times H \times 3$
- Сверточная сеть в виде Encoder-Decoder в качестве генератора в cGAN



Diffuse Rendering



Correspondence



Eye and Gaze Map

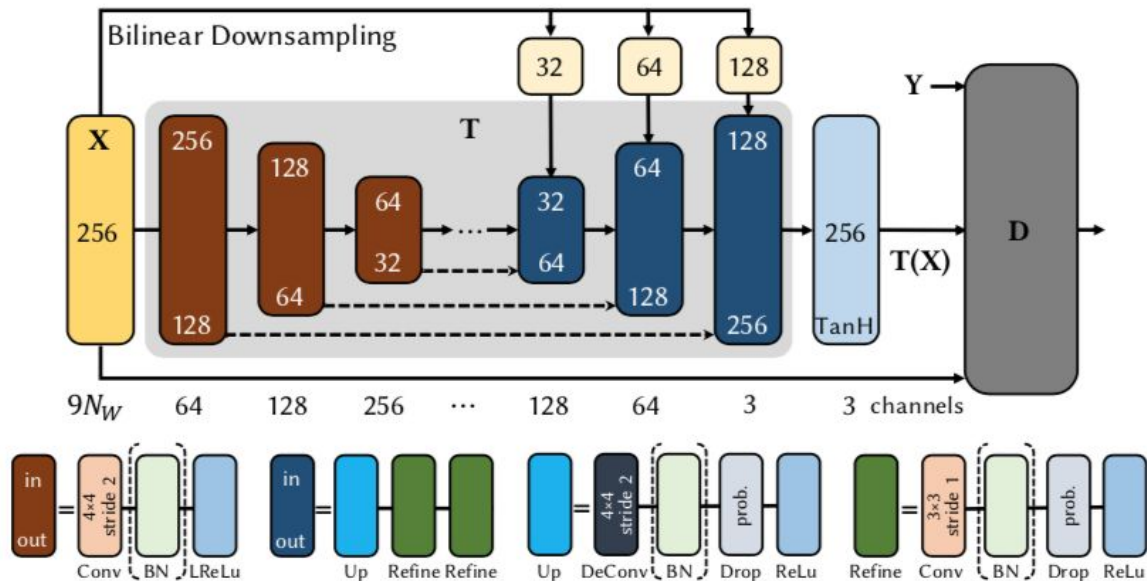
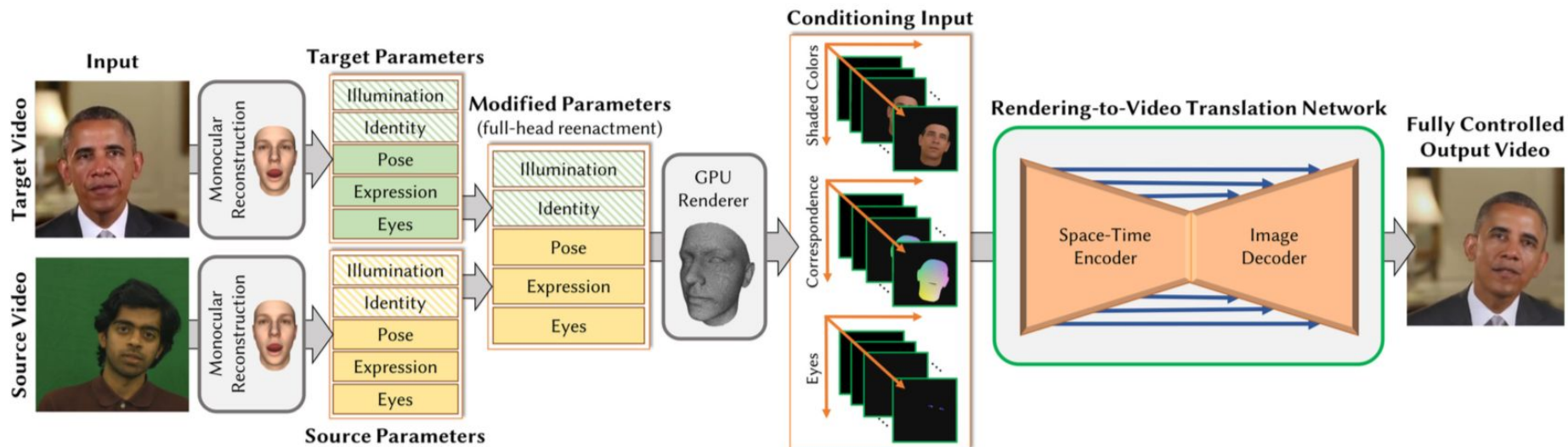


Fig. 4. Architecture of our rendering-to-video translation network for an input resolution of  $256 \times 256$ : The encoder has 8 downsampling modules with (64, 128, 256, 512, 512, 512, 512, 512) output channels. The decoder has 8 upsampling modules with (512, 512, 512, 512, 256, 128, 64, 3) output channels. The upsampling modules use the following dropout probabilities (0.5, 0.5, 0.5, 0, 0, 0, 0, 0). The first downsampling and the last upsampling module do not employ batch normalization (BN). The final non-linearity (TanH) brings the output to the employed normalized  $[-1, +1]$ -space.



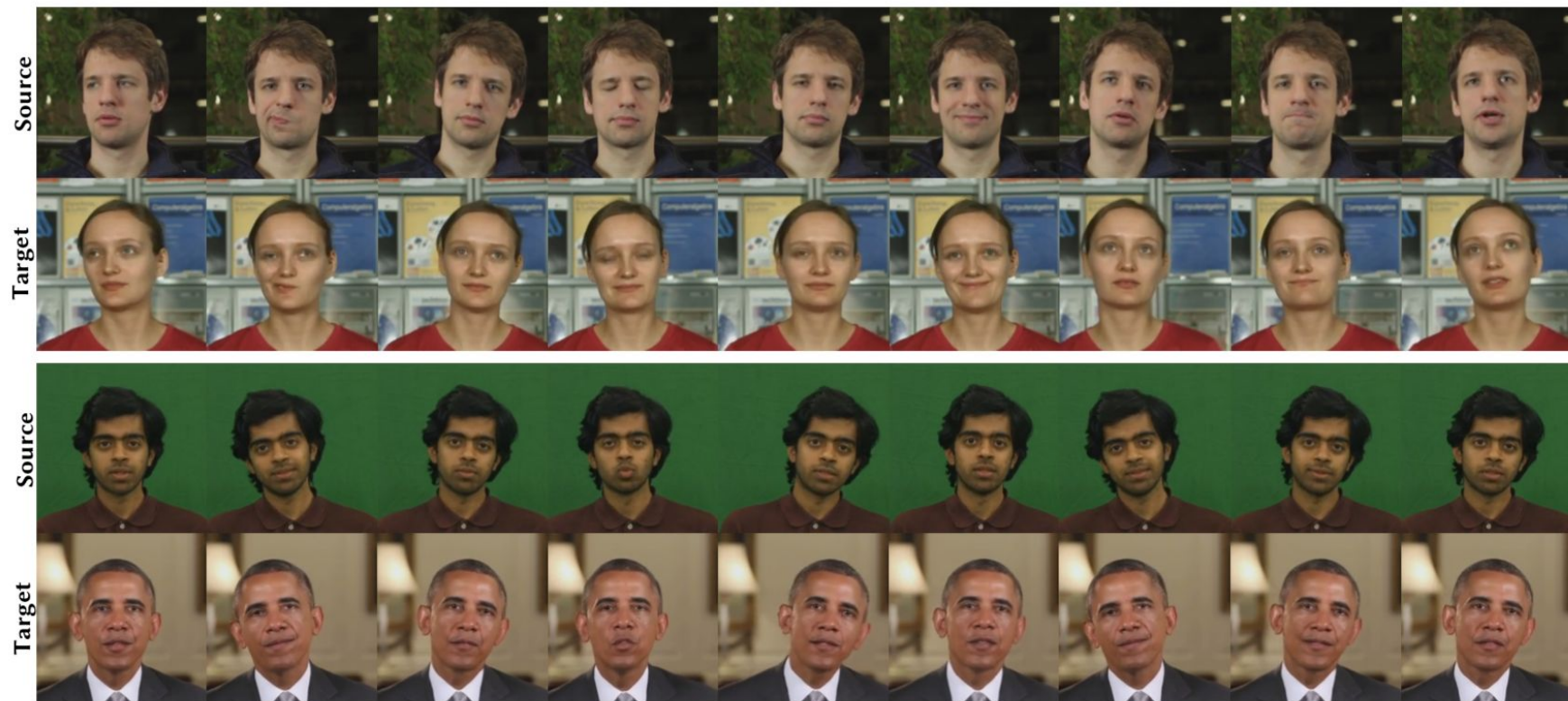
# Общая схема

Deep Video Portraits • 163:3





# Примеры



# Примеры



# Литература

Deep Video Portraits - <https://arxiv.org/pdf/1805.11714.pdf>