📣

Mobile Face Swap Review

Object : Make face swap network light

Contribution

1. Real time framework for video face swapping
2. Present an Identity Injection Module (IIM)
3. For stabilize the learning process, train network using knowledge distillation framework and propose a loss re-weighting module

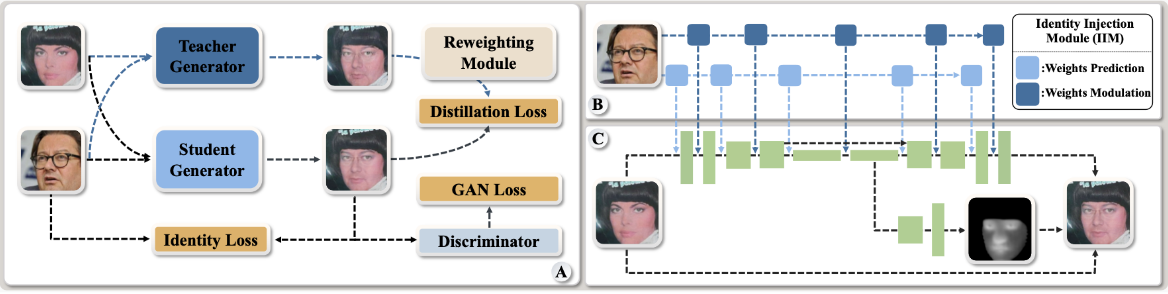
Related Work

- **Face swap method** : source-oriented, target-oriented
    - source - oriented : first transfer the pose and expression of the source image to the target image, and then apply a blending method to obtain the swapped face image. But these methods are unstable and prone to generating artifacts.
    - target - oriented : blend the feature of the source and target image to obtain the swapped face. But this methods have many parameters and computation time.

- **Dynamic neural networks** refers to one that adapts its structure or parameters to the input during inference, which can result in greater computational efficiency.

- **Knowledge distillation** is transferring the knowledge in a larger teacher network to a smaller student network.

Method



- Main two Network Architecture
    - IIN(Identity Injection Network)

      Have several identity injection modules(IIM) to obtain the parameters required.

    - IDN(Identity-aware Dynamic Network)

      Use these parameters to construct a lightweight network for inference process

- IDN = Change U-Net's standard convolution to depthwise and pointwise convolution.
- To keep the background and hair of the generated images, they propose semantic fusion module to merge the back ground of the target image. (reusing the feature map of IDN and exceptionally light weight module)
- **knowledge distillation** → Loss definition

  reconstruction loss → L1

  perception loss → i-th layer with N_i elements of the VGG net.

$$\mathcal{L}_{rec} = ||I'_g - I_g||,$$
$$\mathcal{L}_{per} = \sum_{i=1}^{L} \frac{1}{N_i} ||\mathcal{F}^{(i)}_{VGG}(I'_g) - \mathcal{F}^{(i)}_{VGG}(I_g)||^2$$

- But, there are some failure cases and they divide two case. One is some teacher outputs cannot keep the identity well of the source images. Another is some teacher outputs may have unnatural results or artifacts(noise). → to solve this problem → they propose a loss **reweighting module**.

$$\alpha = Cos(z_{id}, z'_{id})^2 \times Q(I'_g)$$

- **id and mask loss**

- **Total loss** (labda_rec = 30, labda_per = 5, labda_mask = 10)

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{adv} + \alpha(\lambda_{rec}\mathcal{L}_{rec} + \lambda_{per}\mathcal{L}_{per}) \\ +\lambda_{id}\mathcal{L}_{id} + \lambda_{mask}\mathcal{L}_{mask},\end{aligned}$$

Result

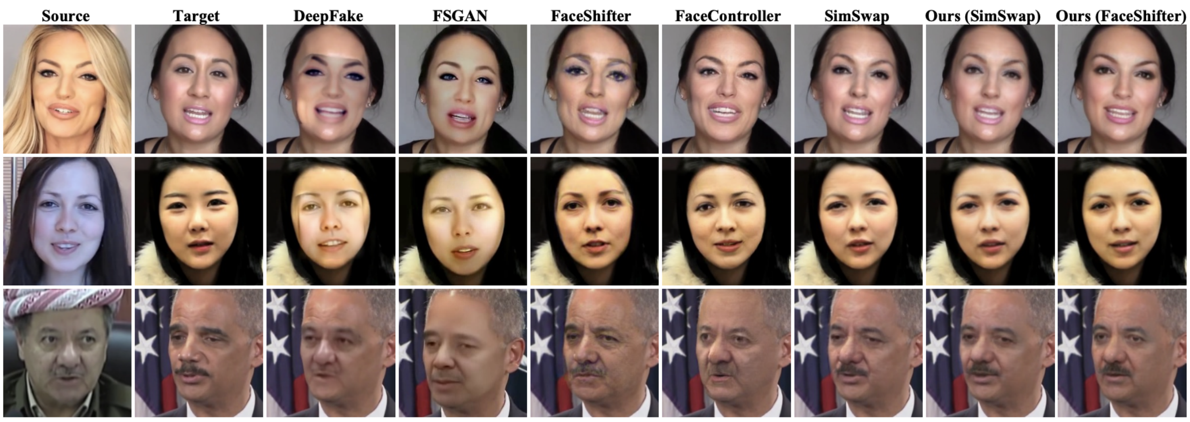


Figure 6: Comparison with DeepFakes, FSGAN, FaceShifter, FaceController, and SimSwap on the FaceForensics++ dataset.

Method	Size	(I) Params (M)	(I) FLOPs (G)	(V) Params (M)	(V) FLOPs (G)	(V) FPS	Id↑	Pose↓
DeepFakes	64	82.1	**1.90**	82.1	1.90	9.5	81.96	4.14
FSGAN	256	226	2440	226	2240	-	57.34	3.81
SimSwap	224	107	55.7	45.6	48.2	0.64	92.83	1.53
FaceShifter	256	421	97.4	350	91.1	-	97.38	2.96
FaceController	224	306	192	236	177	-	**98.27**	2.65
Teacher (SimSwap)	224	107	55.7	45.6	48.2	0.64	95.94	1.39
Teacher (FaceShifter)	256	421	97.4	350	91.1	-	97.15	1.76
Ours (SimSwap)	224	**72.8**	8.07	**0.50**	**0.33**	**25.6**	95.98	**1.32**
Ours (FaceShifter)	256	**72.8**	8.18	**0.50**	0.44	19.7	96.10	1.70
Id Network	112	52.2	7.52	-	-	-	-	-

→ 우리께 param 수도 적지만, 높으면 좋은 id도 높고, 낮으면 좋은 pose도 낮다. 즉 효율좋음!

Contribution

Suggest realtime MobileFaceSwap method. using two main efficient network IIM and IDN. And additional technique knowledge distillation and loss re-weighting module.

어떻게 mobilefaceswap이 forgery detection에 little impact 할수있는지? (결론부분)

Total loss 에 각성분 loss가 의미하는게 뭔지 궁금!(읽었는데 잘 모르겠음)

돌려본 결과 사진

```
import cv2
from google.colab.patches import cv2_imshow

img = cv2.imread(os.path.join('./results', uploaded_images[0]))
cv2_imshow(img)
```