University of Nottingham
UK | CHINA | MALAYSIA

Computer Vision Laboratory
cvl.cs.nott.ac.uk

horizon
CENTRE FOR DOCTORAL TRAINING

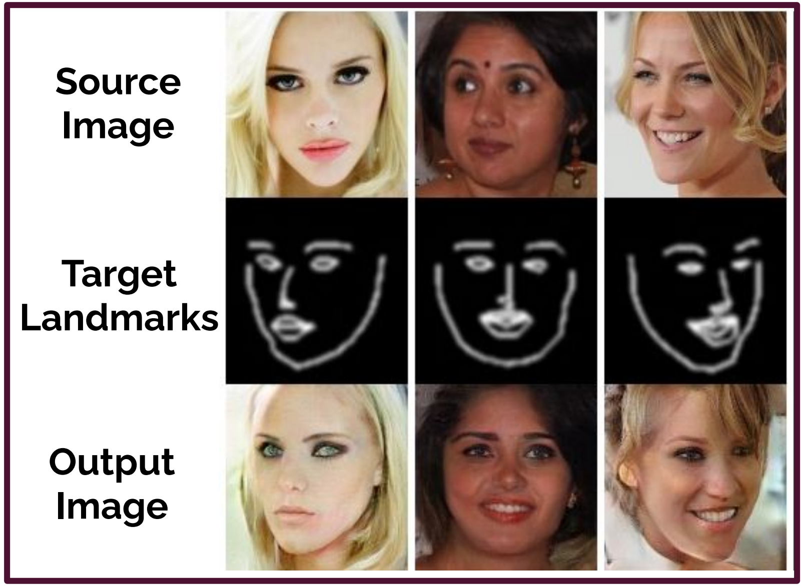
Unsupervised Face Manipulation via Hallucination

Miss. Keerthy Kusumam
Dr. Enrique Sanchez
Dr. Georgios Tzimiropoulos

Introduction

Funded by

- A framework for **face manipulation**, to synthesize novel face images that adhere to a **given pose**.
- Trained under the very general **unpaired/unsupervised** setting.
- Transfers **appearance** and **style** information from an exemplar image, in a semantically meaningful way.
- Models the general appearance, layout and background using a **low-resolution** version of input image, and **progressively** passes through a **hallucination network** to generate features at higher resolutions.
- Does not require "**paired**" data for training or **expensive** identity annotations, unlike most methods.
- **Simpler** method for appearance modelling.



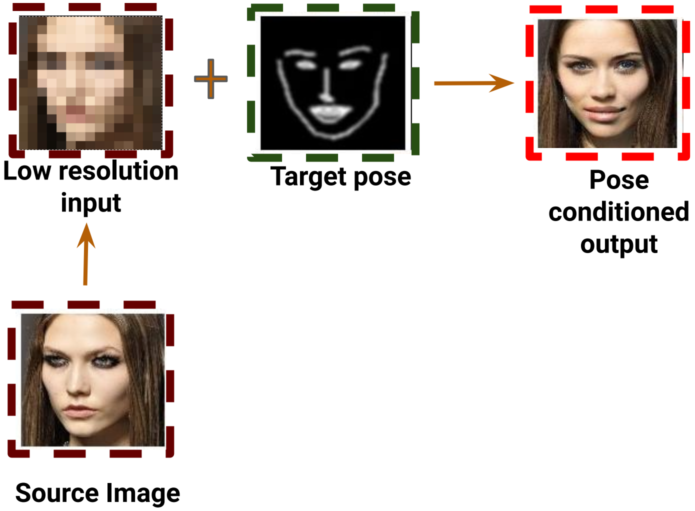
Contributions

Funded by

- Proposes an **unpaired** image-to-image translation method where a **face hallucination** network guides a **pose-synthesis** network to **manipulate** the input low resolution image **according to the target pose**.

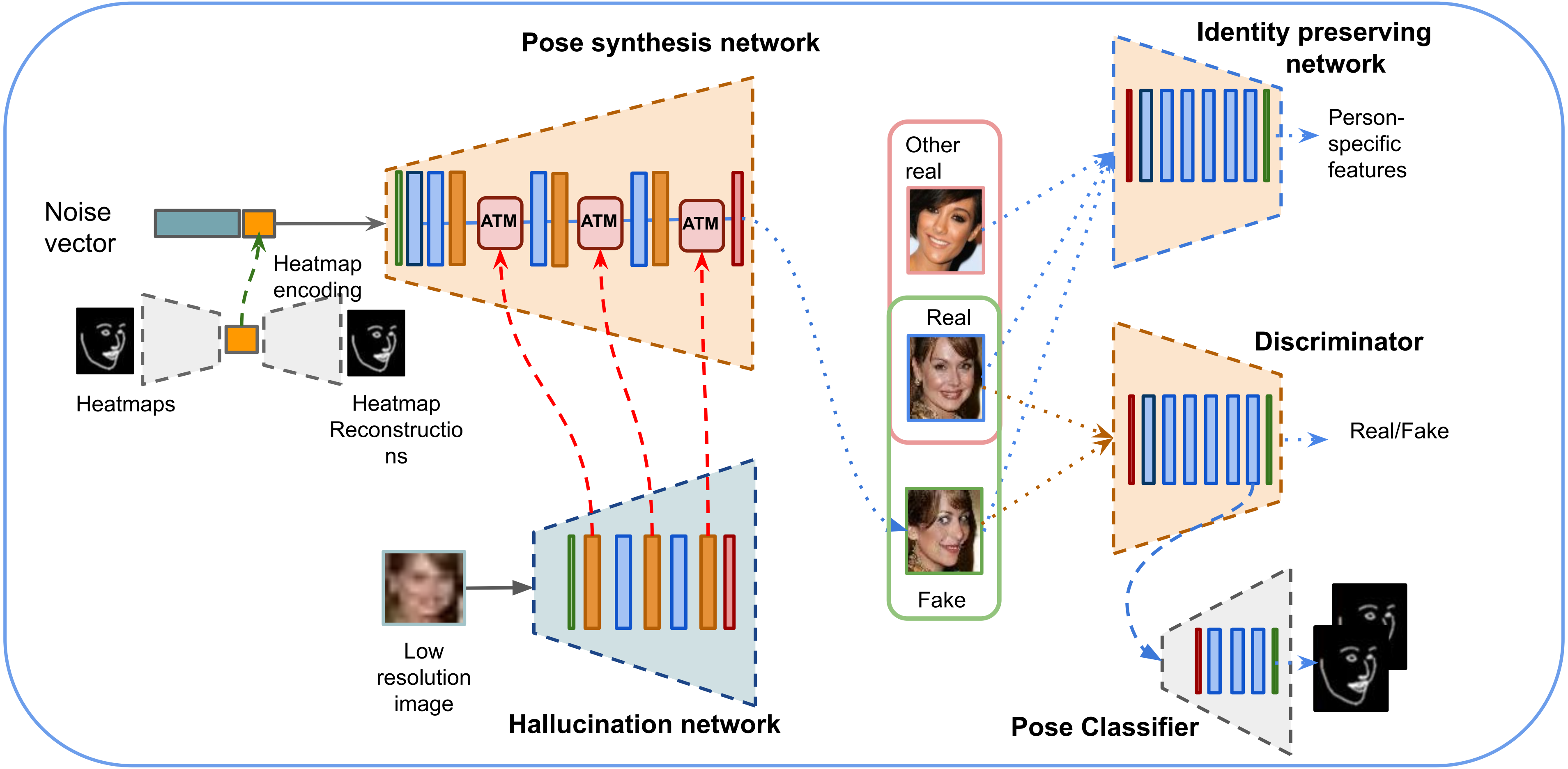
- Introduces an **Appearance Transfer Module**, a fully trainable **spatially-aware** module to deal with the **misalignment** between the **hallucination features** and those generated by the **pose-conditional GAN**.

- Proposes **pose preservation** and **identity preserving** methods that are trained in an **unsupervised** way, using an **auxiliary** pose classifier and identity classifier.



- Demonstrates both **quantitatively** and **qualitatively** the capability of the method to achieve high quality images that are both conditioned on target poses and source appearances.

Proposed Approach

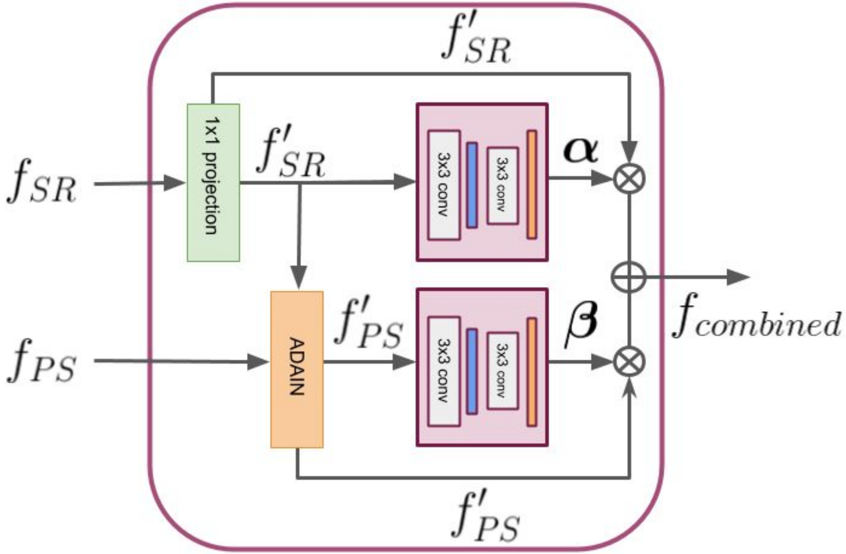


Appearance Transfer Module

- Uses Adaptive Instance Normalisation to align **feature statistics**.
- **Combines** features from the hallucination network with the pose-synthesis network, as the **weighted combination** that is **learned** from the ATM.

$$f'_{PS} = \sigma(f'_{SR})\left(\frac{f_{PS} - \mu(f_{PS})}{\sigma(f_{PS})}\right) + \mu(f'_{SR}),$$

$$f_{combined} = \alpha \odot f'_{SR} + \beta \odot f'_{PS}.$$



Training

Funded by
horizon
CENTRE FOR DOCTORAL TRAINING
University of Nottingham
UK | CHINA | MALAYSIA

Total Loss is given by,

$$L_D = \lambda_{adv} L_{adv}^D + \lambda_p L_p,$$

$$L_G = \lambda_{adv} L_{adv}^G + \lambda_p L_p + \lambda_r L_r + \lambda_{con} L_{con},$$

Discriminator's Adversarial Loss

$$\begin{aligned}\mathcal{L}_{adv}^D &= \mathbb{E}_{z,I_X,Y}[\min(0, -D(G(z, I_X, Y)) - 1)] \\ &+ \mathbb{E}_{I_X}[\min(0, -1 + D(I_X))]\end{aligned}$$

Generator's Adversarial Loss

$$\mathcal{L}_{adv}^G = -\mathbb{E}_{z,I_X,Y}[D(G(z, I_X, Y))]$$

Pose Preservation, Classifier Loss

$$L_p = \mathbb{E}_{z,I_X,Y,X} \|D(G(z, I_X, Y)) - Y\|^2 + \|D(I_X) - X\|^2$$

Reconstruction Loss

$$\mathcal{L}_r = \mathbb{E}_{z,I_X,Y} \|I_X - G(z, \hat{I}_Y, X)\|^2$$

Identity Preservation, Contrastive Loss

$$\mathcal{L}_{con} = (1 - y_{ij})(\Delta f_a^{ij})^2 + y_{ij} \max(0, m - \Delta f_a^{ij})^2,$$

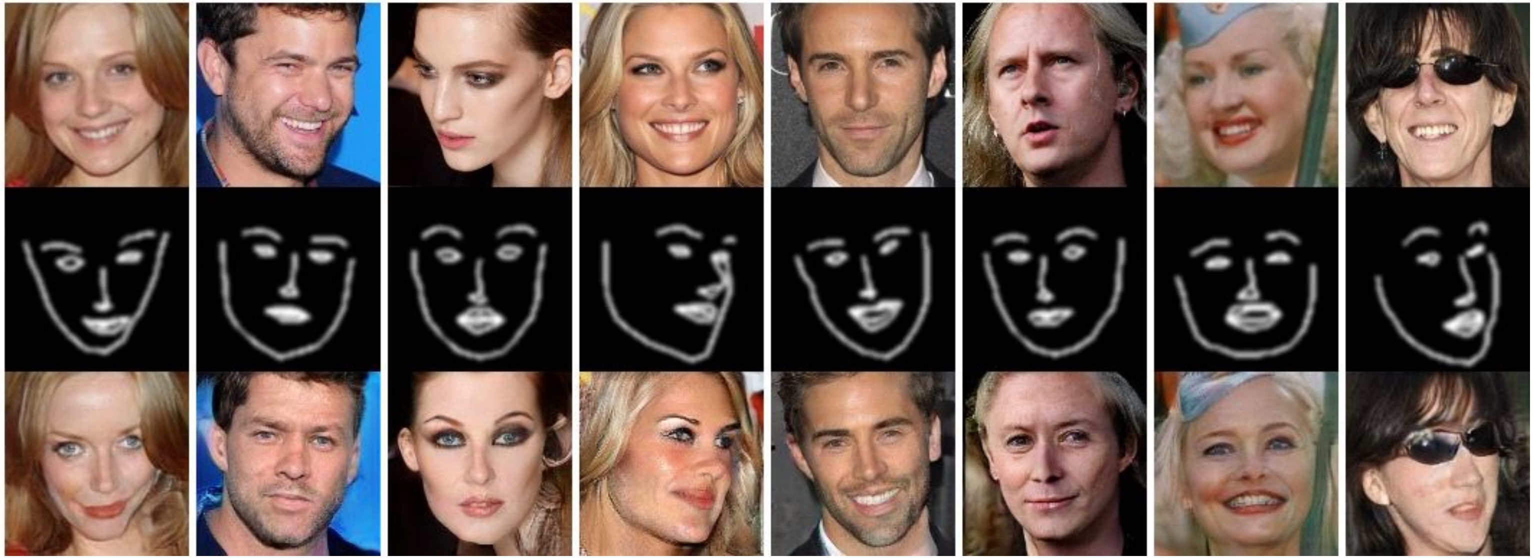
Results - CelebA

Funded by
horizon
CENTRE FOR DOCTORAL TRAINING
University of Nottingham
UK | CHINA | MALAYSIA

Source Image

Target Landmarks

Output Image



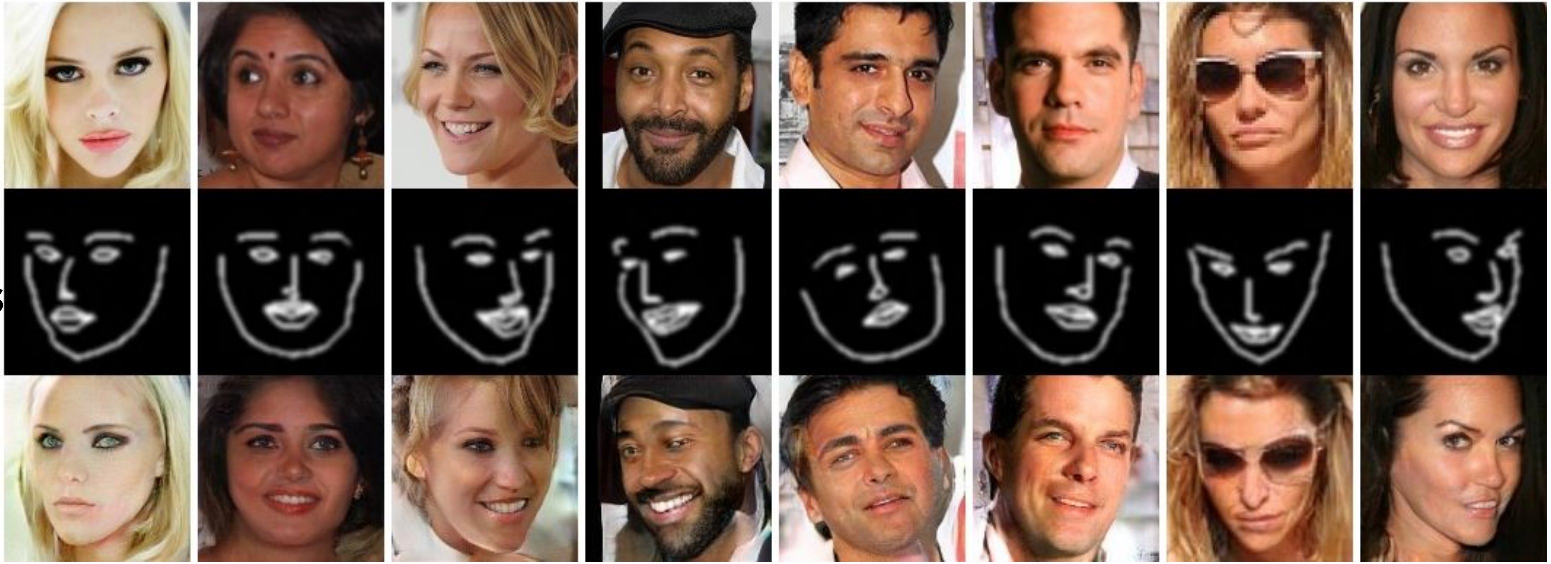
Results - CelebA

Funded by
horizon CENTRE FOR DOCTORAL TRAINING
University of Nottingham
UK | CHINA | MALAYSIA

Source Image

Target Landmarks

Output Image

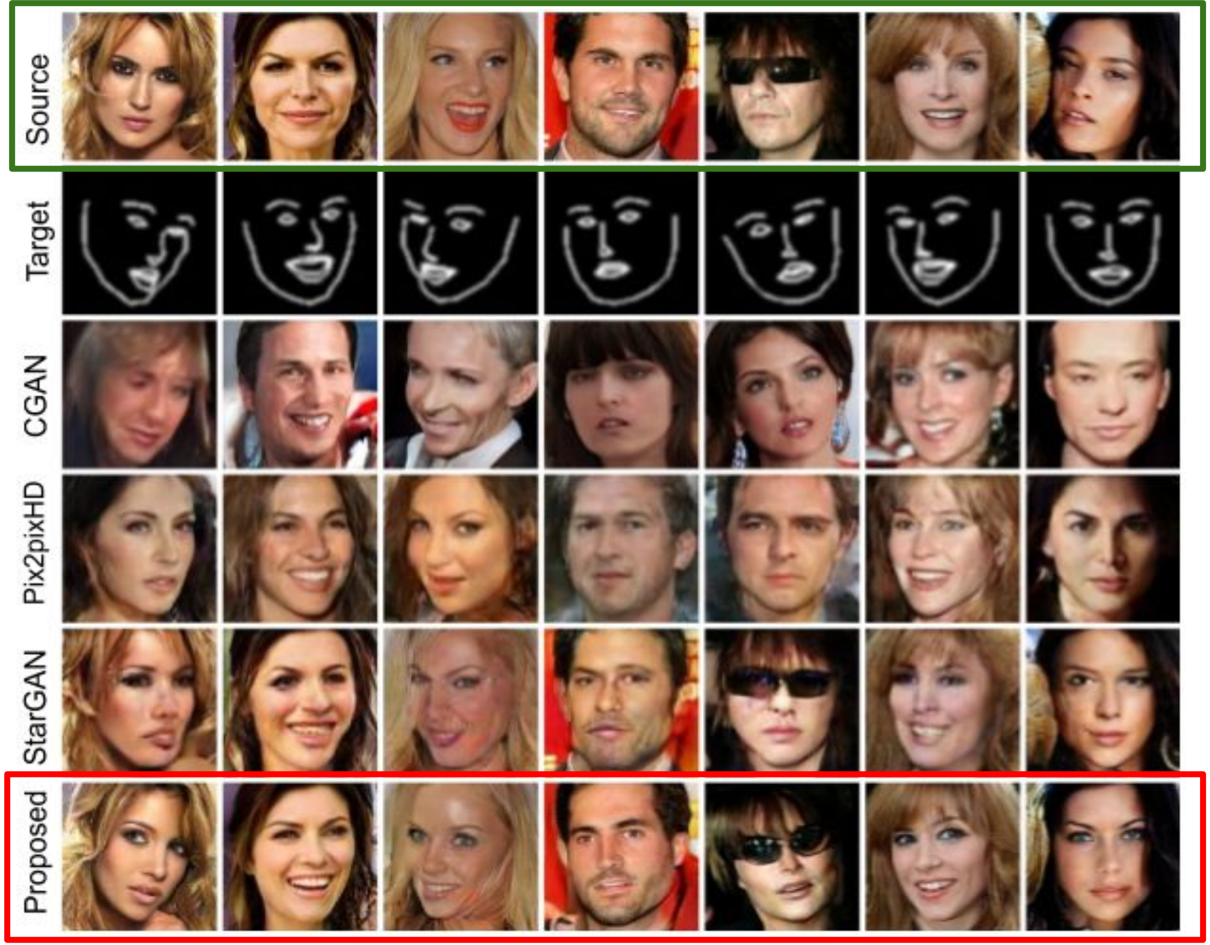


Comparison with SOTA

Funded by

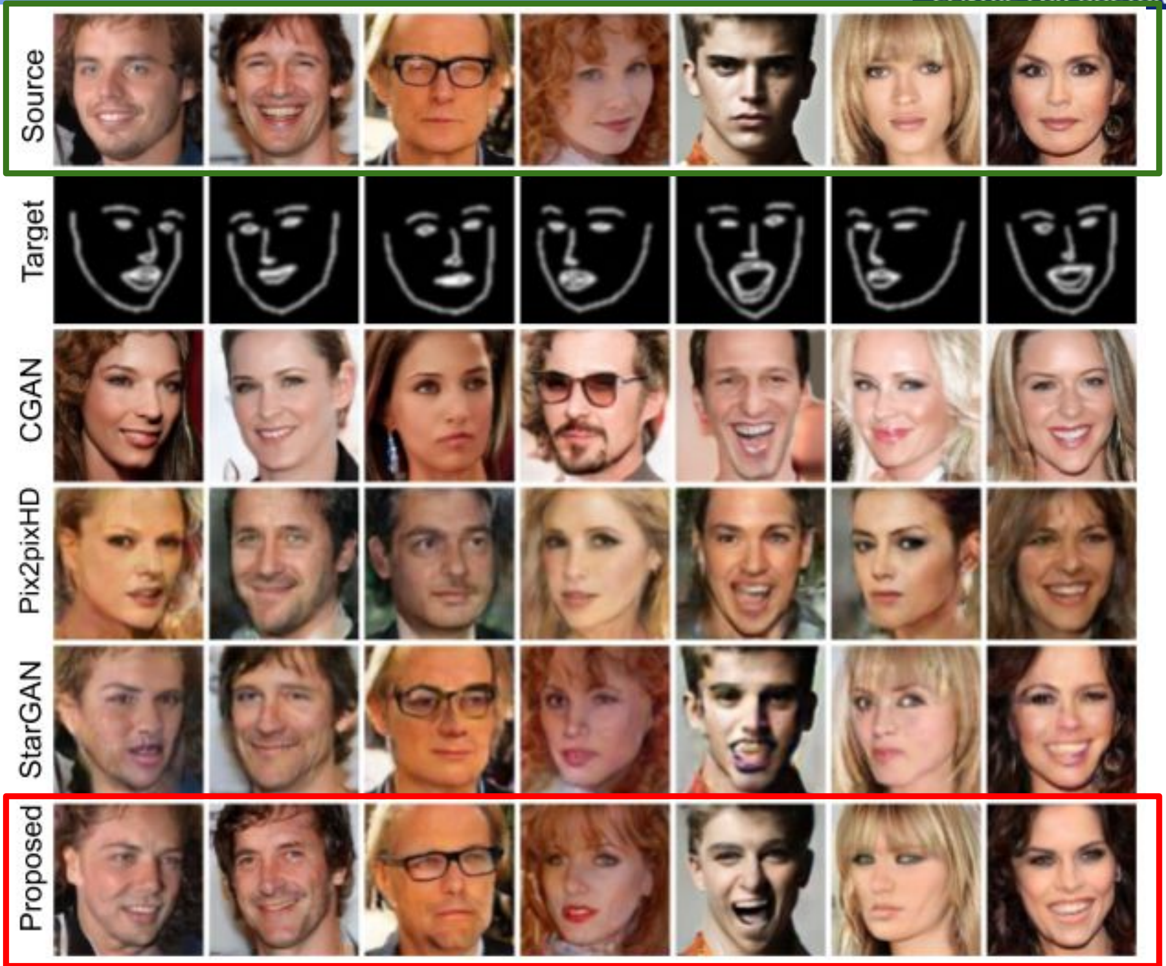
Comparison with SOTA

Funded by



Comparison with SOTA

Funded by



Results - Multi-PIE

Funded by

Simultaneous pose and expression manipulation on Multi-PIE dataset.

Expression only manipulation on Multi-Pie dataset.

Quantitative Results

Funded by horizon CENTRE FOR DOCTORAL TRAINING

University of Nottingham UK | CHINA | MALAYSIA

➢ Evaluation of generated images using FID and inception score metrics for quantifying perceptual quality.

➢ Achieves state-of-the-art performance with respect to other unpaired generation methods.

	Method	FID↓	IS↑		Method	FID↓	IS↑
MultiPIE	Real data	0.00	2.14	CelebA	Real data	0.01	3.49
	CGAN	22.9	1.79		CGAN	7.40	2.42
	Pix2pixHD	19.30	1.58		Pix2pixHD	41.68	2.62
	StarGAN	25.29	**1.81**		StarGAN	12.78	2.55
	Ours	**15.90**	1.78		Ours	**6.14**	**2.65**

➢ Quantitative comparison of different baselines and ablation studies for CelebA.

	Method	FID↓	IS↑
	Real data	0.00	3.01
(A.1.1)	AE-v1	35.06	2.53
(A.1.2)	AE-v2	10.05	2.20
(A.2.1)	Hourglass-v1	5.53	2.49
(A.2.2)	Hourglass-v2	5.18	2.49
(A.3)	Proposed w/o identity	**4.31**	[2.57]
	Proposed	[4.6]	**2.61**

Conclusions

Funded by
horizon CENTRE FOR DOCTORAL TRAINING
University of Nottingham UK | CHINA | MALAYSIA

- ➢ We proposed an **unsupervised face manipulation framework** that synthesizes unseen facial images given a specific pose, whilst **preserving appearance** from an exemplar image.
- ➢ We show how to **integrate appearance features** of the **exemplar image** taken from a pre-trained **hallucination network** into the generation process of a **conditional GAN** using a novel **appearance transfer module.**
- ➢ We demonstrate both **quantitatively** and **qualitatively** the capability of the method to achieve high quality images that are both **conditioned** on **target poses** and **source appearances**.
- ➢ Our method has applications in character **animation**, data **anonymization**, data **augmentation** and **generalisation** techniques.