BMVC 2023

Face Aging via Diffusion-Based Editing

Xiangyi Chen, Stéphane Lathuilière
Télécom Paris - Institut Polytechnique de Paris, Shanghai Jiao Tong University

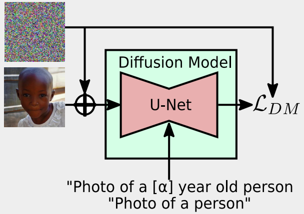
Introduction

- Existing **GAN-based** methods rely solely on face image datasets. Limited scale and bias of the dataset leads to **poor performance on rare cases** (extreme age, facial accesories, etc.).
- Recently proposed **Diffusion Models** exhibit superior generation quality compared to GANs, but no previous work on extending them to specific image-editing tasks.

Method (1/2): Specialization stage

Repurpose the pre-trained text-to-image diffusion model for face-aging task

- For every face image x with estimated age α, perform **fine-tuning** with text-image pair (x, *"photo of a [α] year old person"*)
- **Double-prompt scheme:** add another age-agnostic prompt *"photo of a person"*. Better disentanglement of age information from age-irrelevant features
- Training loss



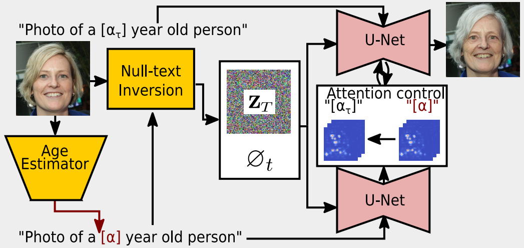
$$\mathcal{L}_{DM} = \mathbb{E}_{\mathbf{z}_0 \sim \mathcal{E}(x), \alpha, \epsilon, \epsilon', t}[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t, \mathcal{P})\|_2^2 + \|\epsilon' - \epsilon_\theta(\mathbf{z}'_t, t, \mathcal{P}_\alpha)\|_2^2]$$

Method (2/2): Age editing stage

Image Inversion

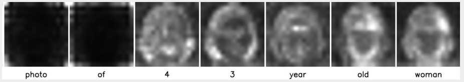
Invert input image with its estimated age to initial noise and optimized null-text embedding

$$\min_{\varnothing_t} \|z_{t-1}^{inv} - z_{t-1}(\bar{z}_t, t, \mathcal{P}_{inv}; \varnothing_t)\|_2^2$$



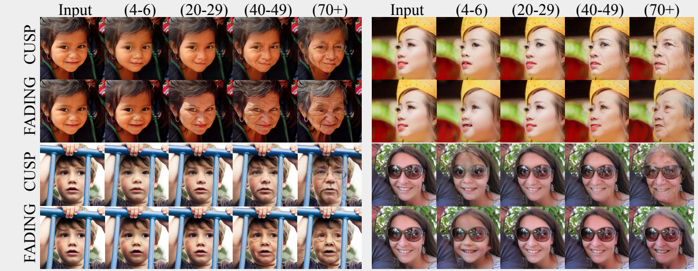
Cross attention control

- **Cross attention maps** contain rich semantic relations between spatial layout and age information in text prompt
- Replace estimated age with target age to guide the new diffusion process while **swapping attention maps**



Cross attention maps during diffusion process

Experiments and results



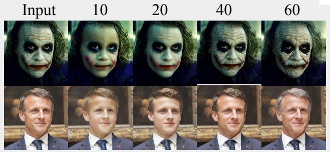
Qualitative comparison with SOTA method (CUSP) on FFHQ dataset

Metrics

- **Aging accuracy:** age MAE
- **Aging quality**: KID
- **Age-irrelevant attribute preservation:** Attribute(%)

Table 2: Quantitative comparison between CUSP and FADING on FFHQ-Aging.

Metric	Method	0-2	3-6	7-9	10-14	15-19	20-29	30-39	40-49	50-69	70+	Mean
MAE	CUSP	9.41	16.28	20.24	18.16	11.88	10.36	12.70	**11.08**	**8.13**	8.05	12.63
	FADING	**5.70**	**11.72**	**13.66**	**11.22**	**6.86**	**6.23**	**9.60**	12.04	8.39	**6.20**	**9.16**
Gender(%)	CUSP	71.5	**73.5**	**74.5**	**78.0**	73.5	80.5	85.5	81.5	82.0	76.0	77.7
	FADING	**72.0**	72.0	67.5	68.0	**88.0**	**96.0**	**98.0**	**97.0**	**95.0**	**87.5**	**84.1**
KID(×100)	CUSP	4.19	3.22	3.14	3.18	3.60	3.63	3.98	4.69	4.07	4.57	3.83
	FADING	**1.41**	**0.11**	**0.45**	**0.25**	**0.52**	**0.16**	**1.00**	**0.59**	**1.50**	**0.61**	**0.66**



Generalization on out-of-dataset examples

Table 3: Ablation study on the Specialization stage

Method	Spec.	DP	MAE	Gender	Smiling	Happy	Neutral	Blur	KID(×100)
Training-free ~[37]	✗	-	9.295	82.40	82.95	78.35	78.80	2.226	0.668
Single prompt	✓	✗	**8.781**	81.95	85.05	81.55	81.05	2.275	0.707
Full	✓	✓	9.162	**84.10**	**86.60**	**81.95**	**81.75**	**2.030**	**0.660**

Conclusion

- **First work to extend large-scale diffusion models for face aging.**
- Successfully leverage attention mechanism for age manipulation and disentanglement
- Qualitatively and quantitatively demonstrate the superiority over state-of-the-art methods in terms of agin accuracy, attribute preservation, aging quality and generalization.
- Code available at https://github.com/MunchkinChen/FADING