SHMT: Self-supervised Hierarchical Makeup Transfer via Latent Diffusion Models

Zhaoyang Sun[1,3] Shengwu Xiong[1,2,5] Yaxiong Chen[1] Fei Du[3,4]
Weihua Chen[3,4] Fan Wang[3,4] Yi Rong[1,2]

[1]School of Computer Science and Artificial Intelligence, Wuhan University of Technology
[2]Sanya Science and Education Innovation Park, Wuhan University of Technology
[3]DAMO Academy, Alibaba Group *[4]Hupan Laboratory* *[5]Shanghai AI Laboratory*

Work done during internship of Zhaoyang Sun at DAMO Academy, Alibaba Group
Corresponding author: Yi Rong (yrong@whut.edu.cn)

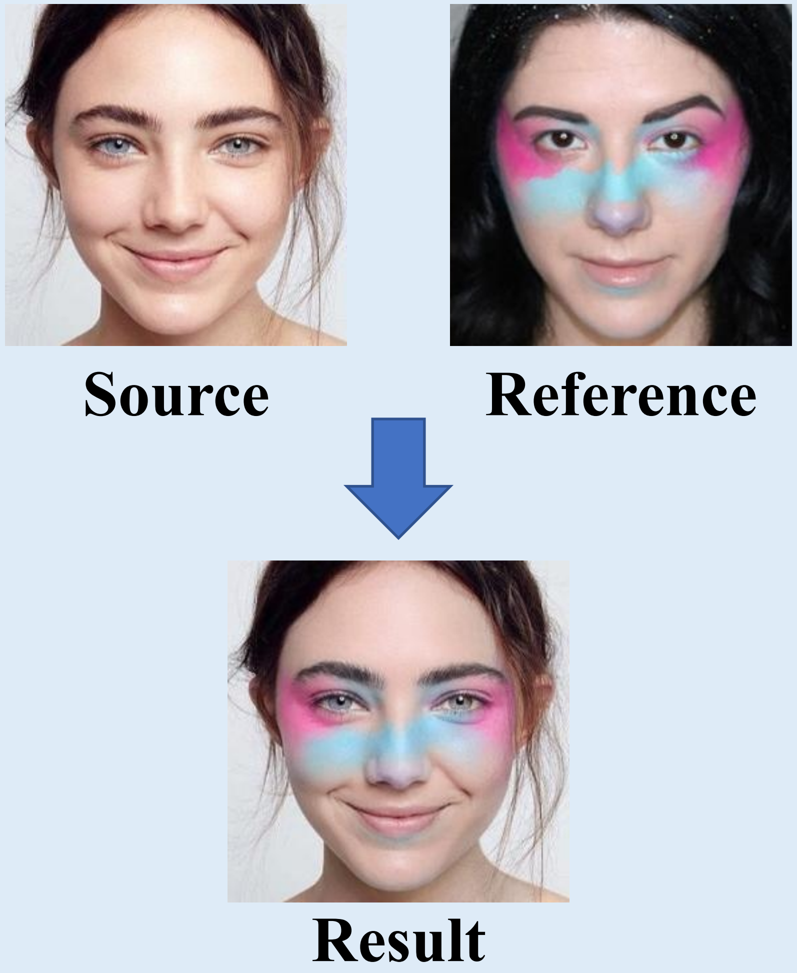
The goal of makeup transfer

Given a pair of source and reference face images, the main goal of makeup transfer is to generate an image that simultaneously satisfies the following conditions:

（1）Containing the makeup styles transferred from the reference image, such as lipstick, eye shadow and powder blush

（2）Preserving the content details of the source image, including identity, facial structure and background.

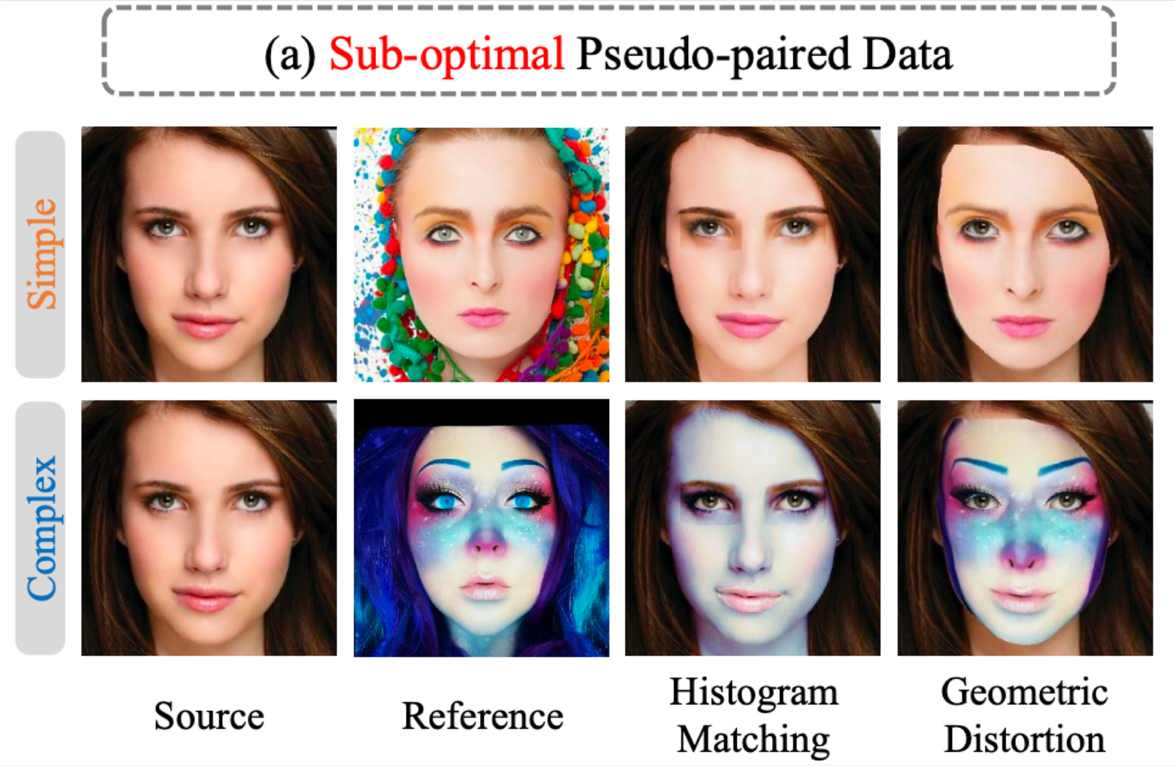
(3) High quality and realistic synthesis results



Challenge1: unsupervised task

Makeup transfer is essentially an unsupervised task, which means that there are no real transferred images that can be used as labeled targets for model training.

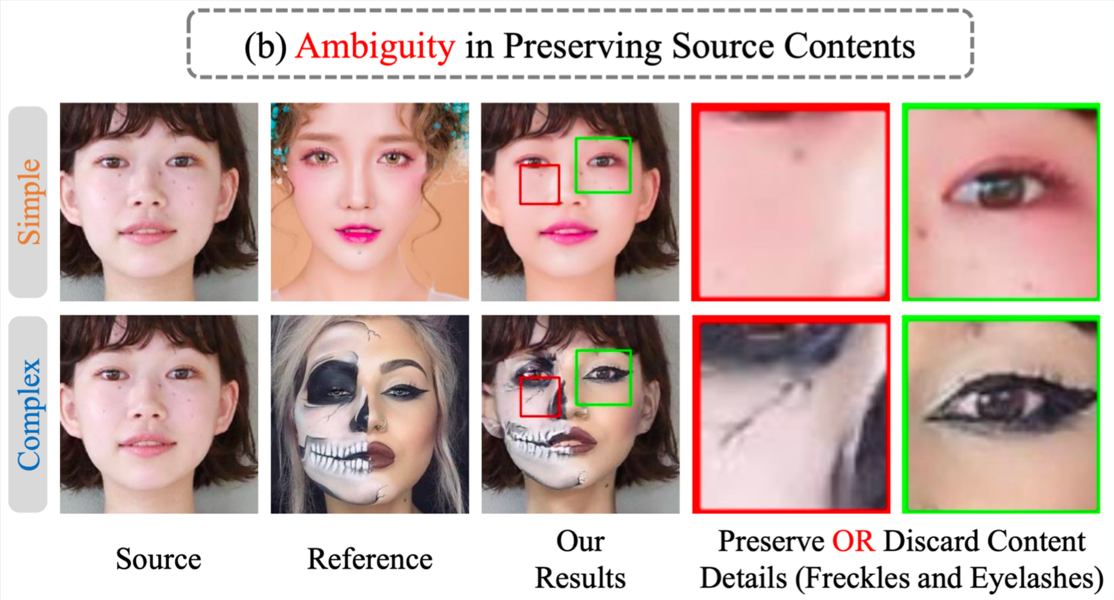
To address this issue, previous methods typically synthesize a "pseudo" ground truth from each input source-reference image pair, as an alternative supervision. Consequently, these **sub-optimal pseudo-paired data** will inevitably misguide the model training.



Challenge2: the diversity of makeup styles

The diversity of different makeup styles can also lead to **ambiguity in preserving source contents.**

In practice, makeup styles can range from natural, barely-there looks to elaborate and dramatic ones, each having a different impact on the person face

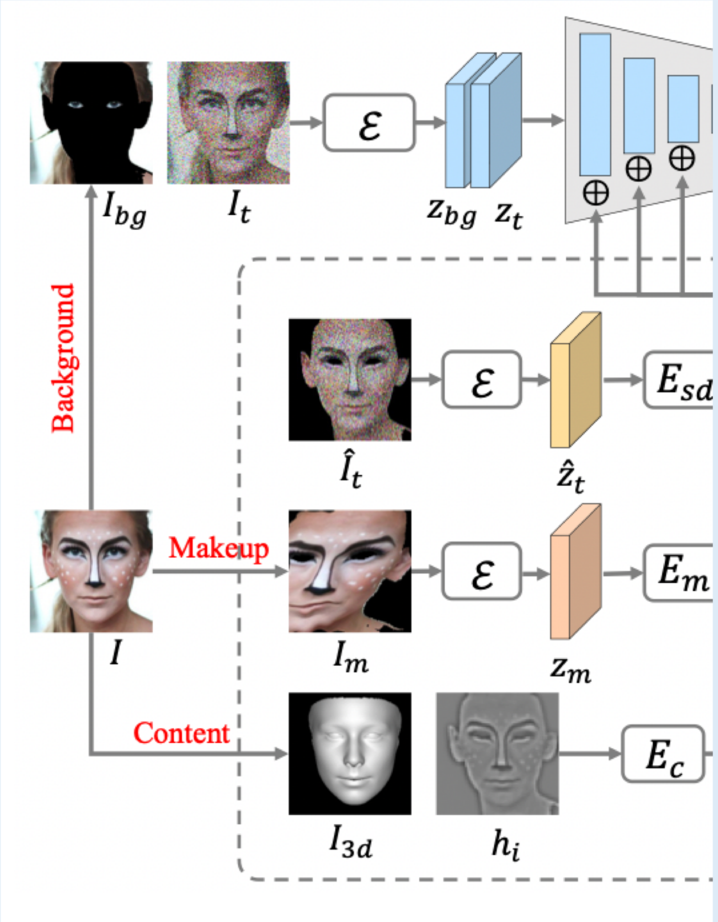


Our Method: SHMT

(1) A self-supervised manner.

Following a "decoupling-and-reconstruction" paradigm, we craft a self-supervised strategy for makeup transfer.

The **main idea** is to separate content and makeup representations from a facial image, and then reconstruct the original image from these components.

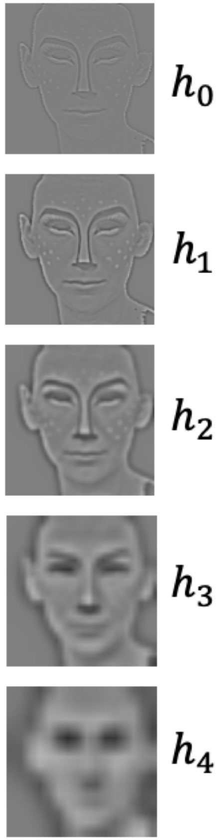


Our Method: SHMT

(2) Hierarchical texture details.

When injecting a **fine texture** detail, our model only needs to distill the low-frequency makeup information from the makeup representation to reconstruct the image. This **is suitable for simple makeup styles**.

When injecting a **coarse texture** detail, our model must also distill the high-frequency makeup information from the makeup representation to ensure the recovery of the image. This **is suitable for complex makeup styles**.



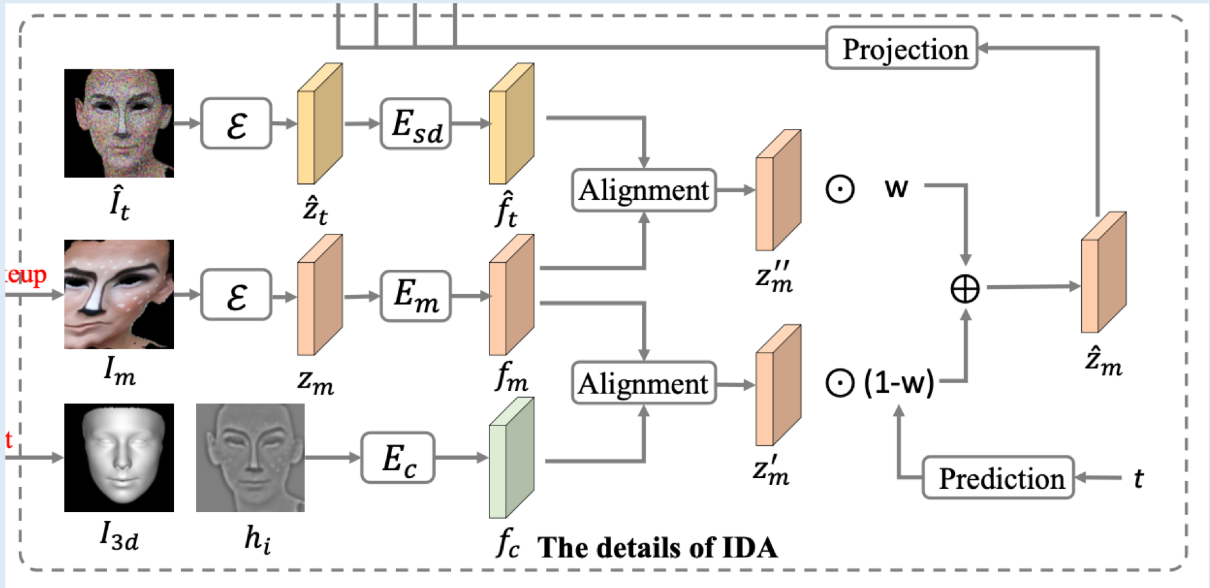
Hierarchical texture details

Our Method: SHMT

(3) Iterative Dual Alignment.

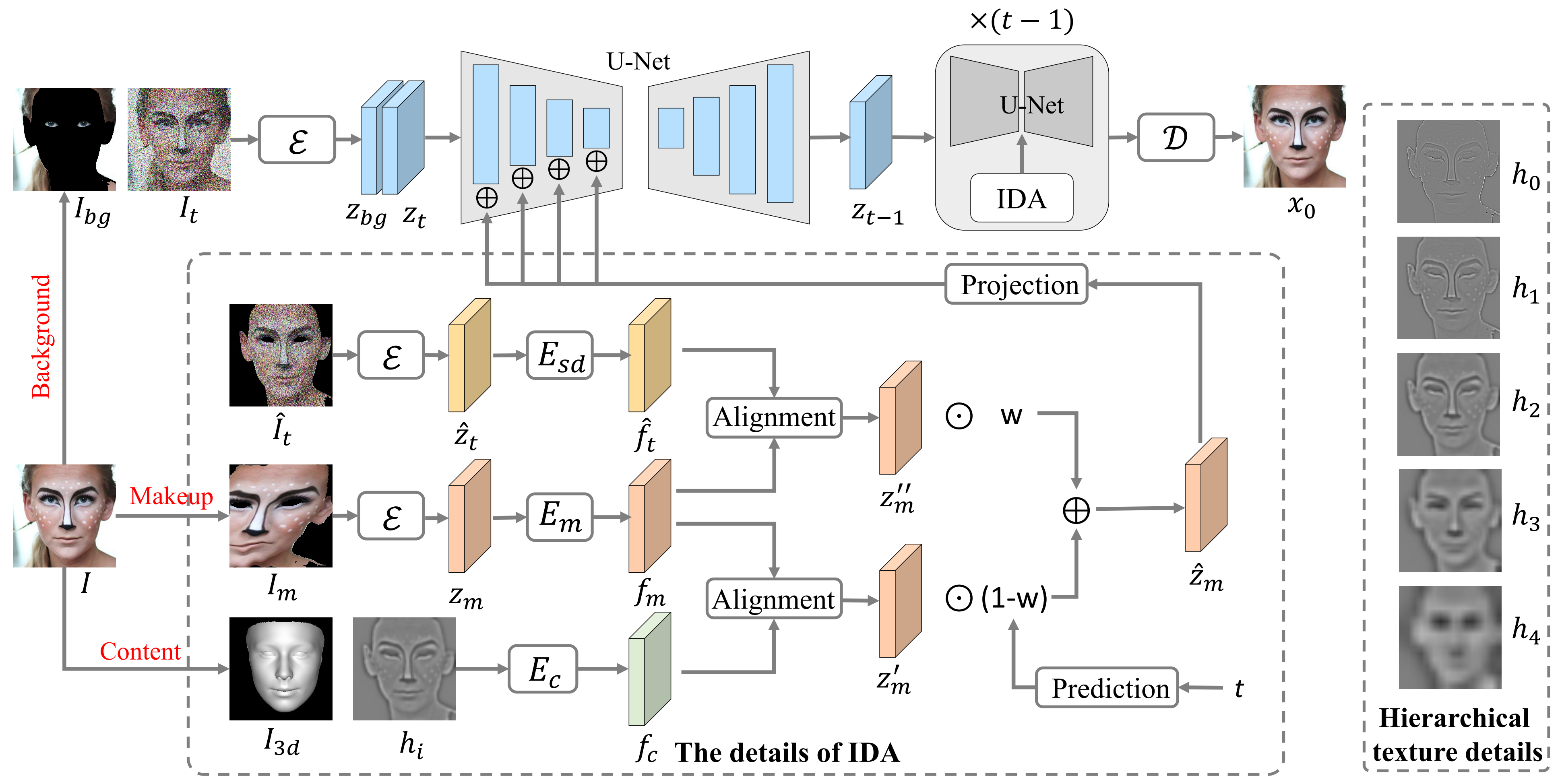
Due to the domain gap between content and makeup representations, we find that alignment errors occur frequently.

Considering the property that **noisy intermediate result is gradually moving closer to the real image domain** (e.g., the makeup representation domain), we propose a Iterative Dual Alignment (IDA) module to address the above issue.



The details of IDA

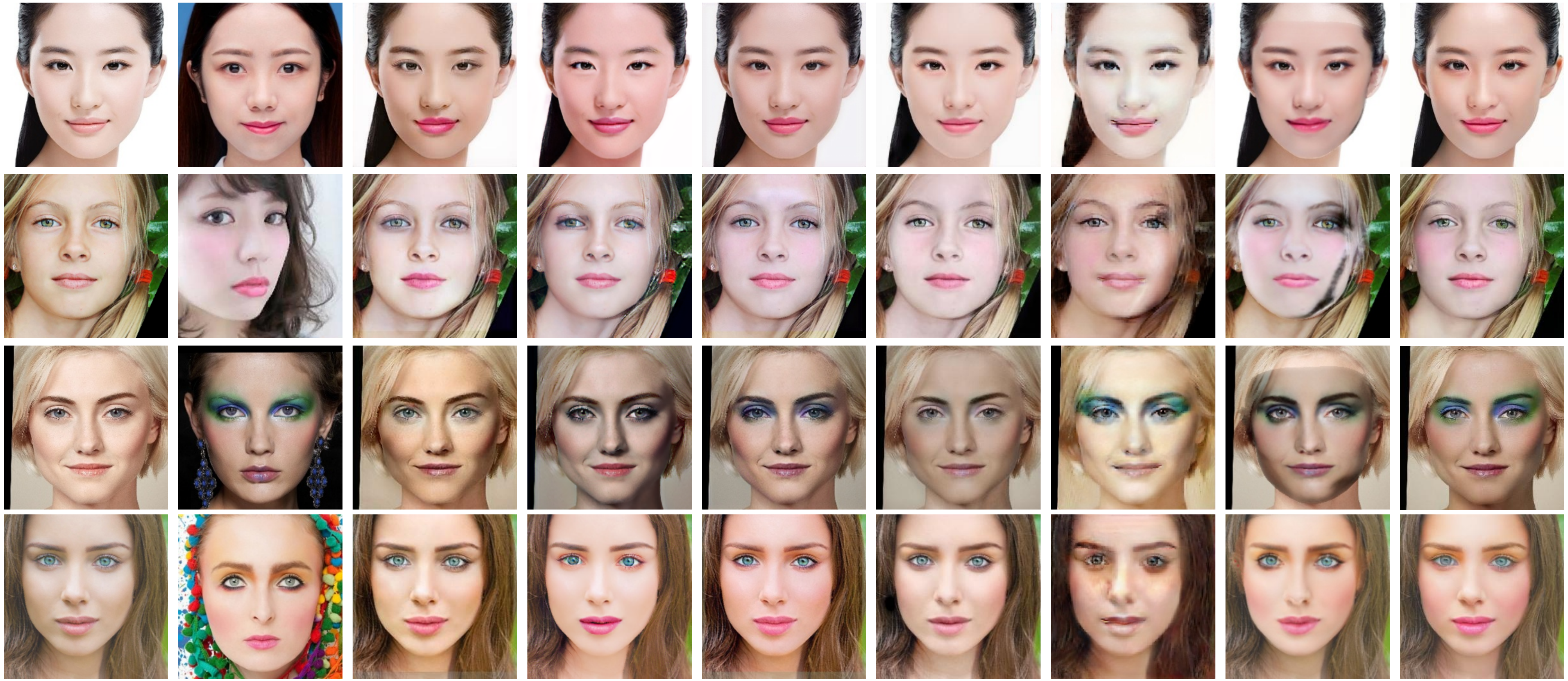
Our Method: SHMT



The overall framework of SHMT

Qualitative experiment

Simple makeup styles (using fine texture detail)



Source	Reference	PSGAN	SCGAN	EleGANt	SSAT	LADN	CPM	SHMT-h_0

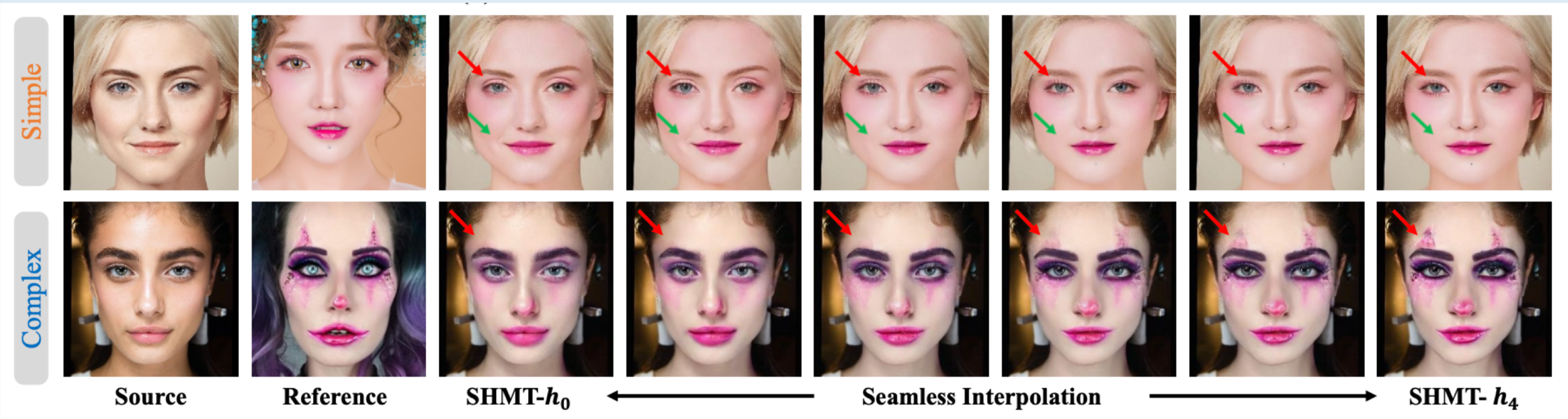
Qualitative experiment

Complex makeup styles (using coarse texture detail)

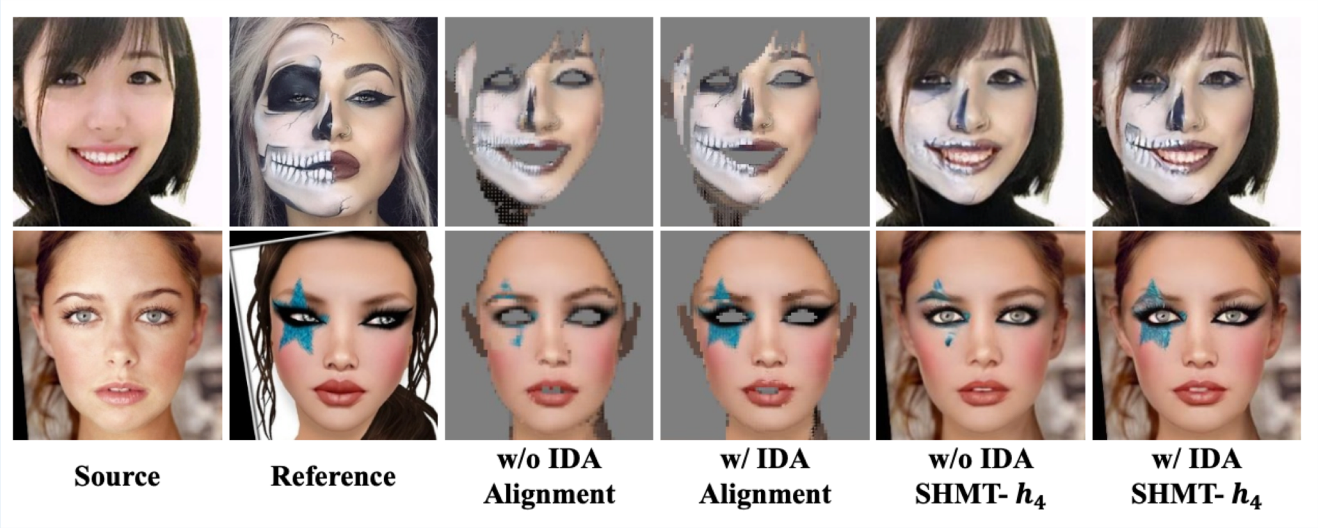
Source	Reference	PSGAN	SCGAN	EleGANt	SSAT	LADN	CPM	SHMT-h_4

Qualitative experiment

The effectiveness of hierarchical texture details

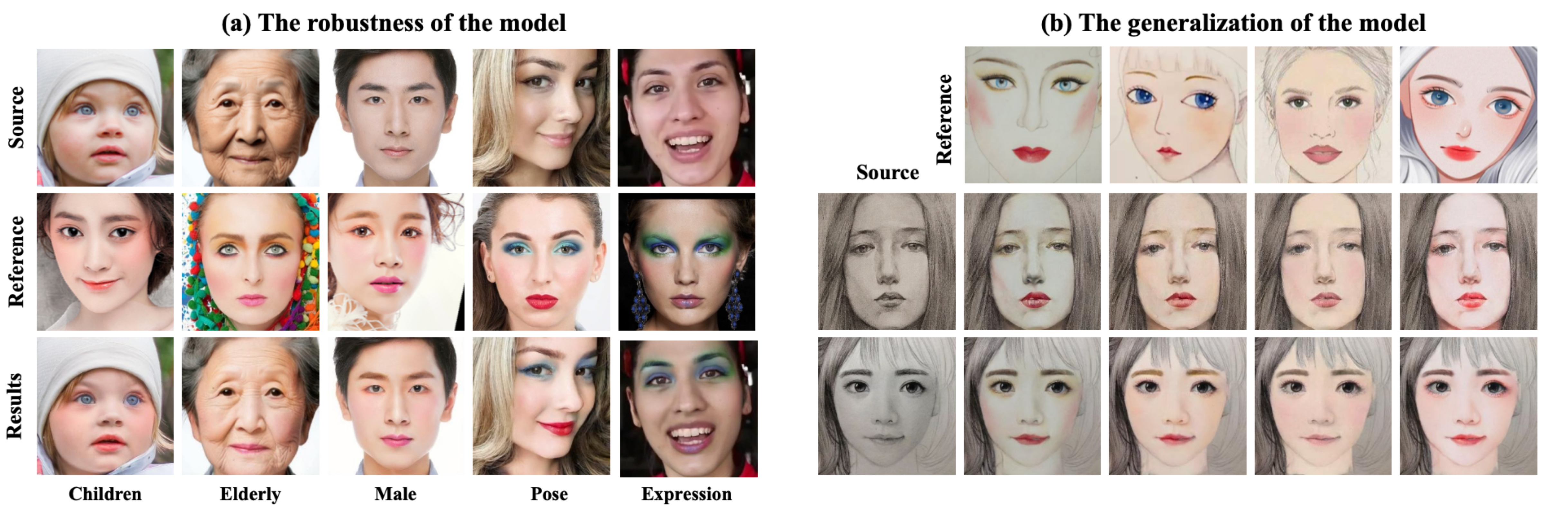


The effectiveness of IDA



Qualitative experiment

The robustness and generalization ability of SHMT



Quantitative experiment

DAMO ACADEMY

NEURAL INFORMATION PROCESSING SYSTEMS

Methods	MT			Wild-MT			LADN		
	FID	*CLS*	*Key-sim*	*FID*	*CLS*	*Key-sim*	*FID*	*CLS*	*Key-sim*
PSGAN [17]	45.02	0.628	0.975	89.92	0.642	0.969	57.80	0.684	0.975
SCGAN [7]	39.20	0.636	0.965	79.54	0.660	**0.976**	51.39	0.685	0.973
EleGANt [47]	54.06	0.634	0.973	86.19	0.651	0.961	61.40	0.693	0.969
SSAT [36]	38.01	0.645	0.975	70.53	0.667	0.973	53.84	0.692	0.976
LADN [11]	73.91	0.620	0.917	104.91	0.634	0.914	65.87	0.688	0.930
CPM [29]	42.76	0.652	0.951	95.61	0.661	0.924	40.57	0.729	0.954
Stable-Makeup [52]	33.26	0.682	0.973	64.64	0.711	0.968	37.33	0.767	0.965
SHMT-h_0	32.24	0.658	**0.976**	51.54	0.668	**0.976**	38.97	0.711	**0.978**
SHMT-h_4	**24.93**	**0.715**	0.953	**45.02**	**0.719**	0.954	**27.01**	**0.786**	0.958

Wechat

Our code: https://github.com/Snowfallingplum/SHMT

Thank you for watching