CVPR JUNE 18-22, 2023 VANCOUVER, CANADA

GP-VTON: Towards General Purpose Virtual Try-on via Collaborative Local-Flow Global-Parsing Learning

Zhenyu Xie[1], Zaiyu Huang[1], Xin Dong[2], Fuwei Zhao[1], Haoye Dong[3], XijinZhang[2], Feida Zhu[2], Xiaodan Liang[1,4*]

[1]Shenzhen Campus of Sun Yat-Sen University, [2]ByteDance, [3]Carnegie Mellon University, [4]Peng Cheng Laboratory

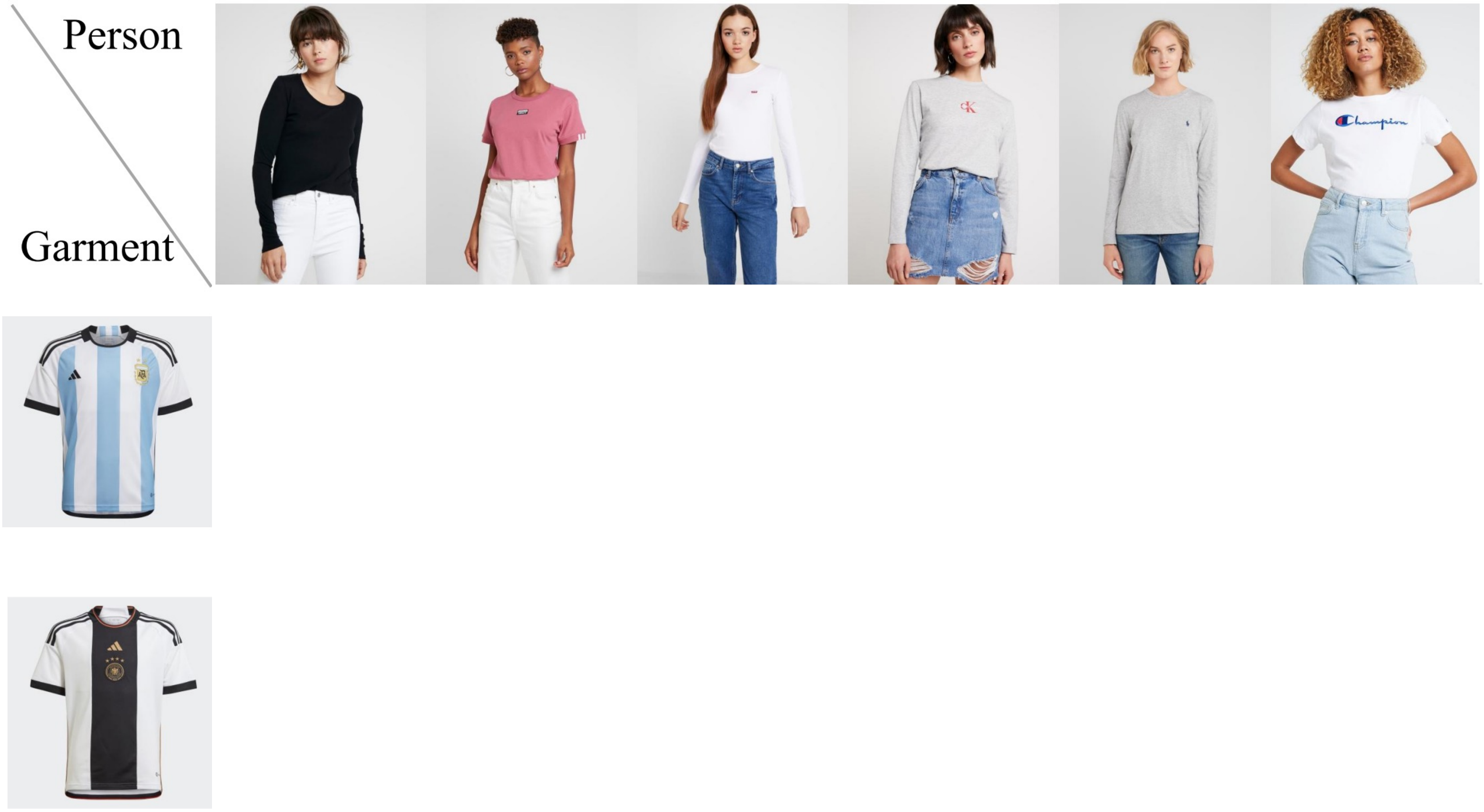
Paper Tag：THU-PM-281

Contributions

- Propose a unified framework, GP-VTON, for ***diverse virtual try-on scenarios***.
- Propose LFGP warping module for ***semantic-correct garment deformation***.
- Introduce DGT training strategy for ***distortion-free garment deformation***.
- Outperform existing SOTAs on two virtual try-on benchmarks.

Task Definition

- Given a **source person** and a **target garment**,



Task Definition

- Given a **source person** and a **target garment**, image-based virtual try-on aims to **transfer the garment onto the specific person**.

Motivation

➢ Problem of Existing Methods

- Fail to handle ***challenging inputs*** (e.g., intricate pose, hard garment)

Motivation

➢ Problem of Existing Methods

- Fail to handle ***challenging inputs*** (e.g., intricate pose, hard garment)
- Suffer from ***texture distortion*** in the warped result

Motivation

- Problem of Existing Methods
  - Fail to handle ***challenging inputs*** (e.g., intricate pose, hard garment)
  - Suffer from ***texture distortion*** in the warped result

- Motivation
  - Exploit local warping flow for ***semantic-correct*** deformation
  - Exploit global parsing to ***combine different local warped results***
  - Introduce a novel training strategy for ***distortion-free*** deformation

Method Overview

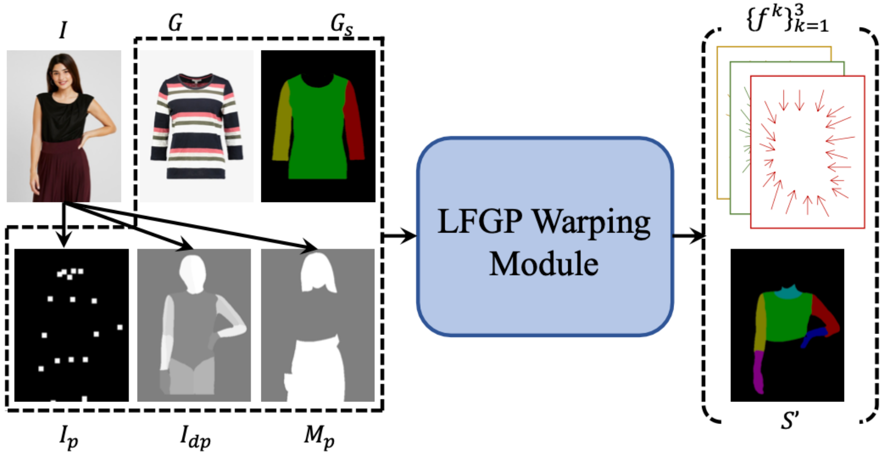
- Garment Warping Module
  - Local-Flow Global-Parsing (LFGP) Warping Module
  - Dynamic Gradient Truncation (DGT) Training Strategy
- Try-on Generator

Method Overview

➢ Garment Warping Module

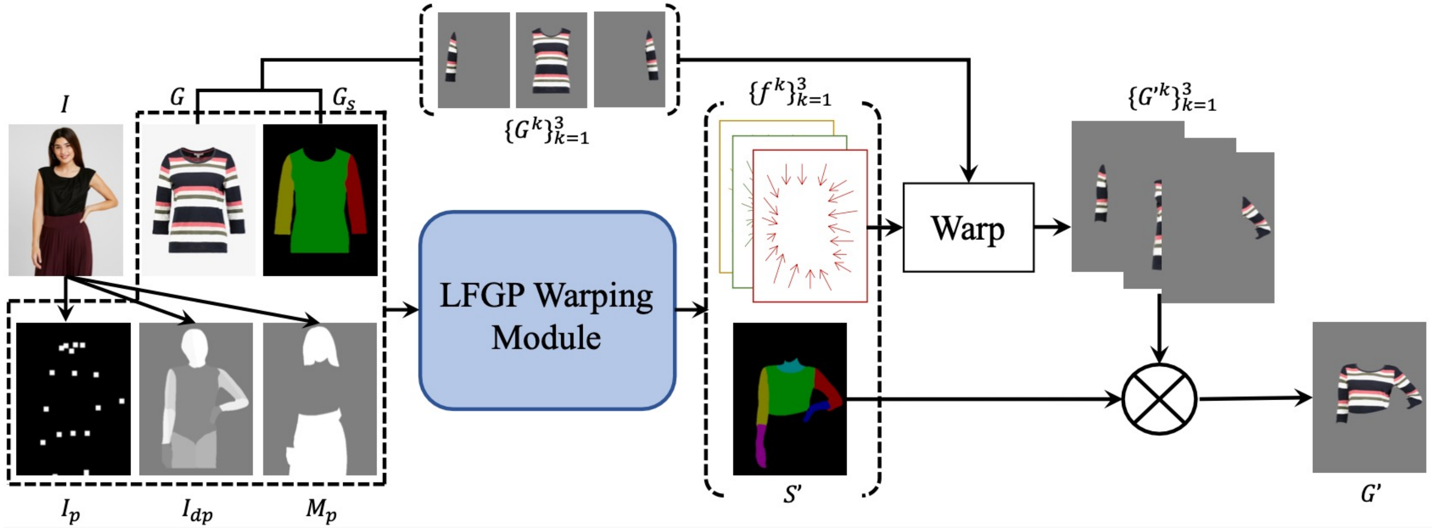
- Local-Flow Global-Parsing (LFGP) Warping Module
- Dynamic Gradient Truncation (DGT) Training Strategy

➢ Try-on Generator



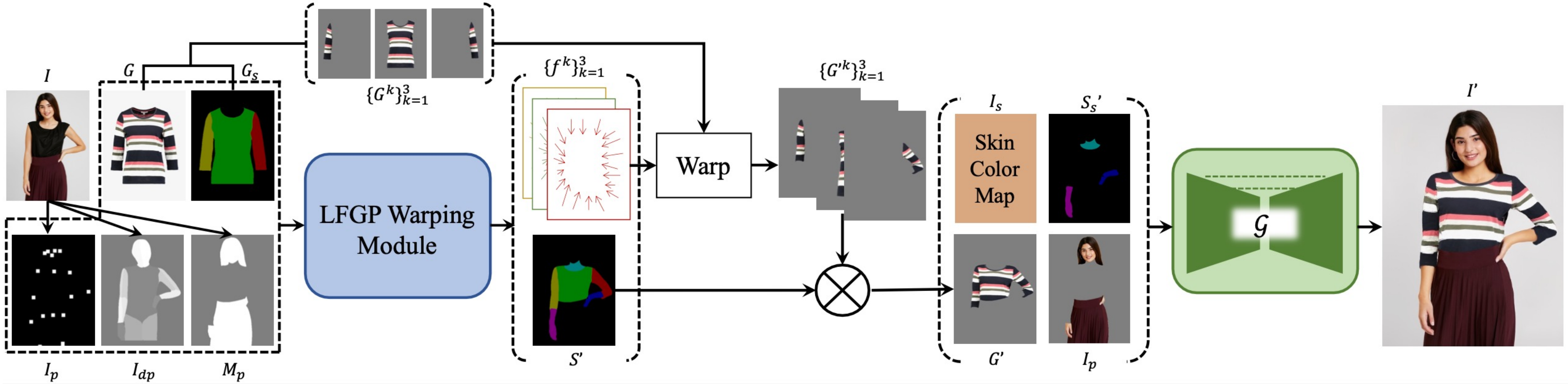
Method Overview

- Garment Warping Module
  - Local-Flow Global-Parsing (LFGP) Warping Module
  - Dynamic Gradient Truncation (DGT) Training Strategy
- Try-on Generator



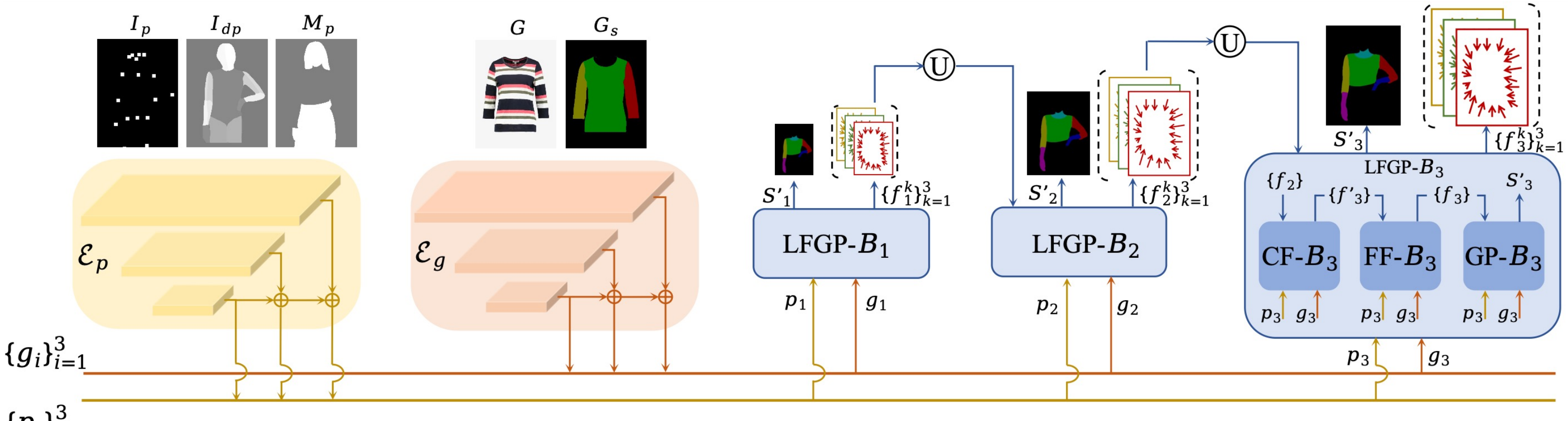
Method Overview

- Garment Warping Module
  - Local-Flow Global-Parsing (LFGP) Warping Module
  - Dynamic Gradient Truncation (DGT) Training Strategy
- Try-on Generator



LFGP Warping Module

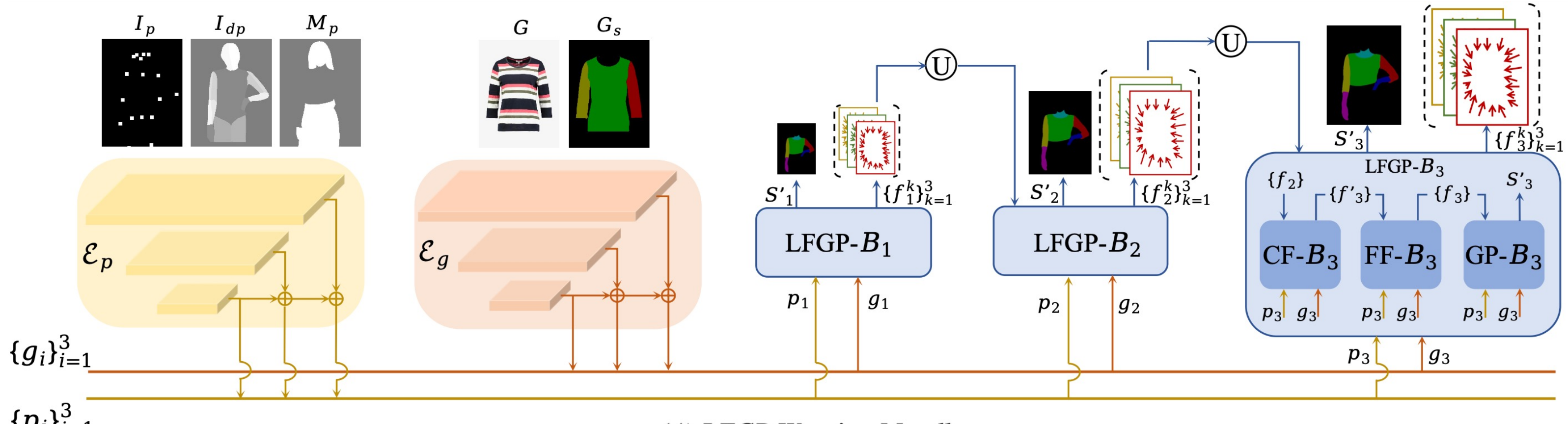
- LFGP Warping Module consists of a person encoder, a garment encoder and cascade local-flow global-parsing estimation blocks.



(A) LFGP Warping Moudle

LFGP Warping Module

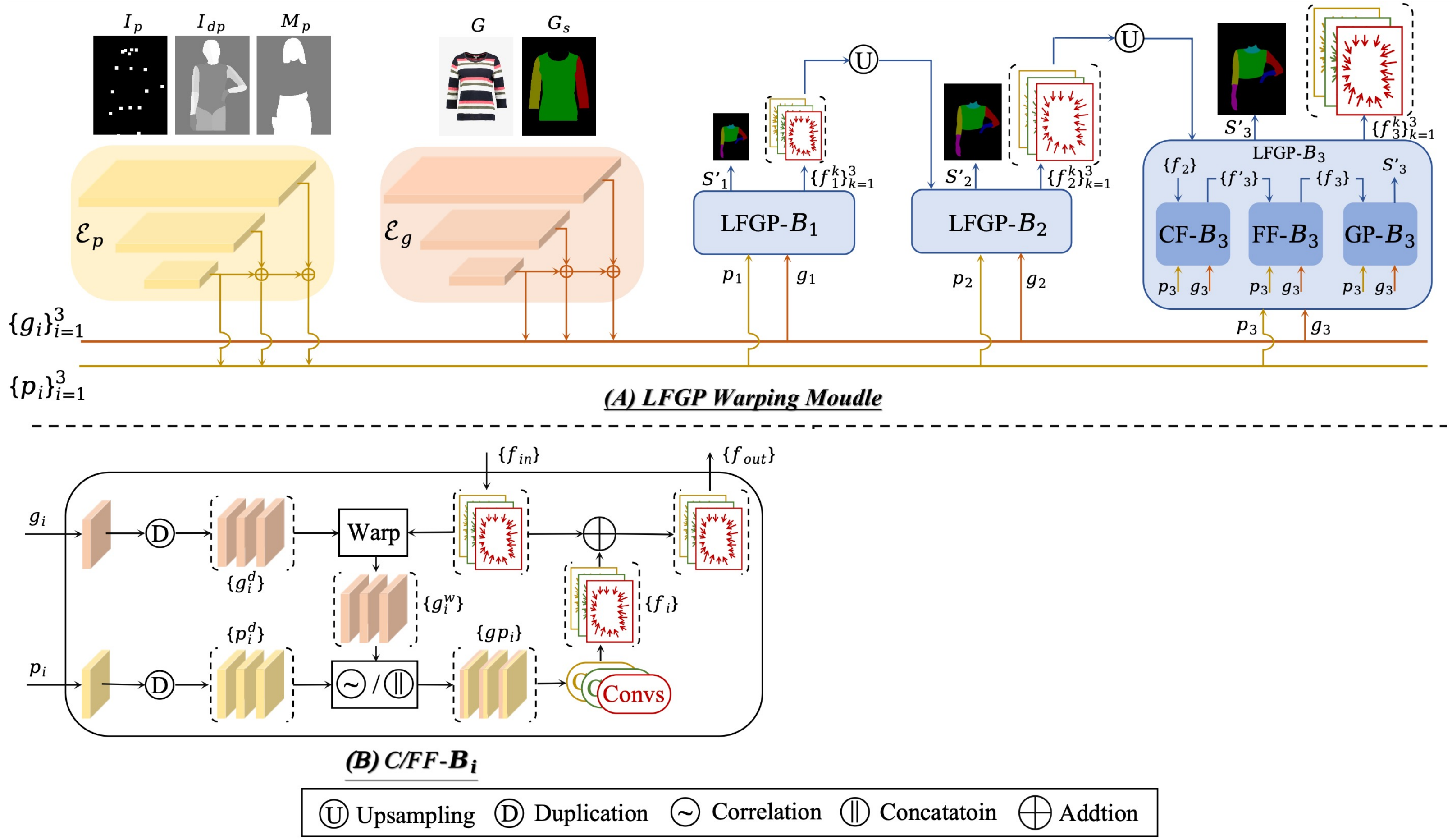
- Each LFGP estimation block consist of the two flow estimation blocks and one parsing estimation block.



(A) LFGP Warping Moudle

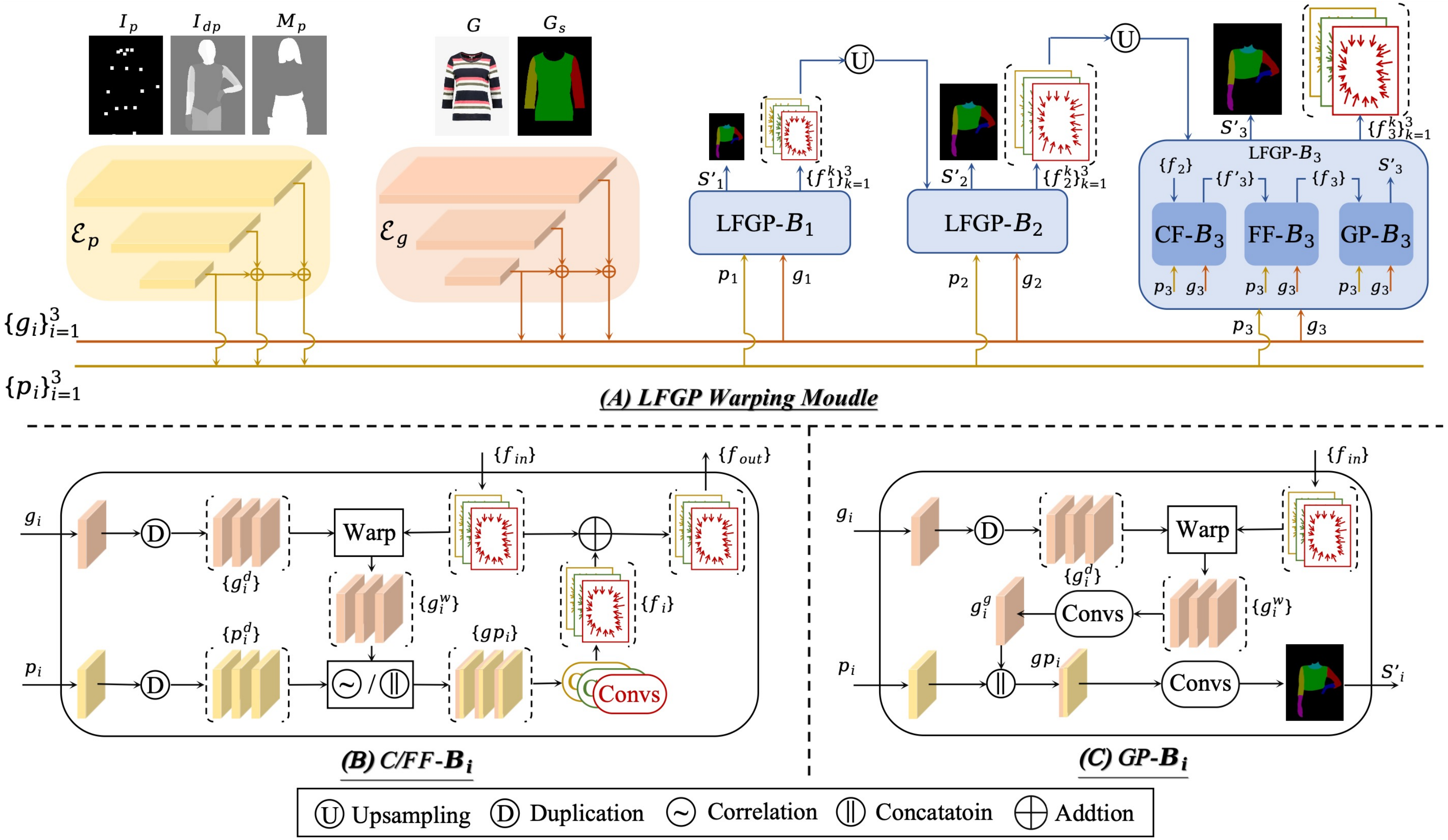
LFGP Warping Module

- The flow estimation block takes as inputs the person feature, garment feature and the previous estimated flow, then outputs the current estimated flow .



LFGP Warping Module

- The parsing estimation block takes as inputs the person feature, garment feature and the previous estimated flow, then outputs the estimated parsing.



DGT Training Strategy

- Out DGT training strategy **dynamically penalized the overlapping region** in warped garment according to the **wearing style** of the person image.

- **Red region:** the warped result will be penalized and the **gradient will be backpropagated**.
- **Green region:** the warped result will be neglected and the **gradient will be truncated**.

DGT Training Strategy

- Out DGT training strategy **dynamically penalized the overlapping region** in warped garment according to the **wearing style** of the person image.

- **Red region:** the warped result will be penalized and the **gradient will be backpropagated**.
- **Green region:** the warped result will be neglected and the **gradient will be truncated**.

$$R_{style} = \frac{R^{warped}}{R^{flat}} \quad (1)$$

$$R^{*} = \frac{H^{*}}{W^{*}}. \quad (2)$$

Experiment Results

- Qualitative Comparisons on VITON-HD dataset

Experiment Results

- Qualitative Comparisons on DressCode dataset

Experiment Results

- Quantitative Comparisons on VITON-HD and DressCode dataset

Method	SSIM ↑	FID ↓	LPIPS ↓	mIoU ↑	HE ↑
PF-AFN [12]	0.8858	9.475	0.0871	0.8412	14.9%
FS-VTON [19]	0.8829	9.552	0.0906	0.8357	8.80%
HR-VITON [29]	0.8623	16.21	0.1094	0.6949	9.10%
SDAFN [1]	0.8821	9.400	0.0922	0.5927	16.3%
GP-VTON (Ours)	**0.8939**	**9.197**	**0.0799**	**0.8764**	**50.9%**

Table 1. Quantitative comparisons on VITON-HD dataset [6]

Dataset	DressCode-Upper					DressCode-Lower					DressCode-Dresses				
Method	SSIM ↑	FID ↓	LPIPS ↓	mIoU ↑	HE ↑	SSIM ↑	FID ↓	LPIPS ↓	mIoU ↑	HE ↑	SSIM ↑	FID ↓	LPIPS ↓	mIoU ↑	HE ↑
PF-AFN [12]	0.9454	14.32	0.0380	0.8392	14.0%	0.9378	18.32	0.0445	0.9463	12.3%	0.8869	13.59	0.0758	0.8743	15.0%
FS-VTON [19]	0.9457	13.16	0.0376	0.8381	5.33%	0.9381	17.99	0.0438	0.9478	14.7%	**0.8876**	13.87	0.0745	0.8760	8.33%
HR-VITON [29]	0.9252	16.86	0.0635	0.6660	3.00%	0.9119	22.81	0.0811	0.8670	2.67%	0.8642	16.12	0.1132	0.7209	2.33%
SDAFN [1]	0.9379	12.61	0.0484	0.5046	11.3%	0.9317	**16.05**	0.0549	0.4543	13.3%	0.8776	**11.80**	0.0852	0.5945	19.3%
GP-VTON (Ours)	**0.9479**	**11.98**	**0.0359**	**0.8766**	**66.3%**	**0.9405**	16.07	**0.0420**	**0.9601**	**57.0%**	0.8866	12.26	**0.0729**	**0.8951**	**55.0%**

Table 2. Quantitative comparisons on DressCode dataset [32]

Experiment Results

- Ablation Study

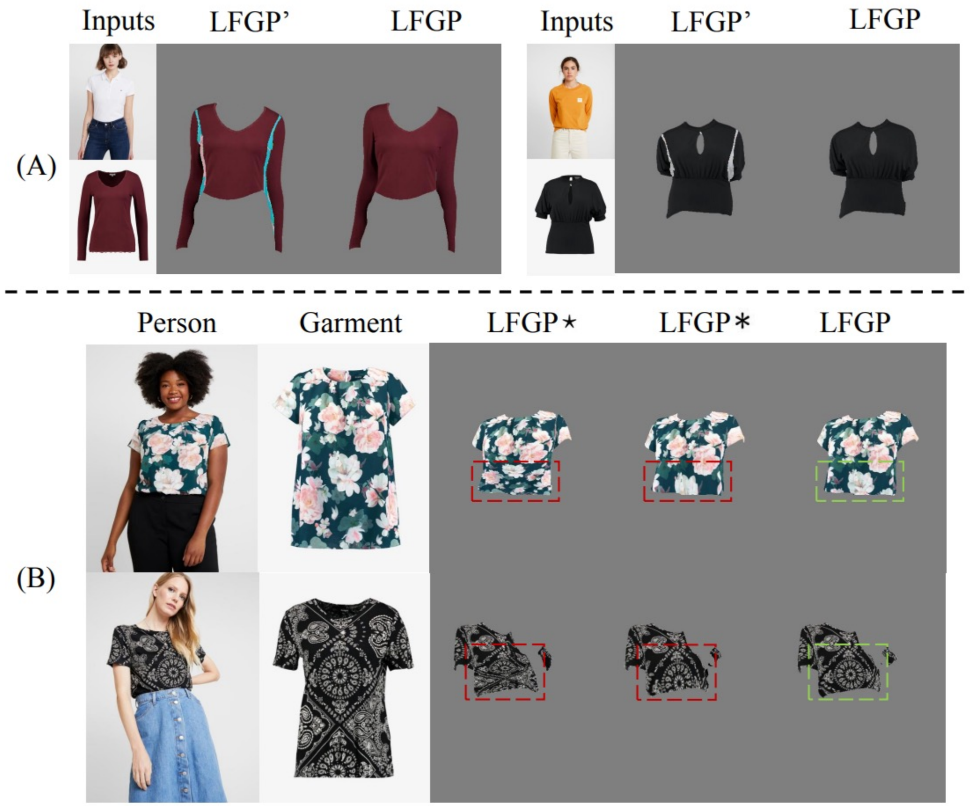


Figure 8. Ablation studies on the effectiveness of (A) the global parsing during the parts assembling process and (B) the dynamic gradient truncation training strategy.

Experiment Results

- Ablation Study

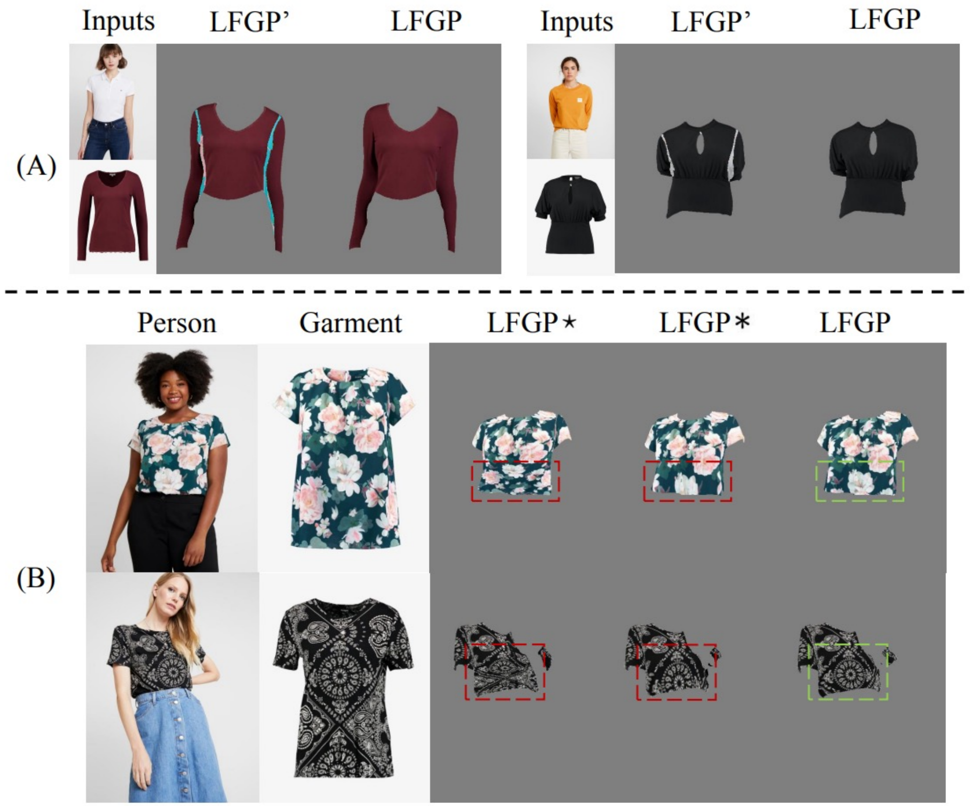


Figure 8. Ablation studies on the effectiveness of (A) the global parsing during the parts assembling process and (B) the dynamic gradient truncation training strategy.

Method	LF	GT	DGT	SSIM ↑	LPIPS ↓	mIoU ↑	R_{diff} ↓
LFGP †	✗	✗	✗	0.9016	0.0950	0.8412	0.3058
LFGP ⋆	✓	✗	✗	0.9039	0.0911	**0.8784**	0.3003
LFGP ∗	✓	✓	✗	**0.9053**	0.0900	0.8774	0.2409
LFGP	✓	✗	✓	0.9050	**0.0884**	0.8764	**0.1655**

Table 3. Ablation study of the Local FLow (LF), Gradient Truncation (GT), and Dynamic Gradient Truncation (DGT) on the VITON-HD dataset [6].

- Hight-width Difference Metric: $R_{diff} = |R^{warped} - R^{flat}|$

SUN YAT-SEN UNIVERSITY 中山大學

Carnegie Mellon University

鹏城实验室 PCL Peng Cheng Laboratory

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thanks for Listening

Paper

Code