

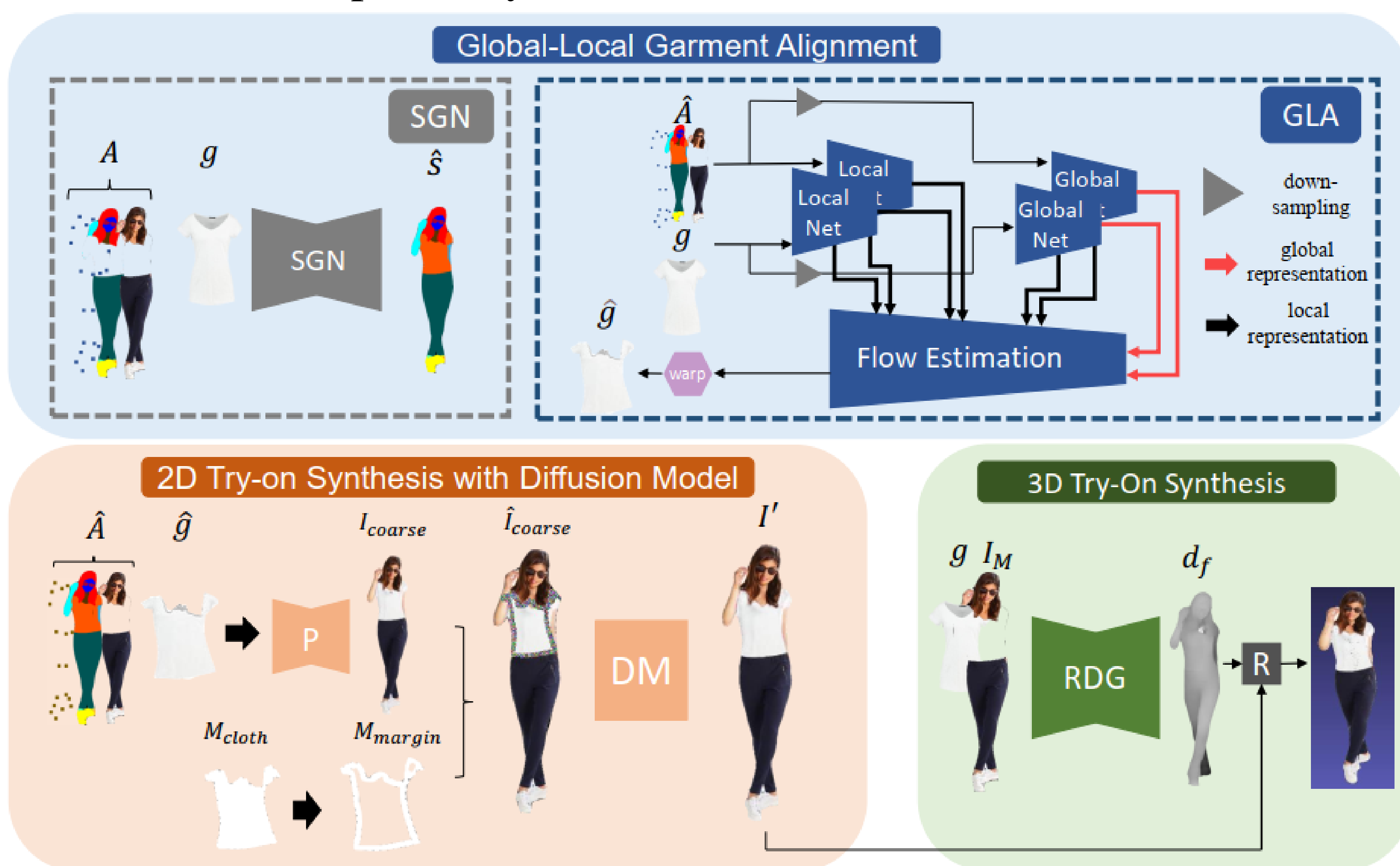
## Contribution

3D virtual try-on has recently received more attention due to its great practical and commercial value. However, there remains the problems that the garment cannot accurately correspond to a human body by geometric transformation and abnormal textures may be produced in the synthesis result. To address these issues, we propose a 3D virtual try-on method with global-local alignment and diffusion model. The main contributions of this paper are summarized as follows:

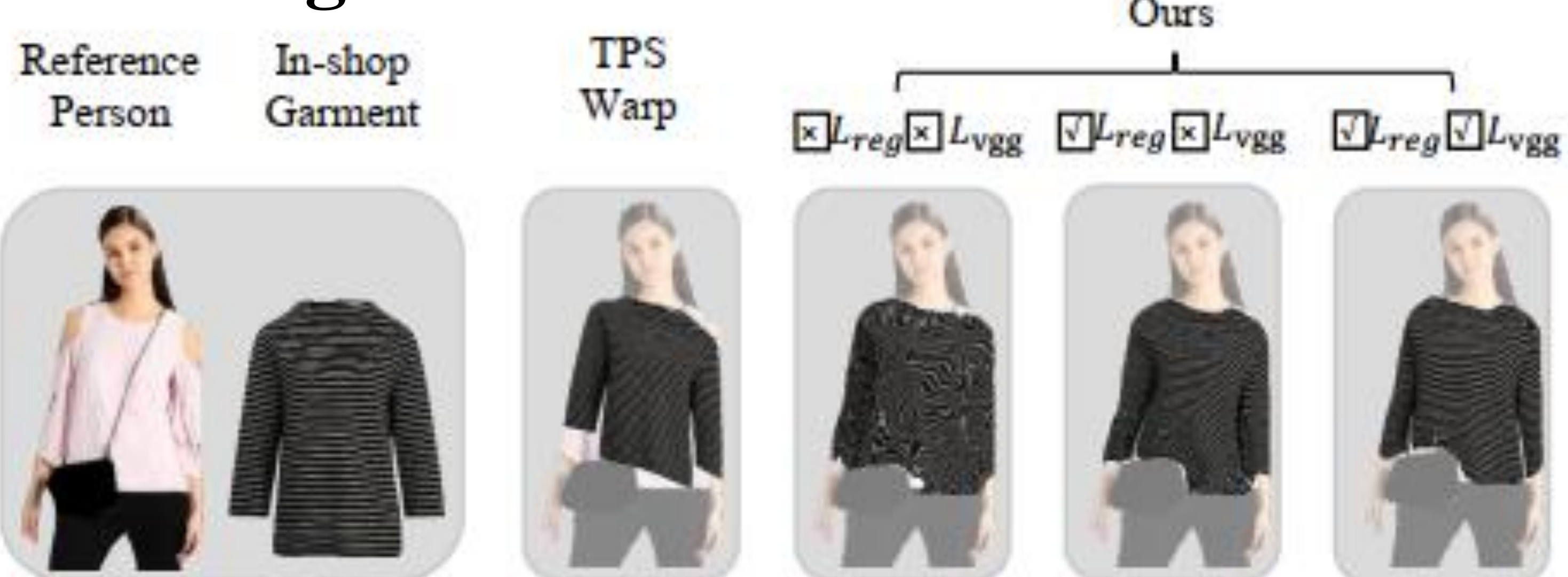
- 1) A 3D virtual try-on network framework through monocular 2D images is proposed to achieve more accurate 3D try-on mesh models;
- 2) A new garment alignment network is proposed to achieve superior warping results with global and local garment appearance flow;
- 3) An image synthesis strategy is proposed using a diffusion model based image generation network to enhance the quality of textures produced in edge regions.

## Overview of our framework

Overview of our framework. The SGN generates a semantic map; the GLA predicts pixel-wise flows for warping garment; the part of 2D synthesis and 3D synthesis generate the try-on result on image and mesh, respectively, where R denotes mesh reconstruction.



## Qualitative evaluation of our global-local alignment



## Challenges in image-based virtual try-on

The deformation of clothes may lead to misalignment and cannot cover the textures of the clothing's back collar.



## Results

Quantitative evaluation on 2D try-on results.

Method	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$
CP-VTON	0.8503	-	-	20.05
ACGPN	0.8924	-	-	20.19
M3D-VTON	0.9373	24.94	0.0395	16.48
CF-VTON	0.7940	20.50	0.1890	15.37
Ours(w/o DM)	0.9501	25.43	0.0374	14.86
Ours(o/ DM)	<b>0.9511</b>	<b>25.49</b>	<b>0.0370</b>	<b>14.81</b>

Quantitative evaluation on 2D try-on results.

Method	Abs. $\downarrow$	Sq. $\downarrow$	RMSE $\downarrow$
NormalGAN	15.41	0.778	18.94
PIFu	8.376	1.813	27.57
M3D-VTON	7.880	0.385	11.27
Ours(w/o DM)	7.839	0.364	10.91
Ours(o/ DM)	<b>7.831</b>	<b>0.358</b>	<b>10.85</b>

A visual comparison of the results between our method and the SOTA methods. The 3D reconstruction meshes are in the blue dashed boxes, and the rest are 2D try-on results. The red dashed boxes represent the defects in the corresponding methods.



## Acknowledgement

This work is supported by the National Natural Science Foundation of China (61828501), and the science and technology project fundings of State Grid Jiangsu Electric Power Co., Ltd. (J2023031).