# VGFlow: Visibility guided Flow Network for Human Reposing Supplementary

Rishabh Jain*
MDSR Adobe

Krishna Kumar Singh
Adobe Research

Mayur Hemani
MDSR Adobe

Jingwan Lu
Adobe Research

Mausooom Sarkar
MDSR Adobe

Duygu Ceylan
Adobe Research

Balaji Krishnamurthy
MDSR Adobe

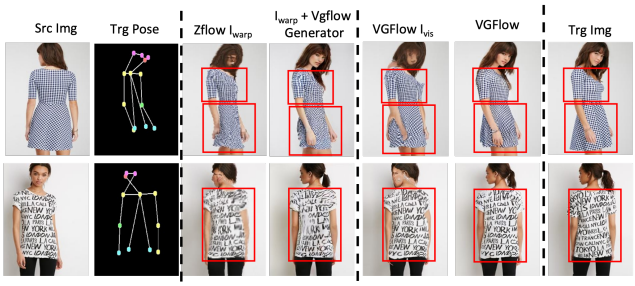


Figure 1. Flow estimation improvement via VGFLow

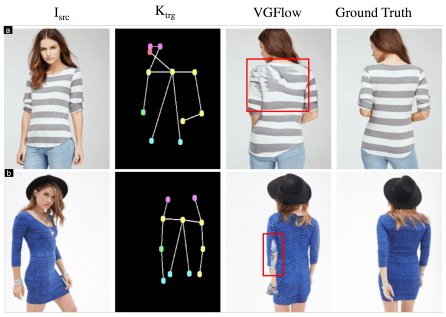


Figure 2. Failure cases of VGFlow due to warping limitations(a) and incorrect target pose segmentation(b)

**Flow ablation** As mentioned in the main paper, we replace our flow network with Zflow [2] architecture, which includes a flow pyramid with GA(Gated Aggregation) to predict a single 2D displacement flow field. Such architectures fail to generate accurate warping results when a significant portion of the final image doesn't have any one-to-one correspondence with input image and a single flow field is insufficient to model both visible and invisible regions. This model produced sub-optimal flow field and lacked the ability to preserve high-frequency components like text & design. To measure the direct effect of our contribution on the flow estimate, we replace our texture encoder's input with Zflow's single warped image(Zflow $I_{warp}$) and finetune our VGFlow generator. The quantiative metrics (SSIM/LPIPS/FID) deteriorate to (0.719/0.205/11.70) against VGFlow's (0.726/0.185/9.29) and we show the qualitative comparison in Fig 1.

**Failures & Limitations** Even though our network is able to produce accurate reposed outputs in majority of the cases, there are still limitations. In Fig 2, we see network artifacts that were caused due to incorrect warping (a) and erroneous target segmentation in VGFlow (b). Some of these difficulties can be alleviated in future by enforcing symmetry [1] or by leveraging target segmentation mask [3, 5].

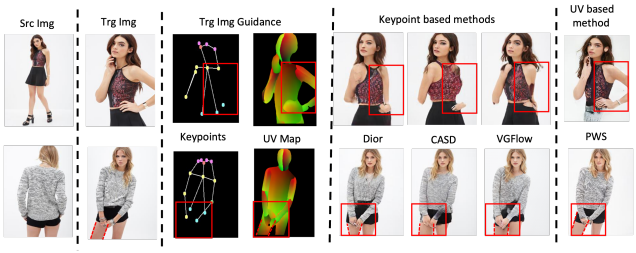


Figure 3. UV map extracts per pixel 3D human body mapping from the target image as guidance. Note missing keypoint for hand in row 1 and leg angle information in row 2

**Comparison with UV based methods** UV based reposing methods like PWS [1] takes additional information in the form of ground truth UV map of the target image as input. UV maps encode per pixel mapping between source and target image, and hence contain a lot more information than keypoints. We also show qualitative effect of this phe-

*rishabhj@adobe.com