### Motivation

- Existing virtual try-on methods warp clothing item to fit human body and fuse warped clothing with human for synthesis.
- Loss functions are not perceptually motivated
- No robustness test
- Does not leverage context

### Failure Cases

- Complex **poses** and **shapes** of person
- Synthesize **skin color**
- **Occluded** body parts (e.g. long sleeve clothing)
- **Logo**, **texture** and **embroidery** of clothing.

These failure cases lead to blurry, unrealistic outputs as well as artifacts.

### Types of Pose

| Easy | Medium | Hard |
|------|--------|------|
| ![Easy Pose] | ![Medium Pose] | ![Hard Pose] |
Fill in Fabrics: Body-Aware Self-Supervised Inpainting for Image-Based Virtual Try-On

Hasib Zunair, Yan Gobeil, Samuel Mercier, A. Ben Hamza

- person body part mask, cloth region
- warp target clothing

- 2D Pose Estimator
- 2D Human Parser

- Mask Compositing

- STN
- Warper

- Fabricator

- MCM objective

- Multi-scale structural constraint (MSC)

\[ \mathcal{L}_{\text{refined}} = \beta_1 \mathcal{L}_{\text{CGAN}} + \beta_2 \mathcal{L}_{\text{VGG}} + \beta_3 \mathcal{L}_{\text{MS-SSIM}} \]
Handles complex poses, retains logo, texture, embroidery, structure of clothing

| Method        | SSIM (↑) | FID (↓) |
|---------------|----------|---------|
| CA-GAN [15]   | 0.740    | -       |
| VITON [10]    | 0.783    | 0.787   |
| CP-VTON [28]  | 0.745    | 0.753   |
| VTNFP [31]    | 0.803    | 0.810   |
| ClothFlow [9] | 0.843    | -       |
| CP-VTON+ [20] | 0.750    | -       |
| SieveNet [14] | 0.837    | -       |
| ACGPN [30]    | 0.845    | 0.854   |
| DCTON [8]     | 0.830    | -       |
| CIT [23]      | 0.827    | -       |
| FIFA (Ours)   | 0.886    | 0.890   |

MCM captures structure of clothing

MSC aptures global context of clothing and shape of person

FIFA is robust to in-the-wild images for virtual try-on