Enhancing Person Synthesis in Complex Scenes via Intrinsic and Contextual Structure Modeling

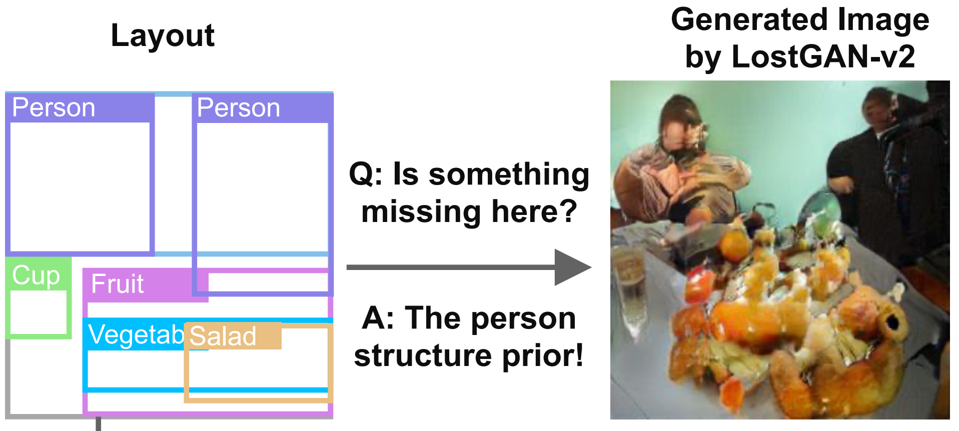
BMVC 2022

Xi Tian[1], Yong-liang Yang[1] , Qi Wu[2]
1 University of Bath; 2 University of Adelaide
xt275@bath.ac.uk, strongyang@gmail.com, qi.wu01@adelaide.edu.au

Purpose

Generating persons in complex scenes is difficult:

- Persons are more **articulated** compared with other objects.
- Existing methods faild, due to lacking of person **structure prior**.
- The person structrues are **intrinsic**, should not be affected by the complex context — scenes and other objects.



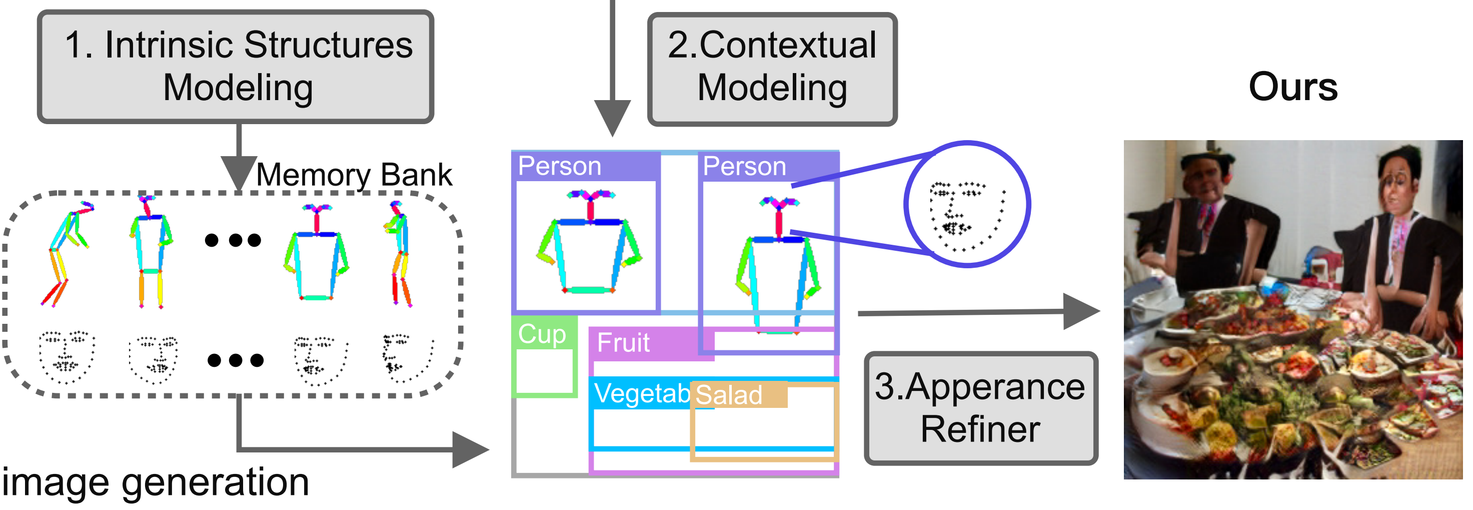
Q: Is something missing here?

A: The person structure prior!

Method

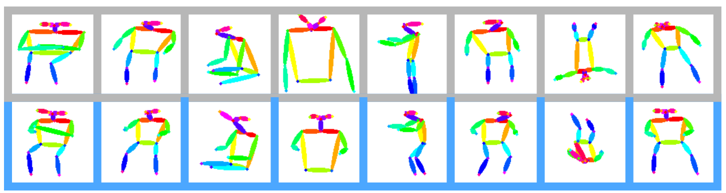
Overview

1. Build the person intrinsict structures: body keypoints & face keypoints
2. Infer the person structures from the context — relation with other objects;
3. Refine the persons together with the image generation



1) Intrinsic structure modelling

Keypoints VAE
- Based on Vector Quantized (VQ) VAE
- Enccding both body/face keypoints

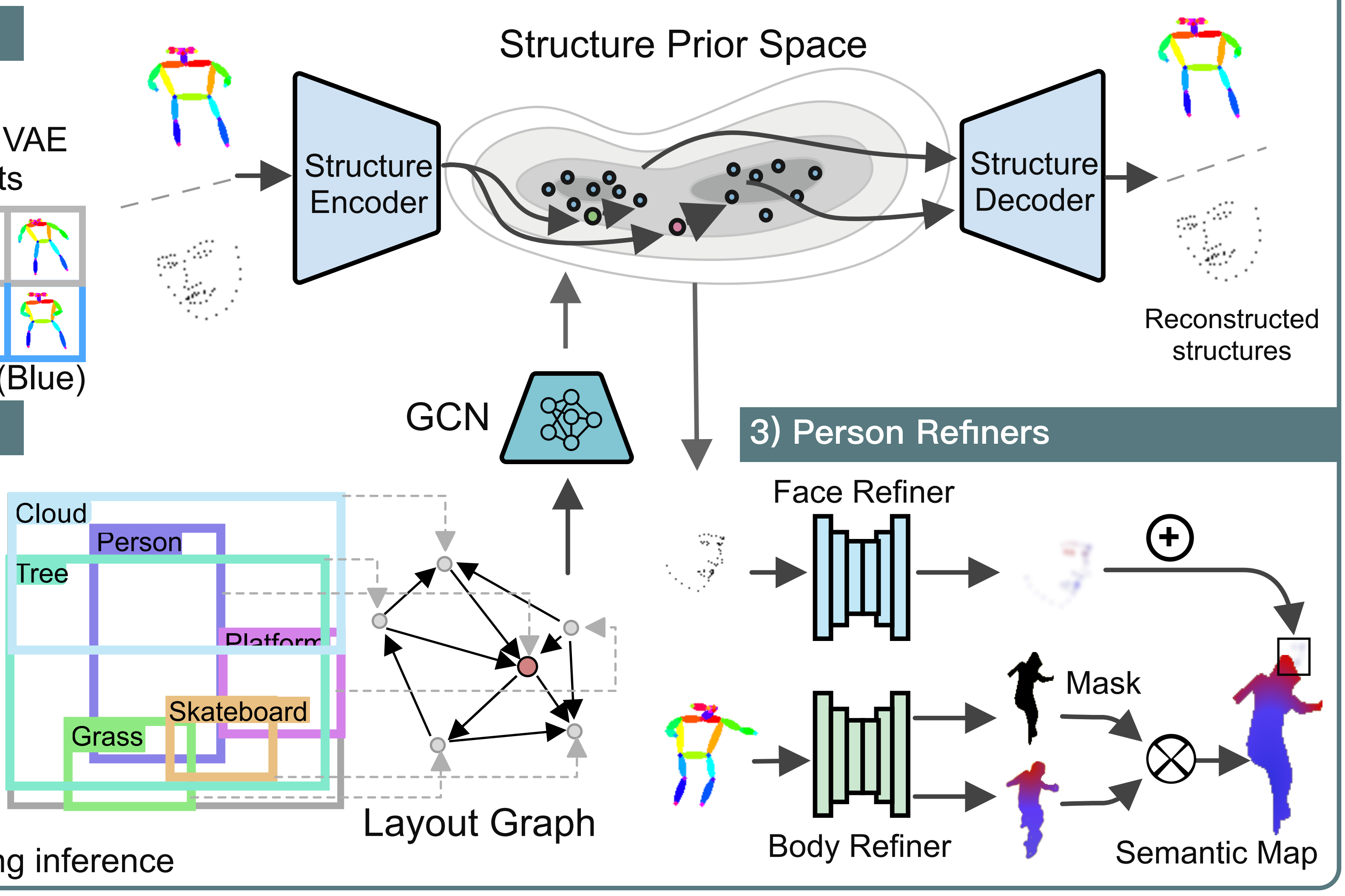


Learnt representational structrues (Blue)

2) Contextual Inference

Layout Graph Representation
- Nodes features:
  - Objects labels
  - Positions
  - Sizes
- Edge featues:
  - Positional Relations

Layout GCN
- Encoding the Person node contextual features
- Predicting person structures during inference



3) Person Refiners

Results

1. Qulitative Results
- Better Person Quality
- Better Crowd generation (row 2)
- Reasonable person structures (Last column)

2. Quantitative Results
- Higher Person Accuracy
- Higher Face Acc
- Higher Face Detection IOU