**Text-Pose-to-Image Model**

We implemented KPE into DALL-E[2] architecture to generate image with text and pose conditions.

**People Count Error (PCE)**

Errors in generated people image often manifest as having missing or additional body parts. PCE makes use of rich human anatomy knowledge embodied in pose estimation models e.g. OpenPose [3] to detect such error. Let’s say we tell the model to generate a person but it ends having 3 arms. OpenPose knows a person can only have 2 arms and allocate the third arm to additional person. As a result, OpenPose think there are 2 people in the image instead of the supposedly one. This discrepancy in people count flags images errors.

Our main contributions:
1. **Keypoint Pose Encoding (KPE)** to use the low-dimensional keypoint directly. This results in over 10x more memory efficiency and it is over 73% faster to train and inference!
2. **People Count Error (PCE)** a novel method to detect people image error.