Phase 5 Capstone Notes

10 December 2024

Pose Estimation and Detection Study Guide

What is Pose Estimation?

- **Pose estimation** uses computer vision techniques to identify and track the movement of a person or object in real-time.
- It involves predicting different poses based on an object's body parts and joint positioning in an image or video. For example, in a squat exercise, pose estimation can detect the positions of joints, arms, hips, and spine.
- It has applications in sports biomechanics, animation, gaming, robotics, medical rehabilitation, and surveillance.

Types of Pose Estimation:

- **Human pose estimation** focuses on detecting human body parts and joint positioning.
- **Rigid pose estimation** works with rigid objects, as opposed to flexible human bodies.
- **2D pose estimation** predicts key points based on pixel values in an image.
- **3D pose estimation** predicts the spatial position of a person or object.
- **Single pose estimation** detects a single person or object.
- **Multi-pose estimation** tracks multiple people or objects.

Pose Estimation Models:

- **Kinematic models** work for both 2D and 3D pose estimation, focusing on joint and limb positions.

- **Planar models** use rectangles to represent human body parts for 2D pose estimation.
- **Volumetric models** use deep learning to derive human body mesh for 3D pose estimation.

Popular Pose Estimation Models:

- **OpenPose**: A real-time model that detects key points of multiple people, including hands, face, and feet.
- **MoveNet**: A lightweight and accurate model from Google that detects 17 key points, with two versions for different speed and accuracy needs.
- **PoseNet**: A real-time model for both single and multi-person pose detection, detecting 17 key points and providing skeletal information.
- **DCPose**: Detects human pose from multiple frames, addressing challenges like motion blur and occlusions.
- **DensePose**: Maps human pixels from an RGB image to the 3D surface of a human body.
- **HigherHRNet**: Addresses challenges in detecting poses of shorter people using feature pyramids to learn from scale-aware representations.
- **Lightweight OpenPose**: An optimized version of OpenPose for real-time inference.
- **AlphaPose**: A multi-person pose estimation model that can track poses across different frames.
- **TransPose**: Uses a transformer encoder to capture long-range spatial relationships between key points.

MediaPipe Solutions:

- **MediaPipe** offers a suite of solutions for applying AI and ML, including **pose landmark detection**.
- It includes:
    - **MediaPipe Tasks**: APIs and libraries for deploying solutions.
    - **MediaPipe Models**: Pre-trained models for each solution.
    - **MediaPipe Model Maker**: For customizing models.
    - **MediaPipe Studio**: For visualizing and evaluating solutions.

- **Pose landmark detection** in MediaPipe identifies key body locations and can be used to analyze posture and categorize movements.
- It outputs body pose landmarks in 2D image coordinates and 3D world coordinates.
- **Python** code examples are available for using MediaPipe pose landmark detection.
- **Configuration options** allow customization of the model's behavior, such as the minimum confidence score for detection and whether to output segmentation masks.

Best Pose Estimation Model for Virtual Try-On:

The sources don't explicitly name the best pose estimation model for virtual try-on. However, they provide insights to inform the selection:

- **Lightweight and accurate models** like **MoveNet** or **Lightweight OpenPose** are suitable for real-time AR applications.
- **3D pose estimation models** like **DensePose** would be necessary to accurately map clothing onto the user's body in 3D space.
- **MediaPipe Pose Landmarker** provides a solution specifically designed for pose landmark detection and offers flexibility through configuration options.

It is important to note that selecting the best model depends on the specific requirements of the virtual try-on application, including accuracy needs, latency requirements, and the complexity of the clothing or accessories being tried on.

11 December 2024

Key Points from the VITON-HD Research Paper

- **Purpose:** The research paper introduces **VITON-HD, a new method for generating high-resolution (1024x768) virtual try-on images.** This technology aims to enhance the online shopping experience by allowing customers to visualize how clothes would look on them.
- **Challenges in Virtual Try-On:** The paper highlights the challenges in creating realistic virtual try-on images, including preserving a person's characteristics,

naturally deforming clothes, maintaining clothing details, and rendering occluded body parts. It specifically addresses the issue of **misalignment artifacts** that become more pronounced at higher resolutions.

- **VITON-HD's Approach:** To overcome these challenges, VITON-HD employs several key techniques:
    - **Clothing-Agnostic Person Representation:** This representation utilizes pose and segmentation maps to remove information about the original clothing from the input image, allowing for a clean replacement with the target clothing item.
    - **ALIAS Normalization:** This normalization technique removes misleading information from misaligned areas, which are regions where the warped clothing does not perfectly match the desired clothing regions on the person. It also helps preserve semantic information throughout the network.
    - **ALIAS Generator:** The ALIAS generator uses ALIAS normalization and a multi-scale refinement process to accurately fill misaligned areas with clothing texture and maintain clothing details.
- **Advantages of VITON-HD:** The paper demonstrates the superiority of VITON-HD over existing methods through qualitative and quantitative evaluations, including comparisons with CP-VTON and ACGPN.
    - **High-Quality Images:** VITON-HD generates visually convincing images at 1024x768 resolution with better detail preservation and more natural body shapes compared to the baselines.
    - **Effective Misalignment Handling:** It effectively handles misalignment, showing better performance as the degree of misalignment increases.
    - **Positive User Feedback:** A user study confirms the effectiveness of VITON-HD in generating realistic virtual try-on images, with users consistently giving it high ratings for photorealism and clothing detail preservation.

Using the VITON-HD Dataset

Unfortunately, the sources provided do not contain specific details on the structure of the VITON-HD dataset or instructions on how to use the test pairs. The information provided

focuses on the technical aspects of the virtual try-on method and its evaluation, rather than dataset usage.

The following preprocessing techniques were used to create the clothing-agnostic person representation:

● A segmentation map (S) and a pose map (P) were predicted for the reference image (I) using pre-trained networks.

● The segmentation map (S) was used to remove the clothing region to be replaced.

● The pose map (P) was used to remove the arms, but preserve the hands, as they are difficult to reproduce.

● Based on S and P, a clothing-agnostic image (Ia) and a clothing-agnostic segmentation map (Sa) were created.

This clothing-agnostic person representation allowed the model to completely remove the original clothing information while preserving information about the person's pose, body shape, and identity

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

- At vero eos et accusam et justo duo dolores et ea rebum
    - Ut wisi enim ad minim veniam.
    - Quis nostrud exerci tation ullamcorper.
    - Suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

Lorem ipsum dolor

- At vero eos et accusam et justo duo dolores et ea rebum
    - Ut wisi enim ad minim veniam.
    - Quis nostrud exerci tation ullamcorper.