

Indice

Link code : <https://github.com/Gengzigang/PCT>

Basado en MMPose

Autores: Zigang Geng, Chunyu Wang, Yixuan Wei, Ze Liu, Houqiang Li, Han Hu

Año: March 21, 2023

Pages: 13

"Human body exhibits inherent compositional structure - it's not just a collection of independent points"

Proceso

- Dos modelos separados pero entrenados secuencialmente
 - Tokenización convierte espacio continuo -> discreto
 - Diffusion actúa en el espacio de tokens, no píxeles
 - Conditioning permite guiar la generación con la imagen de entrada
-

Dataset

- Coco : 149k personas
 - MPII : 40k personas
 - OCHuman : Especificas para oclusiones severas
 - CrowdPose : Escenas muy concurridas
-

Parámetros

- Default data augmentations provided by MMPose
 - Random scale: (0.5, 1.5)
 - Random rotation: (-40° , 40°)
 - Random flip: (50%)
 - Grid dropout and coord jitter: (h=0.2, s=0.4, c=0.4, b=0.4)
- Body augmentation for COCO

- Image size : 256 x 256
- Optimizer: AdamW
- Learning rate: $1e - 2$
- Weight decay: 0.5

Métricas de Evaluación

- AP (Average Precision) - Métrica principal en COCO
- AR (Average Recall)
- AP_m (Medium), AP_i (Large) - Para diferentes tamaños de personas
- $PCK@0.5$ - Porcentaje de keypoints correctos

Figuras

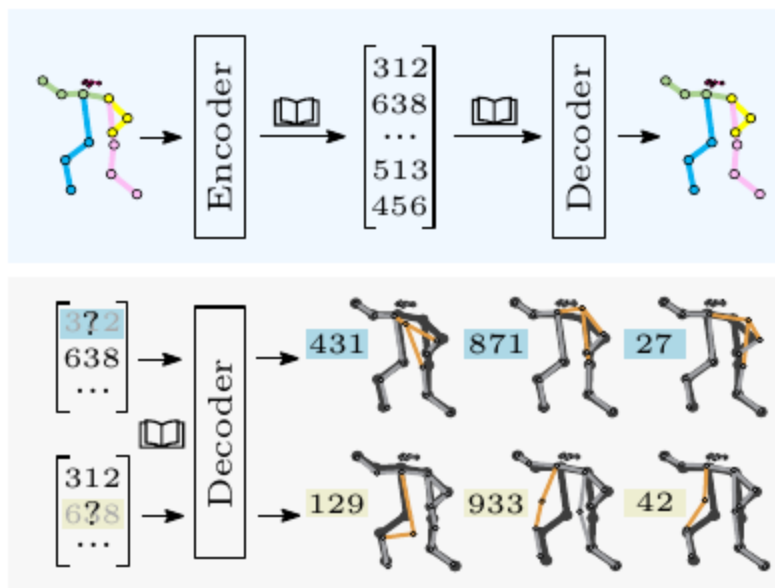


Figura 1: Representa una pose mediante M tokens discretos, que son índices de las entradas del libro de códigos (arriba). Cada token aprende a representar una subestructura. En cada fila, mostramos que si cambiamos el estado de un token a diferentes valores, cambia constantemente la misma subestructura, resaltada en naranja. Las poses en negro son antes del cambio (abajo).

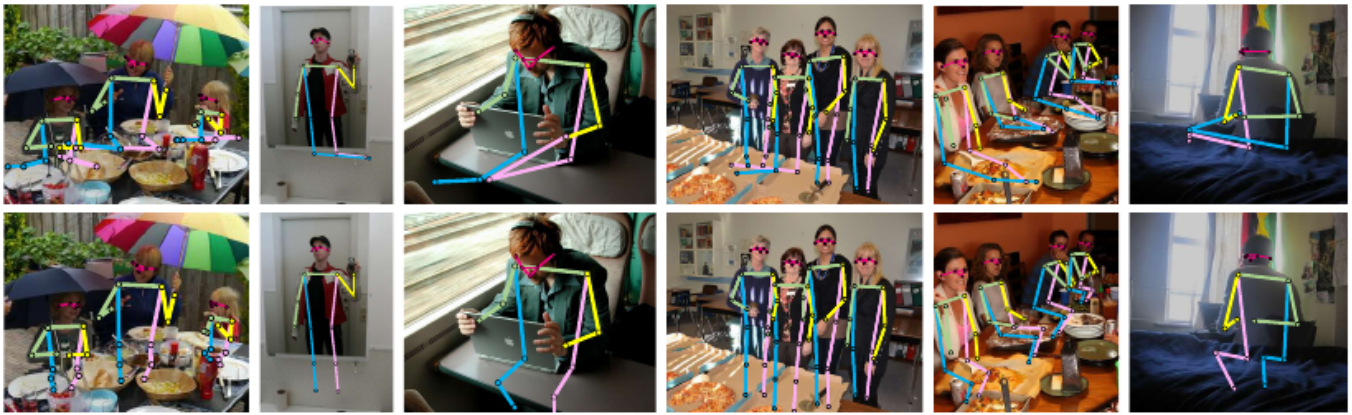


Figura 2. Método basado en mapas de calor (arriba) vs. método PCT (abajo) en escenas ocluidas. El método PCT predice poses razonables incluso en condiciones de oclusión severa. Las imágenes son de COCO val2017.

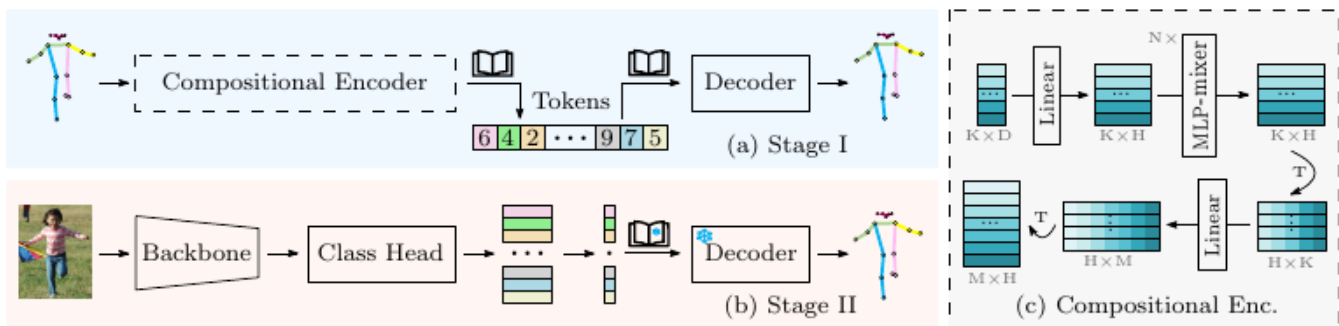


Figura 3. Dos etapas de la representación PCT (a,b) y la estructura del codificador composicional (c). En la Etapa I, aprendemos un codificador composicional para transformar una pose en M tokens, cuantificados por un libro de códigos. Por lo tanto, una pose se representa mediante un conjunto de índices discretos del libro de códigos. En la Etapa II, planteamos la estimación de la pose como una tarea de clasificación mediante la predicción de las categorías de los M tokens, es decir, los índices de las entradas del libro de códigos. Estos serán decodificados por una red de decodificadores para obtener la pose final

ee p e
e ee e ee