

Indice

Link code :

Autores: Miroslav Purkrabek and Jiri Matas

Año: Agost 1, 2025

Pages: 18

"Human pose estimation methods work well on isolated people but struggle with multiple-bodies-in-proximity scenarios"

Proceso

Imagen I

↓

Detección → Enmascara instancias → MaskPose → Keypoints → SAM2

1

Dataset

- 2D human pose
 - COCO
 - MPII
 - AIC
 - Multibody problems such as occlusion and self-occlusion
 - OCHuman
 - CrowdPose

Parámetros

1. Detector: RTMDet-L

- **Entrenamiento: Fine-tuning**

- **Épocas:** 10
- **Conjunto de datos:** COCO-human
- **Aumento de datos:** "instance-removal augmentation"
- **Comparativa:** Se usó el mismo detector en experimentos top-down para una comparación justa.

2. Estimador de Postura: MaskPose (basado en ViTPose)

- **Entrenamiento:**
 - **Épocas:** 210
 - **Conjuntos de datos:** COCO, AIC y MPII
- **Generación de Máscaras:**
 - **Problema:** MPII y AIC carecen de segmentación de *ground truth*.
 - **Solución:** Se generan máscaras de *pseudo ground truth* usando SAM2.
 - **Prompt para SAM2:** Cajas delimitadoras (*bounding boxes*) y puntos clave visibles (*keypoints*) del *ground truth*.

3. Modelo de Segmentación: Segment Anything Model (SAM2)

- **Versión:** sam2-hiera-base+ ("as is")
 - **Configuración de Post-procesamiento:**
 - **Área máxima de hueco (max hole area):** 10
 - **Área máxima de partícula (max sprinkle area):** 50
 - **Procesamiento:** Cada instancia se procesa de forma independiente.
-

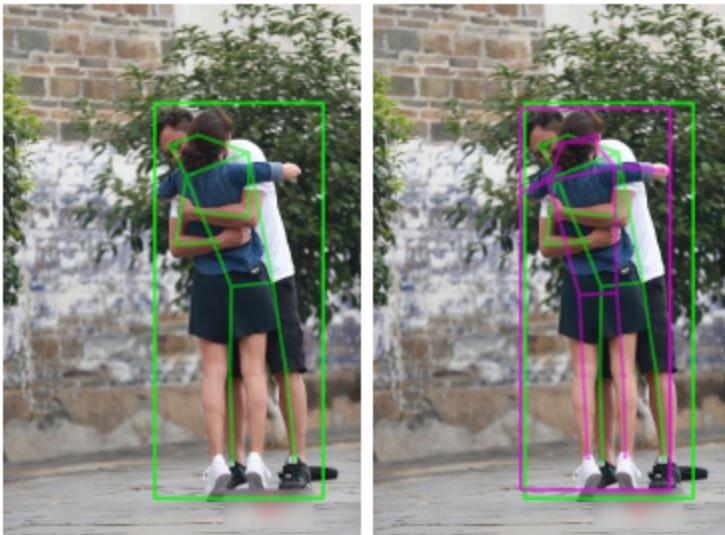
Métricas de Evaluación

- PCKh
 - is used for MPII
 - OKS
 - For COCO and related datasets
-

Figuras



Figura 1. El método BBox-Máscara-Pose (BMP). Los pasos (A) a (D) se repiten hasta que no se detecten nuevas detecciones en el paso (A). En este caso, el jugador de fondo no se detecta en el paso (A1). BMP se ajusta correctamente a la pose del jugador de primer plano (B1), lo que conlleva la corrección de su segmentación y bbox (C1). Tras enmascarar al jugador de primer plano (D1), se detecta al jugador de fondo (A2), se segmenta correctamente su cuerpo y se estima su pose. Derecha: salida de BMP. Nota: el bucle puede comenzar con un cuadro delimitador (A), una pose (B) o una máscara de segmentación (C).



(a) Una instancia omitida que se detecta en la segunda iteración de BMP. Izquierdo: TMDet + MaskPose, derecho: BMP.



(b) Dos casos en una detección se resuelven refinando las máscaras de segmentación con SAM, según la pose detectada. Izquierda: RTMDet, derecha: BMP. Obsérvese que la detección de la mujer ha mejorado, pero su pierna derecha sigue estando mal segmentada.

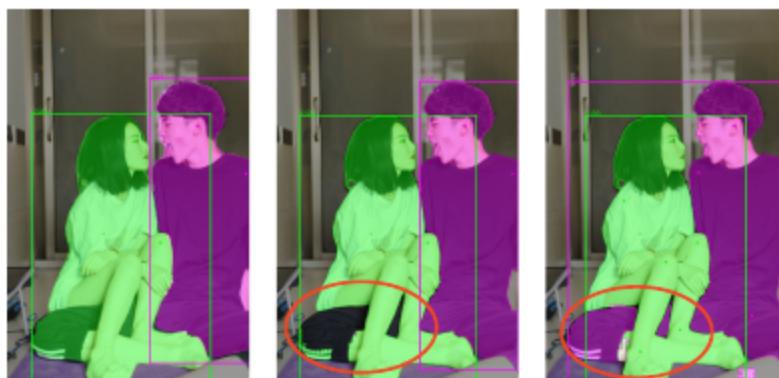


(c) Colapso de las estimaciones de pose para dos instancias con bboxes superpuestos detectados correctamente en un cuerpo. Izquierda: ViTPose-B condicionada por cuadro delimitador, derecha: MaskPose-b condicionada por máscaras.

Figura 2. BMP resuelve errores de detección (arriba y en el centro) y errores de pose (abajo) en OCHuman. Resultados cuantitativos en la Tabla 3.



(a) El número de puntos clave. Demasiados puntos dificultan el rendimiento.
Izquierda: 6 indicaciones para puntos clave; derecha: 13 indicaciones correctas.



(b) Indicación con y sin cuadro delimitador. La indicación con cuadro delimitador impide que SAM corrija las máscaras corporales fuera del cuadro delimitador.
Izquierda: RTMDet, centro: SAM con cuadro delimitador, derecha: SAM sin cuadro delimitador.

Figura 3. SAM: influencia de los parámetros de estímulo.