

[Indice](#)

Link code:

Autores: Yuxuan Liu, Guohui Zhou, Wei He, Hailong Zhu, Yanling Cui

Año: December 9, 2024

Pages: 13

"repeated downsampling → loss of detail information → poor small joints detection"

Proceso

- top_down approach: Detectar personas primero y luego estimar pose
 - DEM (Detail Enhancement Module)
 - Influenciado por: SENet (concepto, importancia de los canales) + ECANet (implementación eficiente para aprender importancia de canales)
 - Sin DEM:
 $\text{features_original} = (\text{canal1}, \text{canal2}, \text{canal3}, \dots, \text{canalN})$
 - Con DEM:
 $\text{features_con_DEM} = (w1\text{canal1}, w2\text{canal2}, w3\text{canal3}, \dots, wN\text{canalN})$
 - dySample (mejor que upsampling)
-

Dataset

- MS COCO dataset 250k anotaciones de personas con 17 puntos principales
 - MPII Human Pose Dataset con una división del dataset en un ratio de 7:6 para train y val
-

Parámetros

- Batch size 32 por GPU
- Optimizer: Adam
- Base learning rate: 1e-3

- drop a 1e-4 en la 170va epoca
- drop a 1e-5 en la 200va epoca
- Número de epocas: 210
- Image resize: 384x288
- Data Augmentation:
 - scale: $\pm 32\%$
 - rotation: ± 45 grados
 - Flipping
 - Half body data augmentation

Métricas de Evaluación

- Se uso Average Presicion (AP) y Average Recall (AR), ambas basadas en las similitud de los keypoints entre objetos (OKS)

$$OKS = \frac{\sum_i \exp\left(\frac{d_i^2}{2SK_i^2} \delta(V_i > 0)\right)}{\sum_i \delta(v_i > 0)}$$

- [PCKh@0.5](#)

Figuras

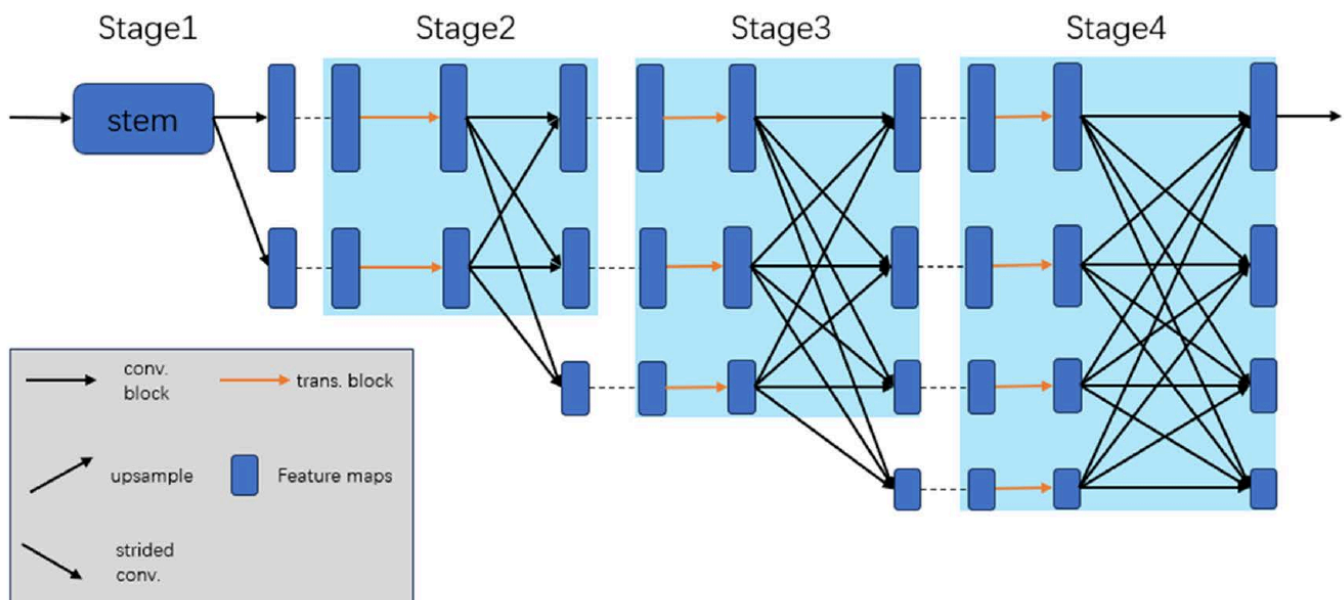


Fig. 1. Arquitectura de HRNet. Los módulos de la etapa **multirresolución** están marcados con

áreas de color azul. El módulo restante de tres etapas consta de subredes multirresolución paralelas con interacciones de información multirresolución.

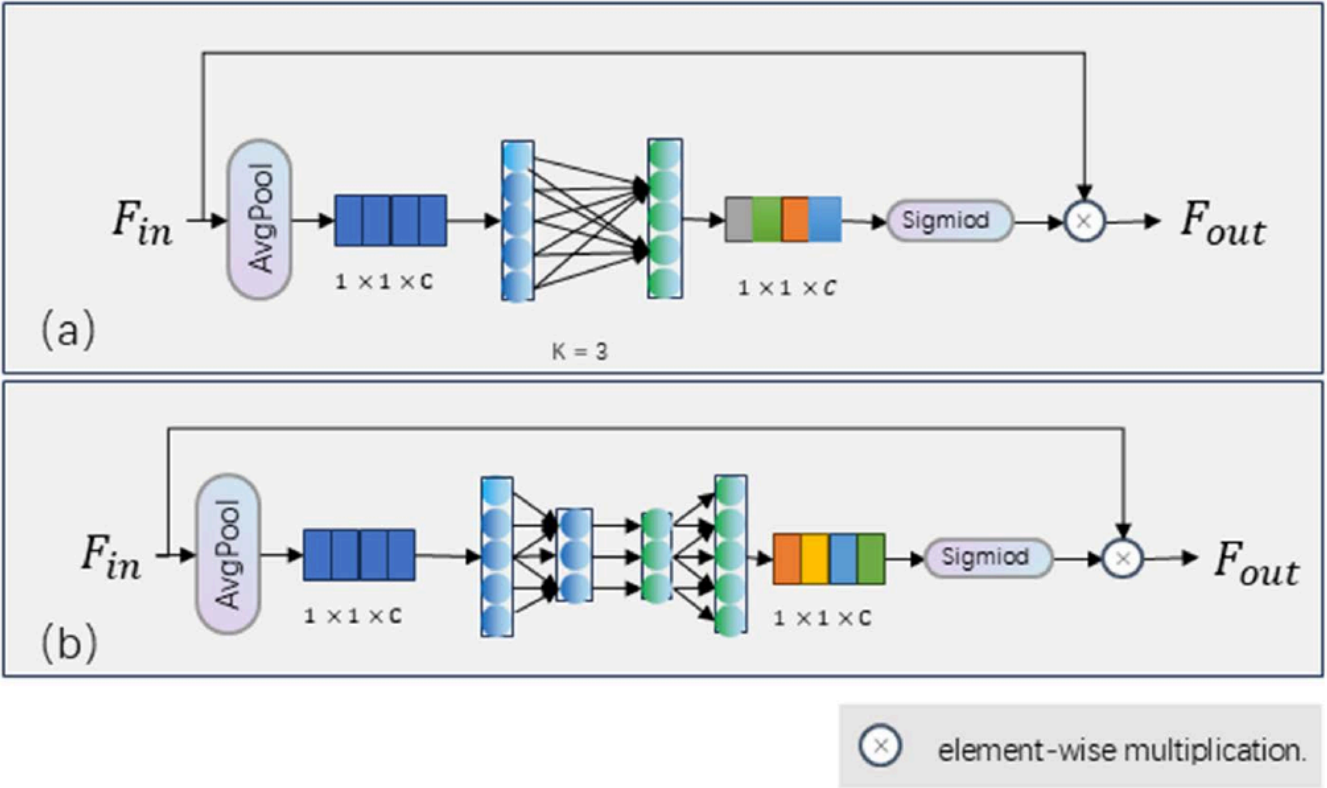


Fig. 2. Mecanismo de atención del canal. El bloque superior (a) es el bloque ECA, que consiste en agrupamiento promedio, una convolución unidimensional rápida de tamaño k y una activación sigmoidea. El bloque inferior (b) es el bloque SE, que consiste en agrupamiento promedio, dos capas completamente conectadas (FC) y una activación sigmoidea.

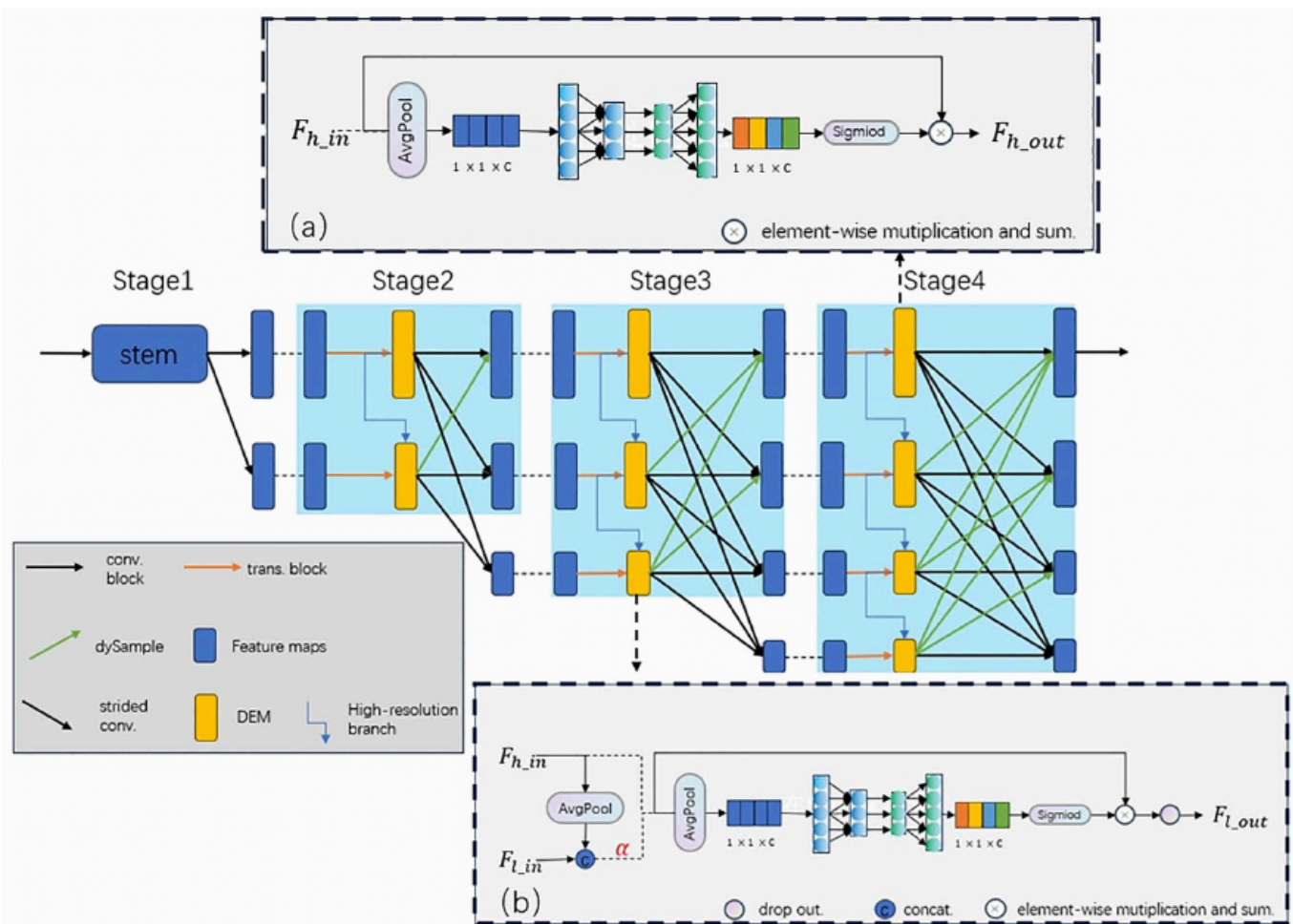


Fig. 3. Estructura de DE-HRNet. dySample y el Módulo de Mejora de Detalles (DEM) se aplican a HRNet.

(For the highest-resolution branch, (a) is adopted to enhance feature representation, which employs a single input and a single output. For other branches, (b) is utilized, featuring dual inputs (from high and low-resolution branches) and producing an output for the low-resolution branch)

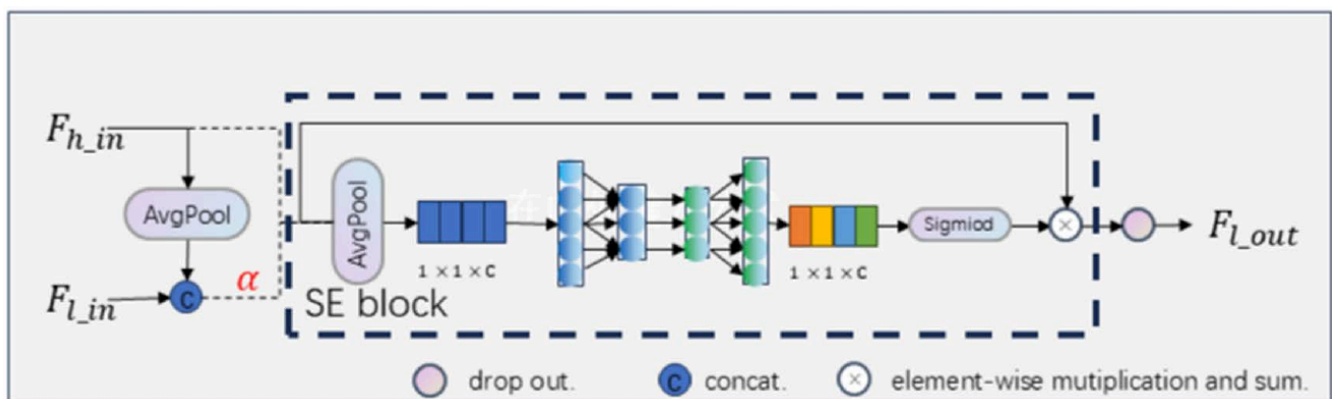


Figura 4. Estructura del Módulo de Mejora de Detalles. El módulo está diseñado con base en el bloque SE. Además, las tecnologías de agrupación de promedio global y de abandono se ubican antes y después del bloque SE, respectivamente. Las líneas discontinuas indican asignaciones de identidad selectiva.

$$SEA(x) = Sigmoid(FC(RELU(FC(GAP(x)))))$$

$$F_{h_{out}} = SEA(F_{h_n}) + 1)F_{h_n} \odot$$

$$F_{i_{out}} = dropOut((SEA(\alpha) + 1) \odot \alpha$$