—Supplementary Material—
Neural Interactive Keypoint Detection

Jie Yang[1,2*], Ailing Zeng[1†], Feng Li[1], Shilong Liu[1], Ruimao Zhang[2†], Lei Zhang[1]
[1]International Digital Economy Academy
[2]School of Data Science, Shenzhen Research Institute of Big Data,
The Chinese University of Hong Kong, Shenzhen
{jieyang5@link,zhangruimao@}cuhk.edu.cn
{zengailing,lifeng,liushilong,leizhang}@idea.edu.cn
https://github.com/IDEA-Research/Click-Pose

Overview

This supplementary material presents more details and additional results not included in the main paper due to page limitation. The list of items included are:

- Inference time comparison between the whole model and the only decoder in Sec. 1.
- Additional experimental analyses on human detection benefits in Sec. 2.
- Discussion the society impacts in Sec. 3.

1. Inference Time Comparison

We compare the time it takes for *Click-Pose*-C0 to generate prediction results with the time required for decoder loop refinement. The results presented in Tab. 1 demonstrate that receiving user feedback at the decoder is more efficient than at the input image.

Methods	Inference time [ms]
Click-Pose-C0	48
Decoder Loop	**12**

Table 1: Comparisons of **the inference time**.

2. Benefit to Human Detection

Besides the significant improvements in human keypoint detection tasks, as shown in the main paper, the proposed interactive human-feedback loop in *Click-Pose* can also help to adjust the positions of human boxes for better human detection. We use the maximum and minimum values of refined keypoints for each person to regularize the width and length of the box. We take the three clicks (C3) and five clicks (C5) as examples, Tab. 2 demonstrates that *Click-Pose* can consistently improve AP_M and AP_L on both in- and out-of-domain datasets.

Click	None	C3	C5	None	C3	C5
	COCO `val`			*Human-Art* `val`		
AP_M	68.6	69.8	**70.3**	3.7	8.2	**9.2**
AP_L	79.0	80.0	**80.4**	14.6	21.3	**22.8**

Table 2: The *Click-Pose*'s impact on **Human Detection**.

3. Society Impact

With the development of deep learning, the ability of large models in many fields has almost reached the level of human knowledge, especially in creative generation types of work (e.g., recent Stable diffusion models on image generation), which has developed much faster than expected. While humans often worry and fear that machines will replace humans and the work they are doing, this article, by exploring how better collaboration with humans on interactive keypoint annotation can reduce repetitive manual efforts and increase the interactive fun of the work, allowing humans to be more productive and spend more time on making critical decisions. Meanwhile, this work explores the problems of the model bias and performance bottleneck, which makes humans/annotators/users essential to high-quality annotations. Introducing human feedback in an interactive way can improve the usability, transparency, and trustworthiness of deep models. This will help to enable better human-machine interaction and a wider range of human-centric applications, like healthcare, avatar creation, and autonomous robotics. At last, using a fast and efficient annotation method can create more high-quality data, making existing large models greater continuously. The data and model will boost each other significantly.

*Work done during an internship at IDEA.
†Corresponding author.