

TITLE: Human pose understanding in video using deep neural networks

TYPE: Research internship in Computer Vision

LEVEL: Master 1 or 2, Fourth or Fifth-year Engineer in Computer Science or close-related field.

SUPERVISION:

Carlos Crispim Junior (LIRIS)

Yoann Blanche (LIBM)

LOCATION OF INTERNSHIP:

LIRIS UMR CNRS 5205,

Université Lumière Lyon 2,

5 avenue Pierre Mendès-France, 69676 Bron CEDEX

KEYWORDS: deep learning, video understanding, human behavior analysis, diffusion models, visual transformers

CONTEXT OF THE STUDY:

This internship offer is proposed in the context of the PoseVertex project, funded by the *Fédération Informatique de Lyon*. The objective of this project is to study computer vision methods for human pose understanding in healthcare applications, in particular, upper limb body movement and related disorders. Recent advances in artificial intelligence have greatly improved methods for human keypoint estimation in 2D images captured by a monocular camera. These advances have democratized pose understanding, giving birth to applications ranging from avatar animation to human-computer interfaces using human body as a medium (Figure 1A). However, transferring such methods to healthcare applications, like body posture and movement diagnosis, requires higher precision and robustness in the extracted measures. However, even if vision foundation models exist for HPE, certain healthcare applications, such as upper limb keypoint estimation, require finetuning existing methods for keypoint classes absent in generic datasets. Another challenge is that the new classes are often depicted in a limited set of available examples. This creates a learning challenge since recent approaches tend to achieve their top performance when trained with large datasets [3-4].

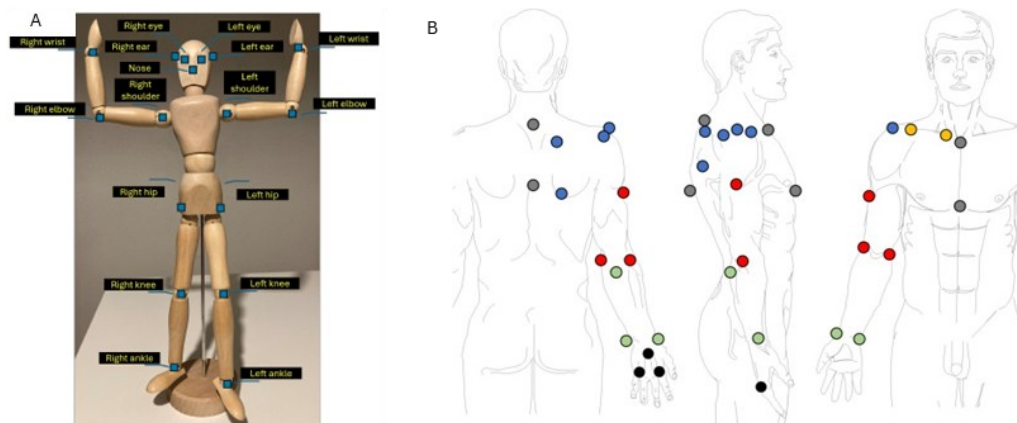


Figure 1. Examples of keypoint classes in human pose understanding: A) generic keypoints (17 classes adapted from [2]) and B) healthcare application-driven keypoints (20 classes, adapted from [1]).

OBJECTIVE

The goal of this internship will be to study, evaluate and adapt existing methods for 2D and 3D human pose understanding in videos of the target healthcare problem. The challenges include new classes of human body keypoints, limited training data, and the need for accurate estimation of keypoint position in space over time.

BROAD VIEW OF THE EXPECTED TASKS:

- Review the state of the art on methods for video based human pose understanding, identifying their success cases and limitations in general and in the context of the target application.
- Build a HPE video benchmark highlighting examples of the open challenges identified.
- Develop a parameter efficient fine-tuning [5] approach to adapt the selected approaches (*e.g.*, Visual Transformer, diffusion models, or MAMBA) to the target problem,
- Write a research article to share the developed work with the computer vision community, accompanied by an open-source repository to foster reproducible research.

PROFILE OF THE CANDIDATE:

We are looking for a motivated candidate with a strong background in computer science, applied mathematics, or related field.

- The candidate must currently be enrolled in a Master 1 or 2 program in Computer Science, or be in the final years of engineering school (Bac+4 or +5 in France) in a related field
- Experience in image processing, computer vision, and/or machine learning will be a plus.

If the internship leads to an international publication, we may study opportunities to pursue the research carried out with a PhD in a similar topic.

LANGUAGE: French or English

EXPECTED SKILLS:

- Mastering of Python language
- OpenCV library
- Versioning tools (GIT)

The following skills would be considered as a plus:

- Framework PyTorch or TensorFlow.
- Docker-like tools and platforms

DURATION: 6 months

EXPECTED INTERNSHIP PERIOD: February to August 2027 (a summer break is imposed mid-July to mid-August)

PERIOD OF EVALUATION OF APPLICATIONS : late October/early November

SALARY: “gratification de stage” in France

CONTACT : carlos.crispim-junior@liris.cnrs.fr

APPLICATION: Send an email to the address above with your curriculum vitae and your university transcripts concerning your last two years of undergraduate studies. **Please be mindful that your application email files do not exceed 2 MB in total.**

RELATED BIBLIOGRAPHIC REFERENCES:

1. F. Lefebvre, I. Rogowski, N. Long, Y. Blache, Three-dimensional markerless pose estimation for anatomical landmarks of the shoulder and upper limb, Journal of Electromyography and Kinesiology, Volume 85, 2025, 103067, ISSN 1050-6411, <https://doi.org/10.1016/j.jelekin.2025.103067>.

2. C. Crispim-Junior, Multimodal Computer Vision: exploiting domain knowledge for visual data generation and understanding, HDR, soutenue le 22 septembre 2025, 183pages.
3. Y. Xu, J. Zhang, Q. Zhang and D. Tao, "ViTPose++: Vision Transformer for Generic Body Pose Estimation" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 46, no. 02, pp. 1212-1230, Feb. 2024, [https://doi: 10.1109/TPAMI.2023.3330016](https://doi.org/10.1109/TPAMI.2023.3330016).
4. R. Khirodkar, T. Bagautdinov, J. Martinez, S. Zhaoen, A. James, P. Selednik, S. Anderson, and S. Saito. 2024. Sapiens: Foundation for Human Vision Models. In Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part IV. Springer-Verlag, Berlin, Heidelberg, 206–228. https://doi.org/10.1007/978-3-031-73235-5_12
5. Yi Xin, Jianjiang Yang, Siqi Luo, Yuntao Du, Qi Qin, Haoxing Chen, Kangrui Cen, Yangfan He, Bin Fu, Yuewen Cao, Junjun He, Xiaokang Yang, Guangtao Zhai, Ming-Hsuan Yang, and Xiaohong Liu. 2026. Parameter-Efficient Fine-Tuning for Pre-Trained Vision Models: A Survey and Benchmark. Int. J. Comput. Vision 134, 6 (Jun 2026).