Vidéos, poses et allures : un jeu de données pour l'analyse de la locomotion du cheval

Benoît Pasquiet[*,**], Faten Chaieb-Chakchouk[**]
Cheima Trabelsi[**] Qiudi He[**], Katarzyna Wegrzyn-Wolska[**]

[*]Institut Français du Cheval et de l'Equitation
170 avenue du Cadre noir, 49400 Saumur, France
benoit.pasquiet@ifce.fr,
[**]Efrei Research Lab, Université Paris-Panthéon-Assas
30-32 Av. de la République, 94800 Villejuif, France

En équitation, la discipline du dressage consiste à apprécier les qualités des mouvements d'un cheval par un jury. Avec une vidéo le cavalier et son entraîneur peuvent revoir les figures exécutées. Le suivi de la pose du cheval permettrait d'enrichir ces vidéos par des indicateurs sur la locomotion du cheval.

De multiples travaux se sont intéressés à l'estimation de pose chez les animaux. Mathis et al. (2018) ont développé DeepLabCut. Le modèle le plus précis pour la reconstitution de pose sur une image 2D pour des animaux quadrupèdes est VitPose+ (Xu et al., 2022).

Plusieurs jeux de données existent pour l'estimation de pose chez les animaux. Horse-10 (Mathis et al., 2021) et PFERD (Li et al., 2024) sont dédiés aux chevaux. D'autres bases, comme APT-36K (Yang et al., 2022) incluent plusieurs espèces de quadrupèdes. Dans ces bases, le cheval n'est pas monté par un cavalier. Dans Horse-10, le cheval se déplace toujours de gauche à droite, perpendiculairement à l'axe de la caméra. Par ailleurs, seul PFERD précise le mouvement que réalise le cheval.

Nous proposons un nouveau jeu de données, associant vidéos, pose 2D sur chaque image, et description de l'allure, de chevaux montés. Ces vidéos proviennent pour l'essentiel de la chaîne YouTube de la Fédération Equestre Internationale (FEI). Des séquences aux trois allures principales (pas, trot, galop) ont été extraites de ces vidéos. Les vidéos ont été choisies afin de proposer une variété dans les sols (sable, herbe), les robes des chevaux, les couleurs de leurs sabots (claires, sombres), la direction du mouvement. La figure 1 présente le squelette appliqué dans notre modèle. Ce squelette est le premier à proposer des points sur la queue du cheval, dont le mouvement est une indication de la souplesse du cheval (Fédération française d'équitation, p. 6).

Nous avons effectué un apprentissage actif, en entraînant un modèle VitPose+ au fur et à mesure de l'annotation des images. Toutes les images ont été contrôlées et corrigées manuellement. Ce jeu de données comprend 49 séquences, d'une durée moyenne de 5.64 secondes, pour un total de 8.696 images annotées. Nous avons 17 séquences pour 3040 images au pas ; 16 séquences pour 2928 images au trot et 16 séquences pour 2738 images au galop.

En utilisant 80% de nos données pour l'entraînement et 20% pour le test, nous avons comparé plusieurs modèles d'estimation de pose, dont VitPose+. Pour ce dernier, la distance entre

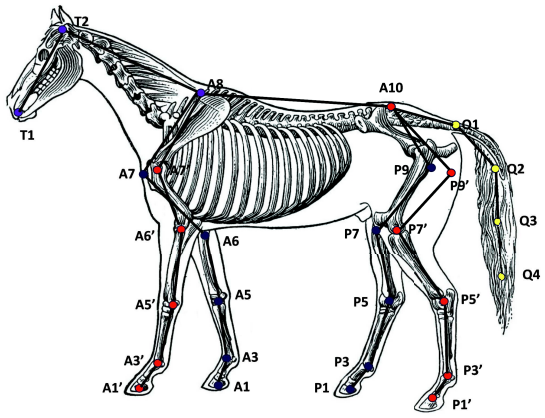

FIG. 1 – *Squelette utilisé pour annoter les images.*

le point réel et le point prédit est inférieure à 5% de la taille de la boîte englobante du cheval pour 86% des points.

Ce jeu de données ouvre la voie à une extraction automatisée de paramètres locomoteurs, tels que la cadence et la régularité des phases d'appui, sur un cheval en compétition ou à l'entraînement, à partir de vidéos. Il pourrait être complété par des mouvements plus complexes, tels que les pirouettes et les piaffers.

Références

Fédération française d'équitation. *Echelle de progression*. Les fondamentaux de la formation du cheval.

Li, C., Y. Mellbin, J. Krogager, S. Polikovsky, M. Holmberg, N. Ghorbani, M. J. Black, H. Kjellström, S. Zuffi, et E. Hernlund (2024). The Poses for Equine Research Dataset (PFERD). *Scientific Data 11*(1), 497, doi: 10.1038/s41597-024-03312-1.

Mathis, A., T. Biasi, S. Schneider, M. Yuksekgonul, B. Rogers, M. Bethge, et M. W. Mathis (2021). Pretraining boosts out-of-domain robustness for pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1859–1868.

Mathis, A., P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, et M. Bethge (2018). Deeplabcut : markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience 21*, 1281–1289, doi: 10.1038/s41593-018-0209-y.

Xu, Y., J. Zhang, Q. Zhang, et D. Tao (2022). ViTPose+ : Vision Transformer Foundation Model for Generic Body Pose Estimation. *arXiv preprint arXiv :2212.04246*.

Yang, Y., J. Yang, Y. Xu, J. Zhang, L. Lan, et D. Tao (2022). Apt-36k : A large-scale benchmark for animal pose estimation and tracking. *Advances in Neural Information Processing Systems 35*, 17301–17313.