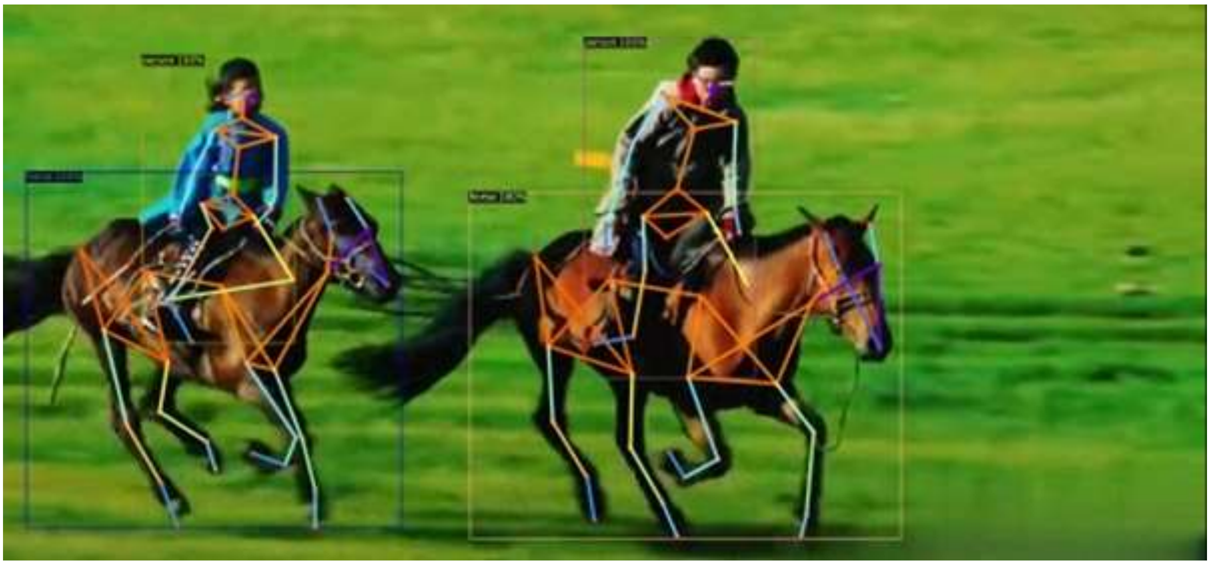


同时进行人体和动物姿态估计：一种增强性能洞察的多目标方法

Simultaneous human and animal pose estimation: A multi-objective approach to enhance performance insights

Tarik Aslan 2022.10 – Beijing Tangta technology

Abstract 文章摘要 :

本文探讨了身体关键点检测的进展，重点介绍了其在人类和动物中的应用。具体而言，本文重点介绍了一个在 2021 年至 2023 年间开发的项目，该项目旨在通过一个针对马匹和骑手的双关键点检测系统来优化赛马训练。该系统利用 Detectron2 和一个定制的 COCO 标注器来检测 25 个关键点，以实现精确的姿态分析。这项工作解决了一个独特的挑战，即如何在一帧中处理多个主体（马匹和骑手），这通过一个将边界框输入姿态估计器的目标检测流程来实现。该项目的成果展示了关键点检测在性能优化、伤害预防和跨物种分析方面的潜力，为未来在体育、医疗保健和动物福利领域的应用铺平了道路。

This paper explores advances in body keypoints detection, focusing on its applications in humans and animals. Specifically, this paper highlights a project developed between 2021 and 2023 that aims to optimize racehorse training through a dual keypoints detection system for horses and riders. The system utilizes Detectron2 and a custom COCO annotator to detect 25 keypoints for accurate pose analysis. This work addresses the unique challenge of handling multiple subjects (horse and rider) in a single frame, which is achieved through an object detection pipeline that feeds bounding boxes into a pose estimator. The results of this project demonstrate the potential of keypoints detection

for performance optimization, injury prevention, and cross-species analysis, paving the way for future applications in sports, healthcare, and animal welfare.

Introduction 文章简介:

识别人体关键点（人体的关键关节和解剖标志）已成为现代计算机视觉领域最重要的功能之一。虽然这个概念看似技术性很强，但其含义却与人类息息相关：理解人类如何移动、互动和表达意图，对于任何试图与我们共存或合作的机器而言，都至关重要。

过去，捕捉人体动作仅限于专业实验室或好莱坞工作室。这需要昂贵的动作捕捉服、光学标记以及精心校准的环境。但随着深度学习和卷积神经网络的兴起，一切都发生了改变。大约在 2010 年代中期，卡内基梅隆大学的 OpenPose 和谷歌的 PoseNet 等突破性进展向世界展示了仅使用标准摄像头和深度神经网络就能实时检测身体姿势的可能性。

Recognizing body keypoints — the critical joints and anatomical landmarks of the human body — has become one of the most important capabilities in modern computer vision. While the concept may seem technical, its implications are deeply human: understanding how people move, interact, and express intention is fundamental for any machine trying to coexist or cooperate with us.

Historically, capturing human motion was confined to specialized labs or Hollywood studios. It required expensive motion-capture suits, optical markers, and carefully calibrated environments. But everything changed with the rise of deep learning and convolutional neural networks. Around the mid-2010s, breakthroughs like **OpenPose** from Carnegie Mellon and **PoseNet** from Google showed the world that it was possible to detect body poses in real time, using only a standard camera and deep neural networks.

这一演变过程中的一个关键里程碑是 Facebook AI Research 的 Detectron 及其后续产品 Detectron2——一个显著提升物体检测和人体姿态估计的开源平台。Detectron 整合了一个强大的模块化、高性能关键点检测模型生态系统，这些模型基于 COCO 等大规模数据集进行训练。它在保持实时性能的同时，突破了准确率的极限，并树立了该领域的新标杆。

A key milestone in this evolution was **Facebook AI Research's Detectron** and its successor **Detectron2** — an open-source platform that dramatically advanced object detection and human pose estimation. Detectron brought together a powerful ecosystem of modular, high-performance models for keypoints detection, trained on large-scale datasets like COCO. It pushed the accuracy frontier while maintaining real-time performance, and set new standards in the field.

这些成就并非仅仅由算法驱动——它们与 2015 年至 2020 年间 GPU 计算能力的崛起息息相关。NVIDIA 的消费级和工作站 GPU 使大规模训练和运行深度学习模型成为可能。这一时期为实时身体关键点估计走出实验室，进入现实世界奠定了计算基础。如今，我们正在见证这一变革对无数领域的影响。

These achievements weren't driven by algorithms alone — they were tightly linked to the rise of **GPU computing power between 2015 and 2020**. NVIDIA's consumer and workstation GPUs made it possible to train and run deep learning models at scale. This period laid the computational foundation for real-time body keypoints estimation to leave the lab and enter the real world.

Today, we’re witnessing the impact of this evolution across countless domains.

在机器人技术中，身体关键点是连接机器和人类行为的桥梁。能够识别人类是在指指点点、挥手示意还是跌倒的机器人，不仅能够感知环境，还能提升社交智能。在协作环境中——例如工厂、智能家居或城市空间——这使得机器人能够安全直观地与人类互动，并展现同理心而非僵硬的反应。

在医疗保健和健身领域，身体姿势追踪能够以非侵入式的方式监测体形、检测损伤并支持身体康复。无论是帮助患者在运动过程中纠正姿势，还是评估康复期间的活动范围，AI 驱动的运动追踪正在让实时健康反馈的获取变得普及。

In robotics, body keypoints are the bridge between machines and human behaviour. A robot that can recognize whether a person is pointing, waving, or falling gains not just situational awareness but social intelligence. In collaborative settings — factories, smart homes, or urban spaces — this allows robots to act safely and intuitively around humans, reacting with empathy instead of rigidity.

In healthcare and fitness, body pose tracking enables a non-invasive way to monitor form, detect injuries, and support physical rehabilitation. Whether it’s helping correct posture during exercise or assessing range of motion during recovery, AI-driven motion tracking is democratizing access to real-time health feedback.

在 AR/VR 和日益壮大的空间计算领域，关键点是全身虚拟形象和自然手势控制的基础。随着数字世界变得越来越沉浸，身体追踪能够确保用户保持临场感并富有表现力——无需手套、套装或昂贵的传感器。

甚至公共安全和无障碍设施也能从中受益。增强了姿势估计功能的监控系统可以检测到诸如倒塌或不稳定运动等异常情况。对于残障人士来说，基于身体追踪构建的手势界面提供了与设备和周围世界互动的全新方式——无需触摸或语音，直观便捷。

In AR/VR and the growing world of spatial computing, keypoints underpin full-body avatars and natural gesture control. As digital worlds become more immersive, body tracking ensures users remain present and expressive — without the need for gloves, suits, or expensive sensors.

Even public safety and accessibility benefit. Surveillance systems enhanced with pose estimation can detect anomalies like collapsing or erratic movement. For individuals with disabilities, gesture-based interfaces built on body tracking offer new ways to interact with devices and the world around them — intuitively, without touch or speech.

至关重要的是，这些功能不再局限于云端或高端系统。MediaPipe、BlazePose 等优化模型以及各种基于 YOLO 的轻量级姿态追踪器，如今已能够在手机和 Orange Pi 或 Jetson Nano 等边缘设备上高效运行。这为隐私保护的实时智能打开了大门，它可以离线运行，功耗极低，并且可轻松掌控。

归根结底，识别身体关键点的意义远不止追踪肢体，它还关乎理解动作、意图和互动。这是迈向机器不仅能“看见”我们，还能“理解”我们的第一步。而这将改变一切。

Crucially, these capabilities are no longer confined to the cloud or high-end systems. Optimized models like **MediaPipe**, **BlazePose**, and various lightweight YOLO-based pose trackers now run efficiently on mobile phones and **edge devices** like Orange Pi or Jetson Nano. This opens the door to **privacy-preserving, real-time intelligence** that can function offline, with minimal power, and in the palm of your hand.

In the end, recognizing body keypoints isn't just about tracking limbs — it's about understanding motion, intention, and interaction. It's the first step toward machines that don't just *see* us — but *comprehend* us. And that changes everything.

https://tangta-technology.com/video/a2b38f40abcce38b282caf6b537b7be9_result.mp4

视频来自友口，关键点检测来自我们的软件

Extending Keypoints Detection to Animals: 将关键点检测扩展到动物：

虽然人体关键点检测已发展成为一个成熟且应用广泛的领域，但动物关键点检测仍处于萌芽阶段，并面临着一系列独特的挑战。其核心原理相似：检测解剖学特征以估计姿势和行为。然而，不同物种的动物生理机能差异巨大，这使得泛化难度加大。与拥有共同身体结构和可预测关节位置的人类不同，动物的形态多样性更为广泛——从四足哺乳动物、有翼鸟类到爬行动物和水生物种。毛发、羽毛、尾巴和非标准的身体姿势进一步增加了检测的复杂性，通常需要针对特定物种的模型或注释。

While human keypoints detection has become a mature and widely applied field, animal keypoints detection is still emerging — and comes with its own set of unique challenges. The core principle is similar: detecting anatomical landmarks to estimate pose and behaviour. However, animal physiology varies widely across species, making generalization far more difficult. Unlike humans, who share a common body structure and predictable joint positions, animals present a much broader morphological diversity — from four-legged mammals and winged birds to crawling reptiles and aquatic species. Fur, feathers, tails, and non-standard body postures further complicate detection, often requiring species-specific models or annotations.

尽管存在这些困难，一些工具和数据集已开始在该领域取得进展。DeepLabCut、LEAP 和 OpenMonkeyStudio 等系统已被开发用于高精度追踪动物姿势，即使在自然环境中也是如此。这些框架通常利用迁移学习，允许研究人员在小型带注释的数据集上微调模型，使其在实验室或野生动物研究环境中发挥强大作用。

Despite these difficulties, several tools and datasets have begun to make progress in this field. Systems like DeepLabCut, LEAP, and OpenMonkeyStudio have been developed to track animal pose with high precision, even in natural environments. These frameworks often leverage transfer learning and allow researchers to fine-tune models on small annotated datasets, making them powerful in laboratory or wildlife research settings.

然而，动物关键点检测的研究兴趣仍然相对小众，尤其是与人体姿态估计相比。部分原因在于缺乏大规模、多样化且标注良好的数据集，这限制了通用模型的训练。此外，商业需求也并非那么迫切或明显：大多数行业驱动的应用都专注于以人为本的领域，例如健身、安全和扩展现实(XR)。因此，动物姿态追踪的大部分工作仍然由神经科学、动物行为学和兽医学领域的学术研究团队主导，而非主流科技公司。

However, interest in animal keypoint detection remains relatively niche — especially compared to human pose estimation. This is partly due to the lack of large, diverse, and labeled datasets, which limits the training of generalized models. Additionally, the commercial demand is not as immediate or obvious: most industry-driven applications focus on human-centric fields like fitness, security, and XR. As a result, much of the work on animal pose tracking is still led by academic research groups in neuroscience, ethology, and veterinary science, rather than mainstream tech companies.

尽管如此，其潜力依然巨大。精准的动物关键点检测可以改变一切，从动物园里的牲畜健康监测和行为追踪，到野生动物保护工作。随着人工智能硬件价格越来越低、移动性越来越强，以及人们对动物福利和环境可持续性的关注日益加深，这一领域在未来几年可能会受到越来越多的关注。

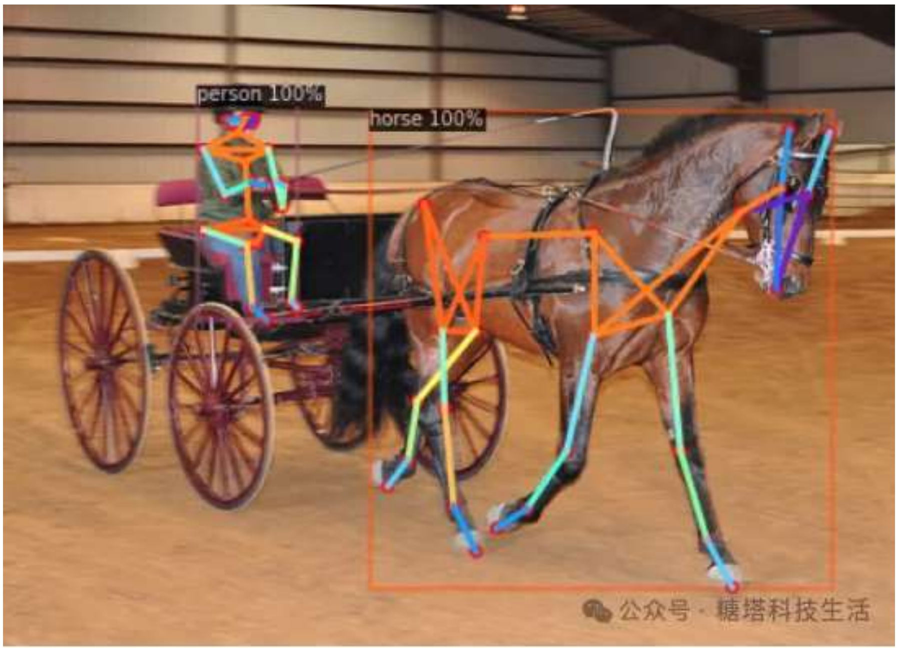
That said, the potential is significant. Accurate animal keypoint detection could transform everything from livestock health monitoring and behavioral tracking in zoos to conservation efforts in the wild. As AI hardware becomes cheaper and more mobile, and as concerns about animal welfare and environmental sustainability grow, this field may yet see a surge of attention in the years to come.

The Road Ahead — Future Applications of Animal Keypoint Detection

未来之路——动物关键点检测的未来应用

随着动物关键点检测技术的成熟，一波新的应用浪潮即将到来。在精准农业领域，姿势追踪可用于持续监测牲畜健康状况——无需人工干预即可检测跛足、疲劳或异常行为的早期迹象。在兽医远程医疗领域，远程诊所可以使用基于摄像头的系统实时分析动物的运动和关节功能，从而实现早期诊断并减轻动物的压力。野生动物研究人员可能很快会部署搭载人工智能模型的自主无人机或嵌入式相机陷阱，这些模型不仅可以检测物种，还可以追踪行为模式——这将彻底改变濒危物种的监测格局。

As animal keypoints detection technology matures, a wave of new applications is on the horizon. In precision agriculture, pose tracking could be used to monitor livestock health continuously — detecting early signs of lameness, fatigue, or abnormal behaviour without human intervention. In veterinary telemedicine, remote clinics could use camera-based systems to analyse an animal’s movement and joint function in real time, enabling earlier diagnosis and reducing stress on the animal. Wildlife researchers may soon deploy autonomous drones or embedded camera traps enhanced with AI models that not only detect species but also track behaviour patterns — a game changer for endangered species monitoring.



展望未来，动物与机器人的互动或许会成为一个独立的研究领域。例如，部署在农场或自然保护区的服务或监测机器人可以利用姿态估计来解读动物的状态或情绪线索，从而改善科技与生物的共存。动物生物力学与人工智能的融合或许还能为仿生机器人技术开辟新的方向，了解动物的运动方式将有助于设计更灵活的机器。

最终，随着技术能力和伦理利益的增长，动物关键点检测可能不仅仅是一种科学工具，更将成为促进人与动物更好共存、动物福祉和环境管理的催化剂。

Looking even further ahead, animal-robot interaction could become a research domain of its own. For example, service or monitoring robots deployed on farms or in natural reserves could use pose estimation to interpret the state or emotional cues of animals, improving how technology coexists with living creatures. The fusion of animal biomechanics and AI may also unlock new paths in bio-inspired robotics, where understanding how animals move informs the design of more agile machines.

Ultimately, as both the technological capability and ethical interest grow, animal keypoints detection may become not just a scientific tool — but a catalyst for better human-animal coexistence, animal welfare, and environmental stewardship.

构建用于赛马优化的双关键点检测器

Building a Dual Keypoints Detector for Racehorse Optimization

2021 年至 2023 年间，我们的团队一直在开展一个并行研发项目，旨在提升竞技体育（尤其是赛马及其骑手）的关键点检测能力。基于 Detectron2，我们使用专有数据集训练了一个自定义姿态估计模型，该数据集使用我们自己改进的 COCO 注释器进行了精心注释。该项目专注于四足动物的检测，特别是训练中的高性能纯种马。通过同时捕捉马匹和骑手的关键点，我们旨在

获取有意义的姿态信息，以优化运动表现、预防伤病并分析比赛。这种双重检测方法旨在弥合生物力学与训练策略之间的差距，为教练和分析师提供强大的工具，帮助他们了解每次训练的动态，并更好地解读运动表现中的差距。

Between 2021 and 2022, our team undertook a focused research and development project aimed at advancing keypoints detection for performance sports — specifically targeting **racehorses and their riders**. Using **Detectron2** as the backbone, we trained a custom pose estimation model on a proprietary dataset that we meticulously annotated through a **self-modified version of the COCO Annotator**. The project centred on quadruped detection, with a primary focus on high-performance **thoroughbred horses in training environments**. By capturing synchronized keypoints from both the horse and the equestrian athlete, our goal was to extract meaningful body pose information that could support **performance optimization, injury prevention**, and **post-race analysis**. This dual-species detection approach aimed to bridge the gap between biomechanics and training strategy, offering trainers and analysts a powerful tool to understand the dynamics of each training session — and better interpret instances of underperformance.

扩展解剖结构——从标准关键点到统一的 25 点模型

Expanding the Anatomy — From Standard Keypoints to a Unified 25-Point Model

在项目的初始阶段，我们采用一种常用的 17 个关键点格式——这种布局由 COCO 推广，并被大多数人体姿势估计系统广泛采用。这包括鼻子、眼睛、耳朵、肩膀、肘部、手腕、臀部、膝盖和脚踝等关键点。虽然这种配置在人体检测方面效果不错，但我们很快发现，它不足以捕捉赛马的详细生物力学特征，尤其是在精确的肢体关节和步幅动态至关重要的情况下。

In the initial phase of our project, we started with a commonly used 17-keypoint format — a layout popularized by COCO and widely adopted in most human pose estimation systems. This included keypoints such as the nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. While this configuration worked reasonably well for human detection, we quickly discovered that it was **insufficient for capturing the detailed biomechanics of a racehorse**, especially when precise limb articulation and stride dynamics were essential.

一个关键限制是背部、腿部和足部检测缺乏精细度。在标准模型中，人体躯干通常被视为单一的刚性块（例如人形机器人的躯干）——这种过于简化的模型无法反映脊柱的自然灵活性，尤其是在动态运动过程中。对于马术运动员和赛马而言，背部的曲度和排列在平衡、力量传递和整体姿势中都起着至关重要的作用。如果背部没有足够的关键点，就几乎不可能分析重心的变化、检测脊柱的拱起或凹陷，或理解运动过程中力量的分布情况。同样，腿部和足部（或蹄子）缺乏详细的关节标记，限制了我们解读步幅力学和地面接触动力学的能力——这两者对于运动表现优化和伤病预防都至关重要。

One critical limitation was the **lack of granularity in back, leg, and foot detection**. In standard models, the human torso is often treated as a single rigid block (like a humanoid robot torso) — an oversimplification that fails to reflect the **natural flexibility of the spine**, especially during dynamic movement. For both equestrian athletes and racehorses, the **curvature and alignment of the back** play a vital role in balance, power transfer, and overall posture. Without sufficient keypoints along the back, it becomes nearly impossible to analyse shifts in center of mass, detect arching or hollowing of the spine, or understand how forces are distributed during motion. Likewise, the lack of detailed joint markers on the legs and feet (or hooves) limited our ability to interpret stride mechanics and ground contact dynamics — both of which are critical in performance optimization and injury prevention.

此外，我们的目标是开发一个能够同时处理马匹和骑手的单一系统。这带来了另一个限制：为了将两个物种的数据输入到一个共享模型中，关键点的数量和顺序需要在人类和动物检测任务中保持一致。标准的人体布局与动物的解剖结构不兼容，反之亦然。

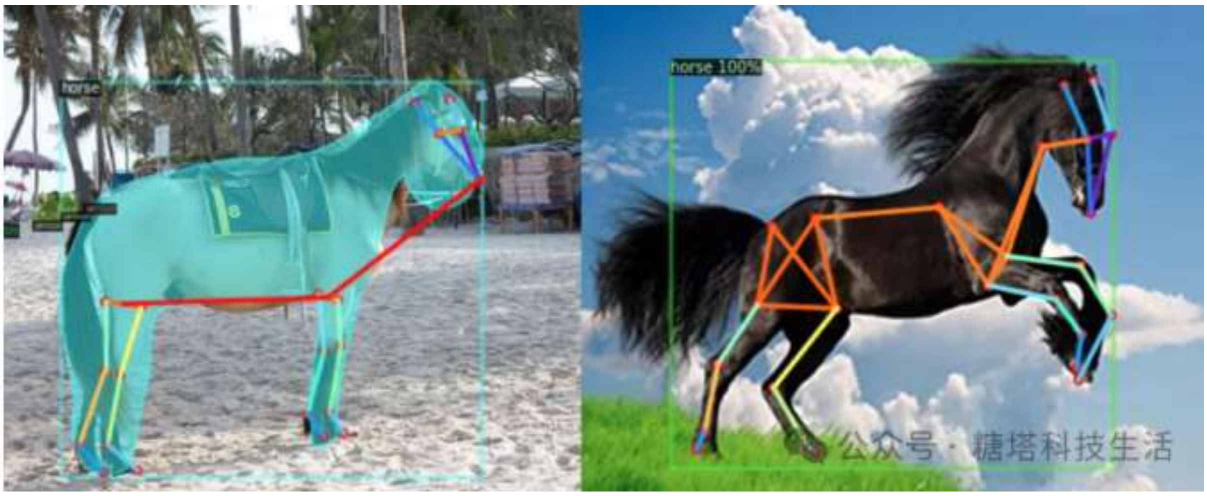
Moreover, we aimed to develop **a single system capable of processing both horse and rider** simultaneously. This introduced another constraint: to feed both species into a shared model, the **number and order of keypoints needed to be consistent** across human and animal detection tasks. The standard human layout was incompatible with animal anatomy, and vice versa.

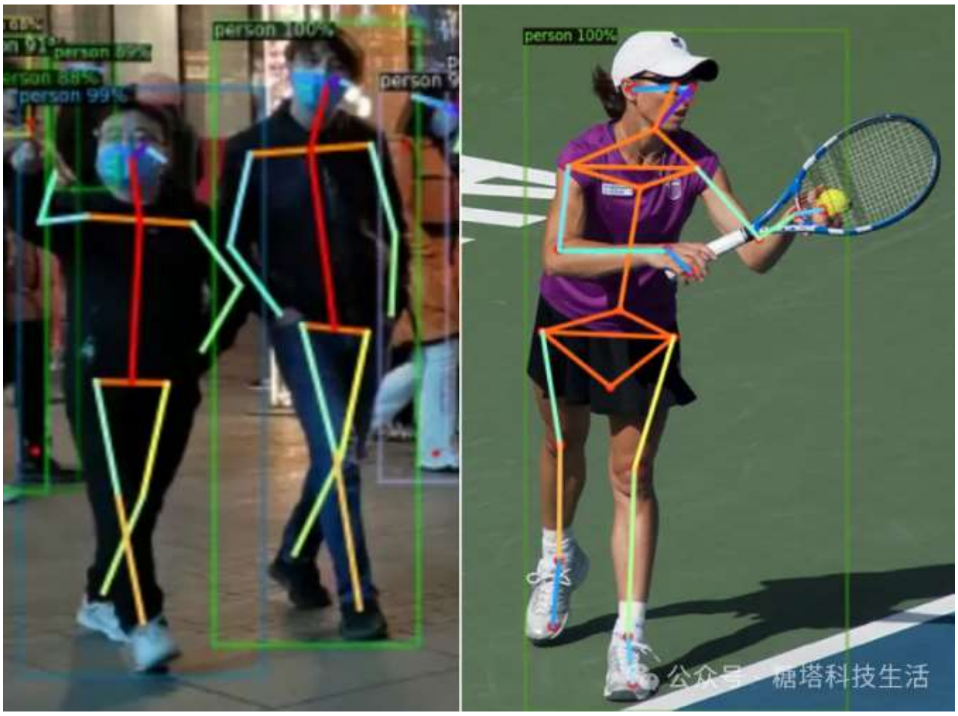
为了应对这些挑战，我们将标注方案扩展至 25 个关键点。每个肢体（动物的前肢和后肢，以及人类的手臂和腿部）都分配了四个关键点，以便更好地捕捉从肩部或臀部到足部或蹄部的完整关节活动。耳朵、眼睛和鼻子的标注与标准关键点保持一致。此外，我们还在背部或脊柱处引入了四个专用关键点，这对于评估骑手的平衡能力和马匹在运动过程中的弯曲度尤为重要。这种统一的 25 点模型能够提供更丰富的姿态表征，并使得两个物种能够使用相同的检测流程，从而显著简化训练和推理逻辑。

To address these challenges, we extended the annotation scheme to **25 keypoints**. Each limb — both forelimbs and hindlimbs in animals, and arms and legs in humans — was assigned **four keypoints** to better capture full articulation from the shoulder or hip to the foot or hoof. The **ears, eyes, and nose** remained consistent with standard landmarks. Additionally, we introduced **four dedicated keypoints along the back or spine**, which proved especially important for evaluating rider balance and horse curvature during motion. This unified 25-point model enabled a **much richer pose representation** and made it possible to use the same detection pipeline for both species, dramatically simplifying training and inference logic.

创建此 25 个关键点配置（针对人类和马）的可视化表示是为了验证注释的一致性并指导我们的模型学习 - 确保从实际训练课程收集的数据可以由教练和分析师直接解释，而无需额外的后期处理。

A visual representation of this 25-keypoint configuration (for both human and horse) was created to validate annotation consistency and guide our model’s learning — ensuring that data collected from real training sessions could be directly interpreted by coaches and analysts without extra post-processing.





突破单物体限制——多主体关键点检测
Breaking through the limitation of single object - multi-subject key point detection

现有大多数身体姿态检测系统的另一个主要限制是，它们每帧只能处理单个人物。许多标准实现只假设单个人物，无法扩展到多人场景——甚至更糟的是，多个物种的场景。这给我们的项目带来了巨大的挑战，因为我们需要同时分析马匹和骑手，或者在健身房或运动队等团体训练环境中可能需要同时分析多名运动员。

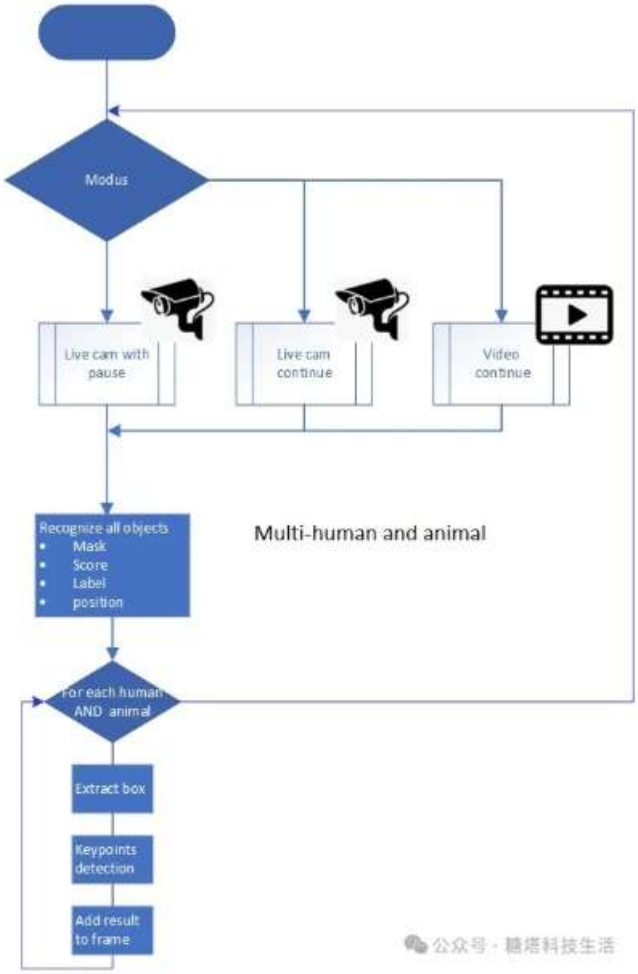
Another major limitation of most existing body pose detection systems is that they can only process a single person per frame. Many standard implementations only assume a single person and cannot scale to multiple people - or even worse, multiple species. This poses a huge challenge for our project, as we need to analyze both the horse and the rider, or perhaps multiple athletes simultaneously in a group training environment such as a gym or sports team.

为了解决这个问题，我们实施了一套两阶段流程。首先，我们使用专用的物体检测器来识别并定位画面中所有相关主体——包括人类和动物。然后，每个检测到的边界框会被单独传入姿态估计模型，该模型会计算出该特定主体的 25 个关键点。最后，这些关键点会被映射回原始画面，从而保留空间上下文，并实现对所有在场个体的同步分析。

To address this, we implemented a two-stage pipeline. First, we used a dedicated object detector to identify and localize all relevant subjects in the frame—both humans and animals. Each detected bounding box was then individually fed into a pose estimation model, which computed 25 keypoints for that specific subject. Finally, these keypoints were mapped back to the original frame, preserving the spatial context and enabling simultaneous analysis of all individuals present.

这种架构不仅能够同时评估骑手和马匹，还为将同一系统应用于更通用的场景开辟了道路——例如在健身房、团体运动或团队生物力学实验室中进行多用户运动追踪。通过将检测与姿势估计分离，我们构建了一个可扩展的解决方案，能够处理涉及多个互动身体的复杂真实场景。

This architecture not only enables simultaneous assessment of both rider and horse, but also opens the way to applying the same system to more general scenarios – such as multi-user motion tracking in a gym, team sports, or a team biomechanics lab. By decoupling detection from pose estimation, we built a scalable solution capable of handling complex real-world scenarios involving multiple interacting bodies.



结论 in conclusion

身体关键点检测已从一种小众研究工具发展成为一项用途广泛、功能强大的技术，对各行各业产生了深远的影响。虽然人体姿势估计已经改变了健身、机器人和增强现实等领域，但将这种能力扩展到动物——尤其是在赛马训练等高性能环境中——开辟了数据驱动洞察的全新前沿。

我们的工作旨在突破这些界限，开发一套适用于人类和动物的统一关键点检测系统。通过提升解剖粒度并构建多学科分析流程，我们创建了一个能够捕捉骑手与马匹之间复杂互动的系统，并适用于团体健身、运动分析或兽医护理等更广泛的用例。

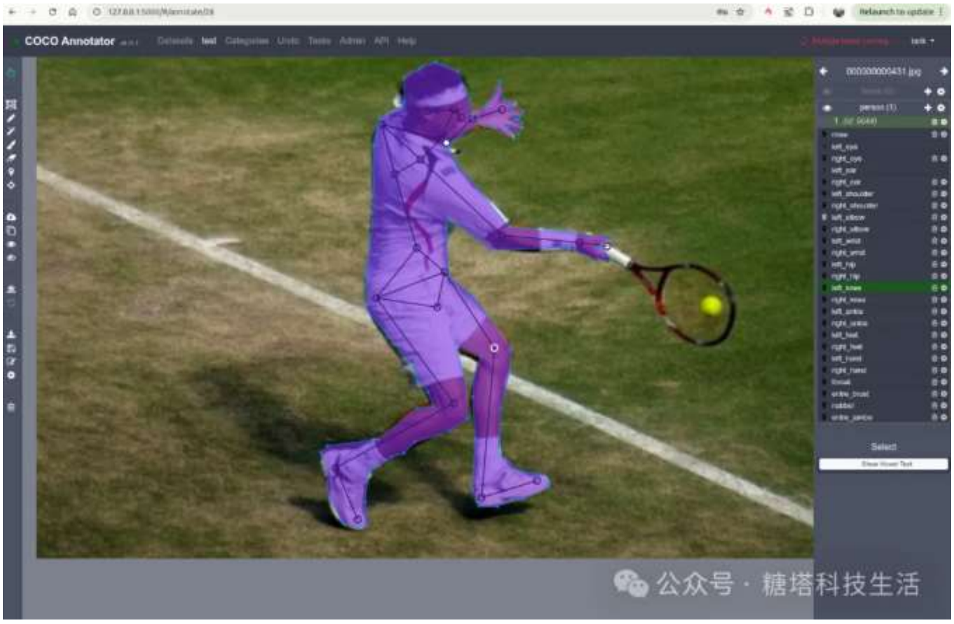
随着硬件效率的提升和人工智能模型的不断改进，我们预计关键点检测将成为性能诊断、人机协作，甚至行为感知自动化的默认工具。曾经局限于实验室环境的技术，如今已准备好逐步应用于现实世界，一步步、一步步地实现。

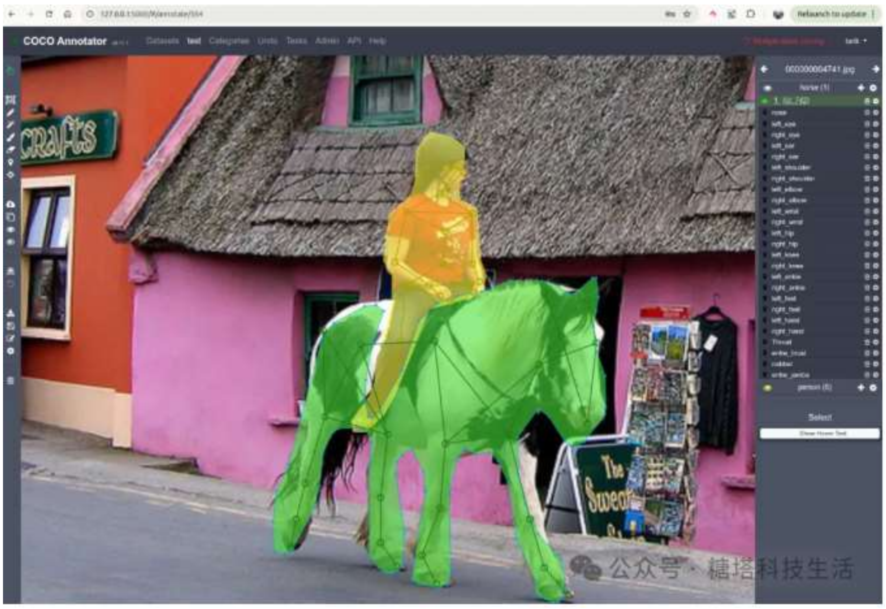
Body keypoints detection has evolved from a niche research tool to a versatile and powerful technology with far-reaching impacts across industries. While human pose estimation has transformed fields such as fitness, robotics, and augmented reality, extending this capability to animals—especially in high-performance environments such as racehorse training—opens up a whole new frontier of data-driven insights.

Our work aims to push these boundaries and develop a unified keypoints detection system for both humans and animals. By increasing anatomical granularity and building a multidisciplinary analysis pipeline, we created a system that captures the complex interactions between riders and horses and is applicable to a wider range of use cases such as group fitness, sports analysis, or veterinary care.

As hardware efficiency increases and AI models continue to improve, we expect keypoints detection to become the default tool for performance diagnostics, human-machine collaboration, and even behaviour perception automation. What was once limited to laboratory environments is now ready to be applied to the real-world step by step, step by step.

附件：coco-annotator GUI





如果您有兴趣投资北京糖塔科技，我们热烈欢迎天使投资者加入我们的创新型创业公司。请联系我们的首席执行官张东伟了解更多信息。

If you’re interested in investing in Beijing Tangta Technology, angel investors are warmly welcomed to join our innovative startup. Please contact our CEO, Zhang Dong Wei, for more information.

- 张东伟 MAIL: 1542370275@qq.com
- 塔里克 MAIL: 3559870647@qq.com 微信: tarikbeijing
- www.tangta-technology.com
- www.tangta-technology.com.cn

同时进行人体和动物姿态估计：一种增强性能洞察的多目标方法