

Recording and Hand Tracking Protocol with Artificial Intelligence and Two Cameras

Seguimiento con inteligencia artificial para el seguimiento de la mano de diferentes especies con dos cámaras

Alberto Lopez Moreno¹, Geovanni Hernández Gómez², Víctor De Lafuente Flores³

¹Universidad de Guanajuato, Departamento de Estudios Multidisciplinarios, División de Ingenierías, Campus Irapuato-Salamanca

³Universidad Nacional Autónoma de México, Blvd. Juriquilla 3001, Juriquilla, la mesa, Juriquilla, 76230, Qro, México

Resumen

El seguimiento preciso de manos, tanto en especies humanas como no humanas, es fundamental en campos como la biomecánica y neurología. La inteligencia artificial (IA) ha emergido como una herramienta clave para lograr un seguimiento sin marcadores con una precisión comparable a la de la anotación manual. En este estudio, empleamos un conjunto de herramientas basado en redes neuronales profundas para rastrear movimientos de manos definidos por el usuario con alta precisión. Al utilizar dos cámaras sincronizadas, obtuvimos datos desde diferentes ángulos, lo que permitió un análisis más detallado y robusto.

Palabras clave: Comportamiento, Redes neuronales profundas, GPU, Neurología.

Introducción

El seguimiento de manos, tanto en sujetos humanos como en especies no humanas, presenta desafíos importantes debido a la variabilidad en el tamaño, forma y color de las manos, así como a las posibles oclusiones causadas por el cuerpo u otros objetos [1]. Tradicionalmente, esta tarea requería el uso de marcadores físicos o sistemas ópticos avanzados, lo cual limitaba la precisión y capacidad de análisis, especialmente en entornos naturales o no controlados. Sin embargo, los avances recientes en inteligencia artificial (IA) han revolucionado este campo, permitiendo un seguimiento robusto y preciso sin la necesidad de marcadores, incluso en escenarios complejos y desafiantes [2].

El uso de redes neuronales profundas, como las redes convolucionales (CNN) y las redes residuales (ResNet), ha facilitado el desarrollo de modelos capaces de aprender patrones de movimiento complejos y de realizar seguimiento en tiempo real. Estos modelos, entrenados con grandes volúmenes de datos, son capaces de adaptarse a nuevas situaciones sin la necesidad de ajustes manuales específicos, ofreciendo una precisión comparable al etiquetado manual. La capacidad de la IA para procesar grandes cantidades de datos de video de manera eficiente hace que sea una herramienta ideal para tareas como el seguimiento de la mano, que requieren velocidad y exactitud en la interpretación de imágenes [3].

En este estudio, implementamos un sistema de seguimiento de manos basado en IA utilizando dos cámaras sincronizadas. Esta configuración de cámaras múltiples permite capturar el movimiento de las manos desde distintos ángulos, lo cual mejora notablemente la precisión y profundidad del análisis. Gracias a la IA, el sistema puede aprender a identificar puntos clave en la mano, como los dedos y la muñeca, y seguir sus trayectorias en tiempo real, incluso en situaciones donde las manos están parcialmente ocluidas o se mueven rápidamente. Además, el sistema se adapta a variaciones en condiciones de iluminación y a diferencias en la apariencia de las manos entre distintos individuos o especies [4].

El potencial del seguimiento de manos mediante IA es vasto y se extiende a diversas aplicaciones, como el estudio del comportamiento animal, programas de rehabilitación y la interacción entre animales y computadoras. Por ejemplo, en la investigación del comportamiento, el seguimiento de manos puede ser usado para estudiar actividades naturales de animales, como el movimiento de manos en primates durante interacciones sociales [5]. En rehabilitación, permite monitorizar la recuperación de la movilidad y la fuerza de las extremidades en animales tras una lesión. Incluso, en el ámbito de la interacción animal-computadora, podría facilitar la creación de experiencias interactivas para animales



mediante gestos con las manos [6]. La combinación de IA y cámaras sincronizadas ofrece así una solución escalable y poderosa para mejorar nuestra comprensión y cuidado de los animales [7].

Metodología Propuesta

Para implementar el seguimiento de manos con inteligencia artificial, utilizamos una serie de bibliotecas y plataformas de software especializadas en procesamiento de imágenes y aprendizaje automático. La programación se realizó en Python, utilizando bibliotecas como OpenCV, TensorFlow y PyTorch. El procesamiento se ejecutó en Google Colab, aprovechando sus capacidades de cómputo en la nube para optimizar el rendimiento. Además, el sistema fue configurado para operar con dos cámaras, lo que permitió obtener información desde dos ángulos simultáneos para un seguimiento más completo.

Tabla 1. Bibliotecas empleadas para el procesamiento de seguimiento

Biblioteca	Enfoque	Lenguaje	Usos principales	Plataforma
OpenCV	Visión computacional	C/C++	Procesamiento de imágenes	Multiplataforma
TensorFlow	Aprendizaje automático	Python	Redes neuronales	Multiplataforma
PyTorch	Aprendizaje automático	Python	Redes neuronales	Multiplataforma

Fuente: Elaboración propia.

Configuración de Cámaras

Para el seguimiento de las manos se utilizaron dos cámaras, colocadas estratégicamente para capturar los movimientos desde diferentes ángulos:

Cámara 1: Orientada hacia la mano derecha del participante, capturando movimientos en los ejes X e Y.

Cámara 2: Dirigida a la mano izquierda o desde un ángulo complementario para proporcionar una vista más completa.

Ambas cámaras fueron sincronizadas para grabar a la misma tasa de fotogramas por segundo (FPS), lo que permite analizar las grabaciones de manera conjunta. Se empleó un estímulo visual (una señal luminosa) al inicio de cada grabación para asegurar la sincronización entre los videos.

Implementación del Seguimiento con Redes Neuronales

Para el proceso de seguimiento, utilizamos una red neuronal ResNet50 entrenada específicamente para detectar y rastrear cinco puntos clave en las manos humanas (dedos, muñeca, etc.). Estos puntos se analizaron desde ambas cámaras, permitiendo generar trayectorias en tiempo real de los movimientos de las manos. Este enfoque nos permitió obtener una precisión robusta en la detección y seguimiento de las manos en diversos ángulos y condiciones [9].

En la imagen 1 mostramos las etapas de la metodología de entrenamiento de redes neuronales

· Etiquetado de Datos

El proceso de etiquetado es fundamental para entrenar modelos de inteligencia artificial en el seguimiento de manos, ya que permite identificar los puntos clave (como dedos y muñeca) que la red neuronal debe aprender a detectar. Se utilizó software especializado para realizar el etiquetado manual de los puntos de interés en las manos en un conjunto representativo de imágenes. Este proceso se realizó en ambas vistas de cámara, asegurando que los puntos clave estuvieran correctamente identificados en cada ángulo. Estos datos etiquetados sirvieron como referencia para el aprendizaje supervisado de la red neuronal [10].



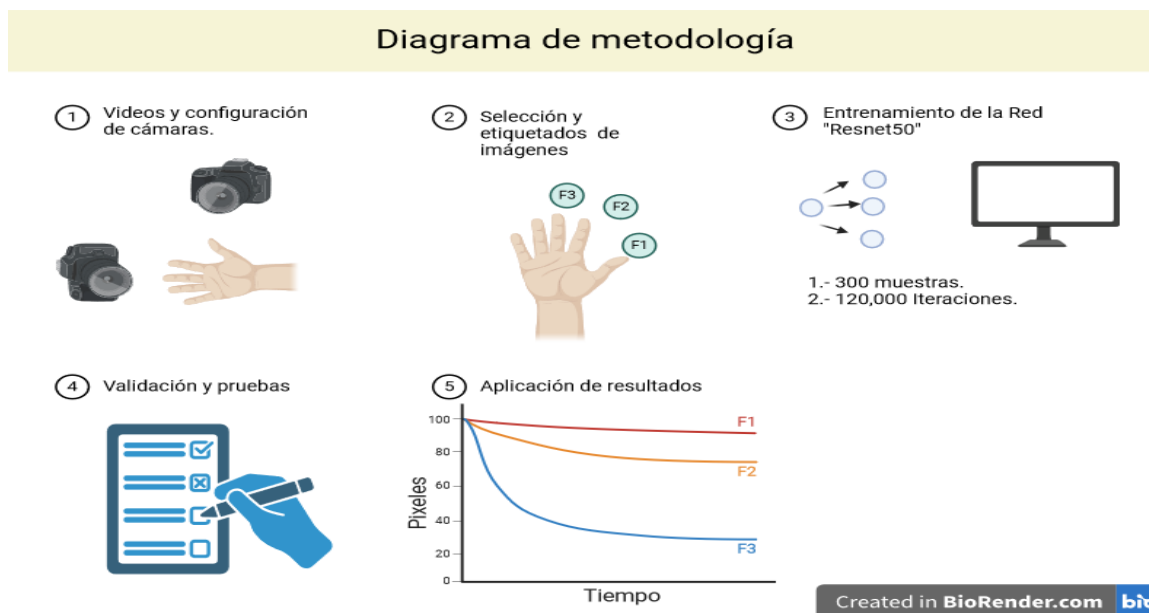


Imagen 1.- Metodología en redes neuronales, Fuente: Elaboración Propia Software: Biorender.

• Adquisición de Imágenes con Cámaras

Para capturar los movimientos de las manos, se configuró un sistema de cámaras sincronizadas. Se emplearon dos cámaras de alta resolución posicionadas para captar la mano desde distintos ángulos:

- Cámara 1: Enfocada principalmente en la mano derecha del sujeto, capturando movimientos sobre los ejes X e Y.
- Cámara 2: Enfocada en la mano izquierda o en un ángulo complementario para obtener una vista completa del movimiento.

Ambas cámaras se sincronizaron para grabar a una tasa constante de fotogramas, lo que facilitó la alineación de las imágenes en la fase de análisis y garantizó que los datos de ambas cámaras fueran consistentes. Para la sincronización, se usó un estímulo visual al inicio de la grabación, el cual permitió alinear los fotogramas en los análisis posteriores.

• Selección de Imágenes

Se seleccionó un conjunto representativo de imágenes para el entrenamiento y la validación del modelo. Este conjunto se eligió cuidadosamente para cubrir una amplia gama de posiciones de la mano, variaciones en ángulos y diferentes condiciones de iluminación, con el fin de aumentar la robustez del modelo. Las imágenes seleccionadas incluyeron tanto movimientos rápidos como lentos y en situaciones de oclusión parcial, para que el modelo fuera capaz de generalizar en situaciones desafiantes [11].

• Entrenamiento del Modelo

Para el entrenamiento, se utilizó una red neuronal profunda basada en ResNet50, que fue previamente configurada para detectar puntos específicos en las manos. Este modelo fue entrenado en Google Colab, donde aprovechamos la capacidad de procesamiento en la nube para manejar el gran volumen de datos. El entrenamiento se realizó utilizando los puntos etiquetados previamente como referencia, de manera que el modelo aprendiera a reconocer patrones de movimiento de la mano y a identificar los puntos clave en tiempo real. Durante esta fase, se implementaron técnicas de aumento de datos, como rotación y escalado de las imágenes, para mejorar la capacidad de generalización del modelo.

• Validación y Pruebas

Una vez finalizado el entrenamiento, el modelo se sometió a una fase de validación para evaluar su precisión en el reconocimiento de puntos de la mano en nuevas imágenes. Para ello, se utilizó un conjunto de datos de validación que

incluía situaciones similares y otras nuevas que no habían sido vistas durante el entrenamiento, lo cual permitió medir la capacidad del modelo para adaptarse a diferentes escenarios. Se analizaron métricas de precisión, sensibilidad y especificidad en la detección de los puntos clave de la mano.

La validación se completó con pruebas en tiempo real, en las que se verificó que el modelo podía seguir el movimiento de las manos de manera continua y precisa, en condiciones similares a las que enfrentaría en aplicaciones prácticas. Además, se realizaron ajustes en el modelo para optimizar su rendimiento en situaciones específicas, como la detección en condiciones de baja iluminación o con oclusiones parciales.

Resultados

Los resultados obtenidos en este estudio demuestran que el sistema de seguimiento de manos basado en inteligencia artificial funcionó de manera eficaz con ambos videos capturados por las dos cámaras. La red neuronal ResNet50 logró realizar un seguimiento preciso y continuo de los puntos clave de la mano en ambas perspectivas, lo que permitió reconstruir trayectorias de movimiento en tiempo real.

Durante el análisis, se observó que el modelo fue capaz de identificar y rastrear consistentemente los puntos de interés (dedos y muñeca) a lo largo de toda la secuencia de video, incluso en condiciones de oclusión parcial y con cambios de iluminación. El uso de dos ángulos de cámara proporcionó una visión complementaria que mejoró la precisión del seguimiento, ya que cualquier pérdida temporal de visibilidad en una cámara fue compensada por la otra.

Estos resultados indican que el sistema es robusto y versátil, y puede ser aplicado en estudios de seguimiento de manos en contextos dinámicos. La precisión obtenida con la red neuronal en ambas cámaras respalda la viabilidad de este enfoque para investigaciones donde la captura de movimiento detallada es esencial, ofreciendo una solución efectiva frente a los desafíos tradicionales del seguimiento manual.

Se realizó un seguimiento exitoso de las imágenes obtenidas desde las dos perspectivas de las cámaras (lateral y frontal). Utilizando la red neuronal profunda implementada, se logró identificar y rastrear los movimientos de la mano en ambas vistas con gran precisión. En las imágenes capturadas desde la vista lateral, el sistema pudo rastrear los movimientos a lo largo de los ejes X y Y, proporcionando datos detallados sobre el desplazamiento lateral y el ángulo de movimiento. En la vista frontal, el sistema detectó los movimientos en profundidad, logrando un seguimiento continuo de los puntos clave en la mano y los dedos como se ve en la **imagen 2,3,4**.

La combinación de estas dos perspectivas permitió reconstruir de manera precisa la trayectoria tridimensional de la mano, con cada cámara complementando la información de la otra. Este enfoque no solo mejoró la robustez del sistema en condiciones variables de iluminación y oclusión, sino que también aportó una visión completa de la dinámica de movimiento en el espacio, validando la efectividad del método implementado para aplicaciones en el análisis de comportamiento y rehabilitación.



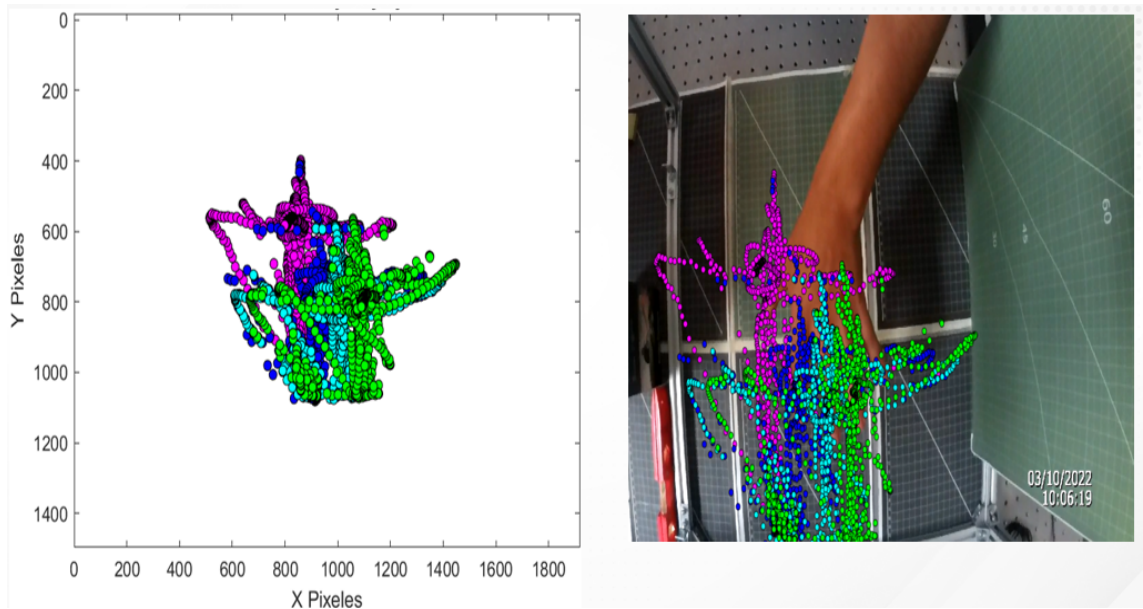


Imagen 2.- Vista área de la imagen con redes neuronales, Fuente: Elaboración Propia.

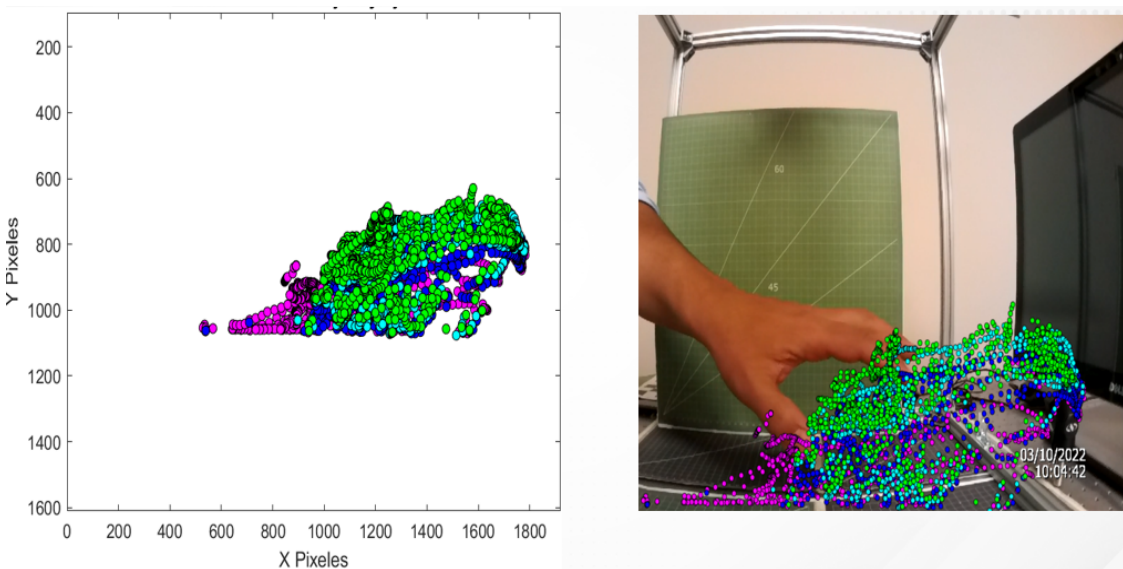


Imagen 3.- Vista lateral de la imagen con redes neuronales, Fuente: Elaboración Propia.

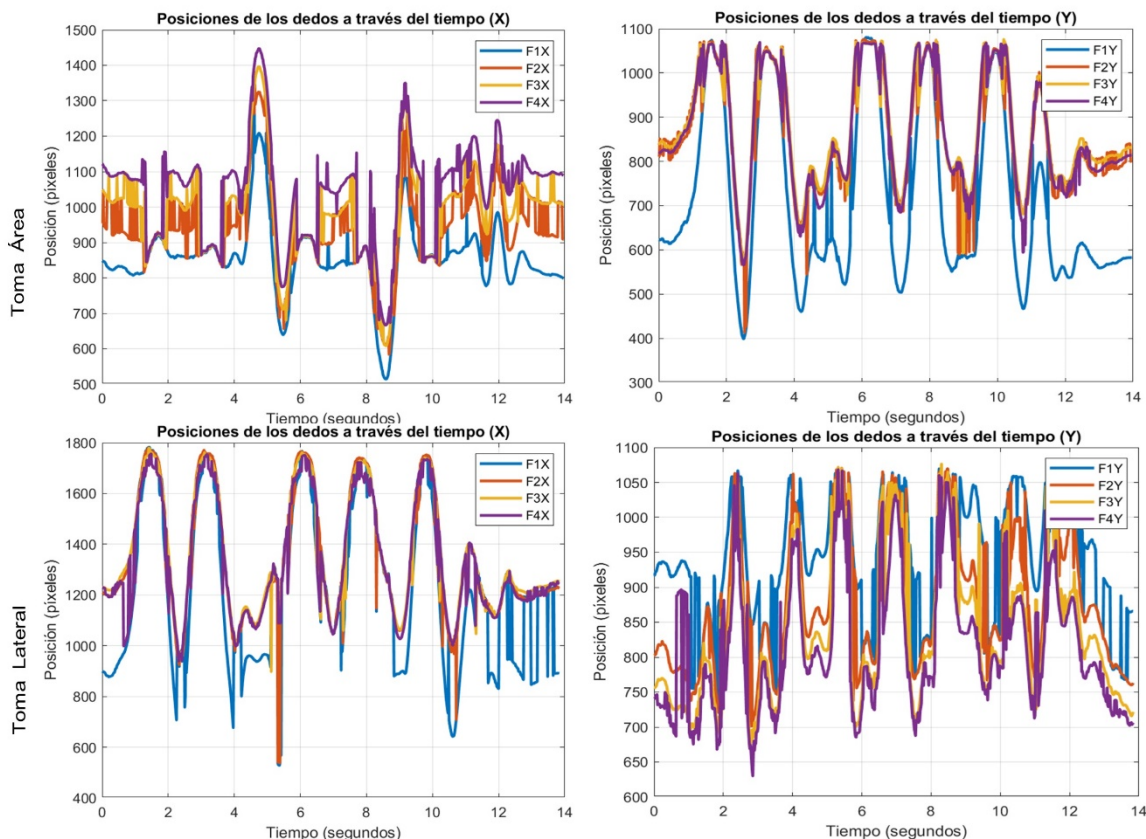


Imagen 4.- Vista lateral de la imagen con redes neuronales, Fuente: Elaboración Propia.

Discusión

Los resultados obtenidos de los experimentos ofrecen información valiosa sobre el rendimiento del sistema de seguimiento para manos de humanos y primates no humanos. A continuación, se analizan algunos aspectos clave:

• Precisión del seguimiento de manos humanas:

El experimento que involucró el movimiento de la mano humana a lo largo de los ejes X e Y evidencia la capacidad del sistema.

• Seguimiento de manos de primates no humanos:

El seguimiento de la mano de un primate no humano durante tareas repetitivas añade una dimensión interesante.

• Uso de herramientas y plataformas:

El uso de Google Colab para realizar los experimentos introduce un elemento interesante.

• Visualización de los resultados:

Las figuras presentadas, como la posición en los ejes X e Y, ofrecen una representación visual de los resultados. ¿Cómo facilita esta visualización la interpretación de los datos por parte del usuario? La representación gráfica de la posición de cada parte de la mano en relación con el cuadro de video es una adición valiosa. ¿Cómo puede aprovecharse esta información para análisis más detallados?

Resumen y conclusiones

En este estudio, hemos explorado la aplicación de técnicas avanzadas de seguimiento para monitorear las manos tanto de humanos como de primates no humanos. Los hallazgos clave y sus implicaciones se resumen a continuación:

• Logros en el seguimiento de manos humanas:

Los experimentos de seguimiento de manos humanas utilizando una red neuronal ResNet50 han demostrado una notable precisión. Esto no solo valida la efectividad del algoritmo elegido, sino que también sugiere su potencial para diversas aplicaciones en biomecánica y neurología.

• Perspectivas del seguimiento de manos en primates no humanos:

El seguimiento de manos de primates no humanos durante tareas repetitivas ha proporcionado información valiosa sobre sus comportamientos. Estos datos podrían ser fundamentales para comprender la cognición de los primates y sus respuestas a estímulos externos.

• Uso de herramientas y plataformas:

La integración de herramientas como OpenCV, TensorFlow y PyTorch dentro de la plataforma Google Colab ha sido exitosa. Esto no solo mejora la eficiencia de los experimentos, sino que también proporciona una base sólida para el desarrollo futuro.

• Representación visual de los resultados:

La representación gráfica de las posiciones de las manos en los ejes X e Y, así como la visualización detallada de las partes individuales de la mano, ofrece una comprensión intuitiva de los resultados del seguimiento. Estos recursos visuales son cruciales para investigadores y profesionales en diversos campos.

• Aplicaciones potenciales y futuras direcciones:

Las aplicaciones discutidas, que incluyen la investigación en comportamiento animal, la evaluación en rehabilitación y la interacción animal-computadora, muestran la versatilidad del sistema de seguimiento. Trabajos futuros podrían explorar aplicaciones adicionales y refinamientos, especialmente en el contexto de otras especies animales.

En conclusión, el sistema de seguimiento presentado no solo cumple, sino que supera las expectativas en la monitorización de movimientos de manos a través de diferentes especies. La combinación de algoritmos avanzados, potentes bibliotecas y plataformas de uso accesible sienta las bases para futuros avances en el campo en expansión del seguimiento del comportamiento animal.

Referencias

- Mann, K.S., Schneider, S., Chiappa, A., Lee, J.H., Bethge, M., Mathis, A., Mathis, M.W. (2021). Out-distribution generalization of internal models is correlated with reward, in: Self-Supervision for Reinforcement Learning Workshop-ICLR.
- Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., Bethge, M. (2018). Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* 21, 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- Mathis, A., Schneider, S., Lauer, J., Mathis, M.W. (2020). A primer on motion capture with deep learning: principles, pitfalls, and perspectives. *Neuron* 108, 44–65. <https://doi.org/10.1016/j.neuron.2020.09.017>
- Mathis, M.W., Schneider, S. (2021). Motor control: Neural correlates of optimal feedback control theory. *Current Biology* 31, R356–R358. <https://doi.org/10.1016/j.cub.2021.01.087>
- Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., Mathis, M.W. (2019). Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature protocols* 14, 2152–2176. <https://doi.org/10.1038/s41596-019-0176-0>



- Theckedath, D., Sedamkar, R., (2020). Detecting affect states using vgg16, resnet50 and se-resnet50 networks. SN Computer Science 1, 1–7. <https://doi.org/10.1007/s42979-020-0114-9>
- Vyas S, Golub MD, Sussillo D, Shenoy KV. Computation Through Neural Population Dynamics. Annu Rev Neurosci. (2020).doi: . <https://doi.org/10.1146/annurev-neuro-092619-094115>.
- Orger MB, de Polavieja GG. Zebrafish Behavior: Opportunities and Challenges. Annu Rev Neurosci. (2017) Jul 25;40:125-147. doi: 10.1146/annurev-neuro-071714-033857.
- Hernández-Pérez R, Cuaya LV, Rojas-Hortelano E, Reyes-Aguilar A, Concha L, de Lafuente V. Tactile object categories can be decoded from the parietal and lateral-occipital cortices. Neuroscience. (2017) Jun 3;352:226-235. doi: 10.1016/j.neuroscience.2017.03.038.
- De Lafuente V. Regaining the senses of touch and movement. Elife. (2018) 10-7. doi: 10.7554/eLife.36137.

