

Estado de la publicación: El preprint no ha sido enviado para publicación

# Detección automática de posición corporal en animales de granja

John Fredy Ramirez Agudelo, Jose Fernando Guarín Montoya, Sebastian Bedoya Mazo

<https://doi.org/10.1590/SciELOPreprints.3705>

Enviado en: 2022-03-02

Postado en: 2022-03-08 (versión 1)

(AAAA-MM-DD)

# **Detección automática de posición corporal en animales de granja.**

Automatic pose detection in farm animals.

***John Fredy Ramírez Agudelo***  
***José Fernando Guarín Montoya***  
***Sebastián Bedoya Mazo***

*Grupo de Investigación en Ciencias Agrarias (GRICA), Facultad de Ciencias Agrarias,  
Universidad de Antioquia, Medellín, Colombia.*

*johnf.ramirez@udea.edu.co*  
*fernando.guarin@udea.edu.co*  
*sebastian.bedoyam@udea.edu.co*

## **RESUMEN**

- **Contextualización:** Los animales utilizan una amplia gama de variaciones en sus poses corporales que puede ser interpretada como información sobre su estado de salud, bienestar o actividad individual. Sin embargo, la observación directa de estas posiciones es una tarea que demanda tiempo y es económicamente inviable en las empresas pecuarias. Gracias a las técnicas de visión por computador es posible que la implementación de sistemas de observación automática en las granjas sea una alternativa viable.
- **Vacío de conocimiento:** Actualmente no existe un modelo de estimación de pose de uso exclusivo con animales de granja que permita el desarrollo de sistemas de detección automática de posturas.
- **Propósito:** El objetivo de este trabajo fue evaluar el desempeño de un modelo de redes neuronales re-entrenado para la detección de poses en algunas especies de rumiantes y en caballos.

- **Metodología:** De la base de datos Imagenet, se descargaron más de diez mil imágenes de rumiantes y equinos. De estas imágenes, se seleccionaron 2000 de vacunos y 591 de otras especies para el re-entrenamiento y evaluación del modelo, respectivamente. Estas imágenes fueron etiquetadas con el programa COCO Annotator. Este proceso consistió en la identificación manual de ocho puntos claves de la anatomía de los animales en cada imagen. El proceso de re-entrenamiento fue realizado con la librería detectron2 en Python. Para cuantificar la precisión del modelo se utilizó la similitud de puntos claves de objeto.
- **Resultados y conclusiones:** El índice de similitud de puntos claves de objeto permitió establecer que el aprendizaje desarrollado por el modelo para identificar puntos clave en vacunos puede ser utilizado para la misma tarea en otros animales de granja. Caballos y búfalos presentaron los mejores resultados de detección. En conclusión, un conjunto de datos relativamente pequeño de la posición en animales permite evaluar la generalización de la inferencia de los modelos dentro (vacunos) y fuera del dominio (otros rumiantes y equinos). Este tipo trabajos sirven como línea base para el desarrollo de sistemas de monitoreo automático de animales de granja.

**Palabras clave:** Aprendizaje profundo, Ganadería de precisión, Automatización.

## ABSTRACT

**Contextualization:** Animals use a wide range of variations on their body poses that can be interpreted as information about their health, welfare, or activity level. However, direct observation of these poses is a great time-consuming activity and economically unfeasible task for farmers. Although, with the use of computer vision techniques, it is possible to implement automatic observation systems on farms making this complicated and costly activity a viable alternative.

**Knowledge gap:** Currently, there is a lack of a pose estimation model exclusively for farm animals, which allows the development of automatic posture detection systems.

**Purpose:** The objective of this work was to evaluate the performance of a re-trained neural network model for the detection of poses in some ruminant species and horses.

**Methodology:** More than ten thousand images of ruminants and horses were downloaded from the Imagenet database. From these images, 2000 cattle and 591 other species were selected for re-training and evaluation of the model, respectively. These images were labelled with the COCO Annotator software. This process consisted of the manual identification of

eight key points on the animals' anatomy in each image. The retraining process was carried out with the detectron2 library in Python. Object Keypoint Similarity was used to quantify the precision of the model.

**Results and conclusions:** The Object Keypoint Similarity index established that the learning developed by the model to identify key points in cattle can be used for the same task in other farm animals. Horses and buffaloes had the best detection results. In conclusion, a relatively small data set of position in animals allows evaluating the generalizability of the inference of the models within (cattle) and outside the domain (other ruminants and equines). This type of work serves as a baseline for the development of automatic farm animal monitoring systems.

**Keywords:** Deep Learning, Precision Livestock, Automation.

## 1. INTRODUCCIÓN

Recientemente, los científicos han comenzado a utilizar nuevas herramientas de aprendizaje automático llamadas redes neuronales profundas (LeCun et al. 2015) para estimar el comportamiento de los animales. Estas herramientas aprenden a determinar la posición de las partes del cuerpo de un animal en el espacio, a partir de imágenes o videos. Esto permite que los algoritmos de aprendizaje profundo determinen automáticamente el comportamiento de los animales de manera más rápida y precisa que técnicas anteriores (e.g. acelerómetros). Esta capacidad de aprender de las imágenes también elimina la necesidad de colocar marcadores físicos en los animales, lo que puede alterar su comportamiento natural (Graving et al. 2019).

La estimación de pose o posición corporal se refiere a la identificación de una serie de puntos clave del cuerpo. Esta es una tarea fundamental de visión por computadora, que es útil para el reconocimiento de actividades (Boualia y Amara, 2019), comprensión del comportamiento (Marginean et al. 2019) y la interacción con objetos (Wu et al. 2020) en humanos. Basados en los trabajos con humanos, la estimación de la postura de los animales es un campo importante que ha recibido una atención creciente en los últimos años y que puede ser utilizada para fines de investigación, productivos, así como para aplicaciones prácticas como la zoología y la conservación de la vida silvestre.

Las redes neuronales, una clase de algoritmos de aprendizaje automático que aprenden relaciones estadísticas complejas a partir de datos (Mathis et al. 2018, Pereira et al. 2019), son herramientas matemáticas muy eficaces para la estimación de poses. Esta tecnología ha sido ampliamente utilizada en la detección de poses en humanos (Insafutdinov et al. 2016). La estimación de pose en humanos se ha realizado con éxito en conjuntos de datos a gran escala (Fang et al. 2017, Cao et al. 2017, Chen et al. 2018). Existen algunos modelos que se enfocan en la pose de partes específicas del cuerpo, como las manos (Yang y Yao, 2019, Kim et al. 2021) y la cara (Albiero et al. 2021). En animales, Mathis et al. (2020) proporcionan una discusión sobre las técnicas y tecnologías disponibles para la captura de movimiento de las posiciones de los animales sin la necesidad de marcar a los sujetos, donde el principal desafío para esta tarea es la falta de datos etiquetados.

En los últimos años se han hecho públicos varios conjuntos de datos de etiquetados para la determinación de pose en animales. Cao et al. (2019) proporcionaron alrededor de cuatro mil imágenes de varios cuadrúpedos domésticos. El conjunto de datos MacaquePose proporciona más de 13 mil imágenes anotadas de macacos (Labuguen et al. 2021). Biggs et al. (2020) proporcionó 20 mil posturas de perro 2D basadas en el conjunto de datos de perros de Stanford (Khosla et al. 2011). AnimalWeb proporciona 21 mil imágenes de caras (Khan et al. 2020). El conjunto de datos Horse10 comprende ocho mil imágenes de caballos (Mathis et al. 2021). OpenMonkeyStudio proporciona un excelente conjunto de datos de 195 mil fotogramas de primates con varias cámaras en condiciones de laboratorio (Bala et al. 2020). Además, existen bases de datos de cebras (Graving et al. 2019) y tigres (Li et al. 2019).

El principal problema en la mayoría de los trabajos anteriormente señalados es que estos se enfocan solo en algunas especies animales, ignorando la diversidad en la naturaleza, lo que limita la capacidad de generalización de estos modelos. Sin embargo, gracias a que los humanos comparten similitudes de esqueletos con muchos animales, principalmente cuadrúpedos, es posible que los modelos entrenados para la detección de pose humana proporcionen un conocimiento previo importante para mejorar el rendimiento en la estimación de la postura de los animales, ya que esos modelos han sido entrenados con una escala mucho mayor de datos. El objetivo de este trabajo fue abordar la detección de pose en animales de granja como un problema de adaptación de dominio en el que un modelo puede aprender a estimar la pose los animales nunca antes vistos.

## **2. MATERIALES Y MÉTODOS**

### **Selección de imágenes**

Cerca de seis mil imágenes de vacunos fueron descargadas de la base de datos Imagenet (Deng, et al. 2009). Imagenet es una base de datos de más de 14 millones de imágenes organizada en jerarquía, en la que cada nodo de la jerarquía está representado por cientos o miles de imágenes. Este proyecto ha sido fundamental en el avance de la visión por computadora y la investigación de aprendizaje profundo. Los datos están disponibles de forma gratuita para los investigadores para uso no comercial (image-net.org). De estas seis mil imágenes se seleccionaron dos mil imágenes que cumplen con dos condiciones: a) al menos un animal, b) cuerpo completo. Estas dos mil imágenes fueron utilizadas para el re-entrenamiento del modelo rcnn\_R\_50\_FPN\_3x.

Adicionalmente se descargaron más de ocho mil imágenes de otros rumiantes y equinos: búfalo, yak, cabra, oveja, alpaca, camello, bisonte americano, ciervo y caballo. De igual manera, se seleccionaron imágenes de cada especie que cumplen con las dos condiciones señaladas para vacunos. Estas imágenes fueron utilizadas para comprobar si la capacidad del modelo para identificación puntos claves en vacunos puede ser utilizado para esta misma tarea en otras especies. En la Figura 1 se presentan algunos ejemplos de las imágenes seleccionadas.

### **Etiquetado de imágenes**

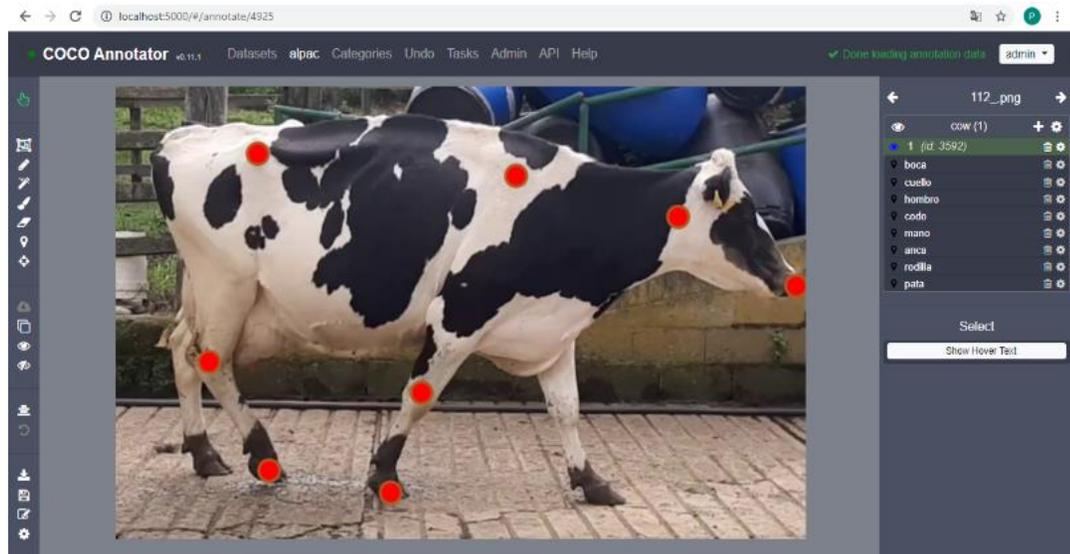
En "Machine Learning", el etiquetado es el proceso manual de identificación de los datos sin procesar (imágenes, archivos de texto, videos, etc.), donde se agregan una o más etiquetas informativas que proporcionen contexto a los datos para que un modelo de aprendizaje automático pueda aprender del trabajo realizado por un humano. Al construir un sistema de visión por computadora, primero se deben etiquetar las imágenes, mediante la identificación de píxeles o puntos clave, o de bordes que encierran el área de interés dentro de la imagen, para generar su conjunto de datos de entrenamiento.

Todas las imágenes, tanto las de vacunos como de otros rumiantes y equinos fueron etiquetadas con el software COCO Annotator (Brooks, 2019). Como se presenta en la Figura

2, en cada imagen se identificaron ocho puntos claves: boca, cuello, hombro, codo, mano, cadera, rodilla y pata. Se seleccionó este número de puntos ya que la vista lateral del animal puede ofrecer información suficiente para estimar su posición corporal.



**Figura 1.** Ejemplos de las imágenes seleccionadas.



**Figura 2.** Interface gráfica de COCO Annotator.

El etiquetado del conjunto de datos de imágenes es la parte que consume la mayor cantidad de tiempo en todo el proceso de entrenamiento de un modelo de aprendizaje automático. Como alternativa Pereira et al. (2019) propuso usar el aprendizaje activo, donde un modelo se entrena con una pequeña cantidad de imágenes anotadas, luego este modelo se usa para inicializar las etiquetas en un conjunto más grande de datos no anotados. Estos datos preinicializados son luego corregidos manualmente por el anotador humano. El modelo se vuelve a entrenar con este nuevo conjunto de datos. El anotador aplica este proceso de forma interactiva a medida que el conjunto de entrenamiento crece hasta proporcionar sólo ajustes menores a los datos preinicializados. Este proceso de anotación de estilo "humano en el bucle" fue utilizado en este trabajo para acelerar el proceso de etiquetado de imágenes. Las dos mil imágenes de vacunos fueron divididas en tres grupos de 100, 500 y 1400 imágenes, es decir que el proceso de etiquetado se llevó a cabo en tres etapas. En la primera etapa se etiquetaron manualmente 100 imágenes; en las siguientes dos etapas, el trabajo del anotador consistió en corregir las detecciones realizadas por el modelo entrenado en la primera etapa. Las imágenes de otros rumiantes y equinos fueron etiquetadas por un anotador humano sin la utilización del etiquetado asistido por el modelo. La Figura 2 presenta la interface gráfica de COCO Annotator (Brooks, 2019) para el proceso de etiquetado manual.

Durante el proceso de etiquetado se pudo constatar que las imágenes varían en el color del animal, el fondo, las condiciones de iluminación y el tamaño relativo del animal dentro de la imagen. La variación de la luz solar se sumó a la complejidad del desafío de aprendizaje, además de que algunos animales aparecen usando accesorios como arneses o acompañados de personas. Algunos animales estaban bajo la luz solar directa, mientras que otros tenían la luz detrás de ellos, y otros estaban bajo sombra. En algunas imágenes las patas fueron excluidas del etiquetado ya que la hierba las ocultó; pero se garantizó que, en todas las imágenes, se pudieran identificar los puntos desde los tobillos hasta la cabeza del animal.

Dependiendo de la variabilidad de los datos, las redes neuronales generalmente requieren miles o decenas de miles de ejemplos anotados manualmente para alcanzar una precisión a nivel humano. Sin embargo, teniendo en cuenta la similitud anatómica entre las especies de animales seleccionados, el tamaño del animal dentro de la imagen y la similitud en los entornos donde las imágenes fueron tomadas, consideramos que es posible minimizar la cantidad de ejemplos de entrenamiento necesarios (2000 imágenes) para lograr predicciones precisas. Mathis et al. (2018) y Pereira et al. (2019) han demostrado que se puede lograr

una precisión cercana al nivel humano con pocos ejemplos de entrenamiento. Parece que 100 ejemplos de entrenamiento es un tamaño mínimo razonable para un grupo de entrenamiento, aunque el número exacto probablemente cambiará dependiendo de la variación de los datos de la imagen que se anotan (Graving et al. 2019).

## **Entrenamiento del modelo**

Los métodos de aprendizaje automático de última generación, como las redes neuronales, utilizan los datos etiquetados por un humano para el proceso de entrenamiento de un modelo. Este entrenamiento consiste en parametrizar un modelo que describe las relaciones estadísticas entre un conjunto de datos de entrada (imágenes etiquetadas) y la distribución de salida deseada (puntos clave). Después de un entrenamiento adecuado, se puede usar un modelo para hacer predicciones sobre datos nunca antes vistos, datos que no formaban parte del conjunto de entrenamiento, lo que se conoce como inferencia.

Específicamente, utilizamos el modelo rcnn\_R\_50\_FPN\_3x, el cual es un modelo de tipo ResNets previamente entrenados para la detección de puntos claves en humanos. La popular arquitectura ResNet (He et al. 2016) es un modelo de aprendizaje profundo de última generación utilizado para la clasificación de imágenes. Elegir este modelo es ventajoso porque se trata de una arquitectura previamente entrenada, lo cual mejora el rendimiento del re-entrenamiento y reduce la cantidad de datos requeridos para este último (Mathis et al. 2018). Este proceso es conocido como transferencia de aprendizaje (Pratt, 1993). La transferencia de aprendizaje se refiere a transferir el conocimiento aprendido de una tarea de origen a una tarea de destino, las cuales están relacionadas de algún modo. Por lo general, el dominio de origen tiene una gran escala de datos etiquetados, mientras que solo hay un número limitado de datos etiquetados en el dominio de destino.

Todo el proceso de entrenamiento se realizó con la librería Detectron2 en Python. Detectron2 es la biblioteca de Facebook AI Research que proporciona algoritmos de detección y segmentación de imágenes de última generación. Es el sucesor de Detectron y maskrcnn-benchmark. Es compatible con una serie de proyectos de investigación de visión por computadora y aplicaciones de producción en Facebook. ([github.com/facebookresearch/detectron2](https://github.com/facebookresearch/detectron2)).

## Evaluación del entrenamiento del modelo

Los modelos de aprendizaje automático generalmente se evalúan por su capacidad para generalizar a nuevos datos, lo que se conoce como rendimiento predictivo, usando una muestra de datos anotados que no se usaron para el entrenamiento o validación. Para evaluar dicha capacidad, se utilizaron las imágenes de las clases de animales invisibles (otros rumiantes y equinos). Aquí "Invisible" significa que las muestras de animales para la prueba provienen de dominios que no participaron en el entrenamiento.

Para comparar las anotaciones humanas con las predicciones del modelo, utilizamos la similitud de puntos claves de objeto (Object Keypoint Similarity, OKS), la cual es la métrica de uso común en el desafío de puntos clave de COCO (Ruggero y Perona, 2017) y otros autores (Wei, et al. 2020). Esta métrica oscila entre 0 y 1. Cuanto más se acerque el punto clave predicho del punto clave verdadero, más se acercará el OKS a 1. Para calcular esta métrica se utiliza la siguiente ecuación.

$$\frac{\sum_i \exp(-d_i^2/2s^2k_i^2) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$

Dónde,

$d_i$  es la distancia euclidiana entre el punto clave etiquetado por un humano y el punto clave predicho por el modelo

$s$  es la raíz cuadrada del área de la imagen donde se detectó el animal

$k$  es la constante para cada punto clave.

$v_i$  se considera una bandera de visibilidad que puede ser 0, 1 o 2 para no etiquetado, etiquetado no visible, y etiquetado visible, respectivamente.

El término  $s$  se refiere a la escala de un objeto, simplemente es la raíz cuadrada del área del objeto. Cuanto más grande sea el objeto, menor debe ser la penalización, en otras palabras, mejor será el OKS. Está bien predecir un punto clave ligeramente alejado del punto clave verdadero, si el objeto es grande. Sin embargo, si el objeto es pequeño, una ligera desviación

podría hacer que el punto clave predicho se encuentre por fuera del animal. La ecuación hace que tales casos sean severamente penalizados.

El término  $k$  es un factor constante para cada punto clave y permanece igual para todas las muestras.  $k$  es una medida de la desviación estándar de un punto clave en particular. Esencialmente, el valor de  $k$  para los puntos clave más pequeños (ejemplo boca, pata) tiene una desviación estándar relativamente menor que los puntos claves de más tamaño (ejemplo hombro, caderas). Basado en los valores de  $k$  utilizados en la inferencia de puntos claves en humanos (Ruggero y Perona, 2017), se utilizaron los siguientes valores de  $k$ . Para boca, codo, mano, rodilla y pata se utilizó un  $k=0.35$ , para cuello  $k=0.5$ , y para hombro y anca  $k=1.0$ .

Otra técnica ampliamente utilizada en el análisis del desempeño de este tipo de modelos es la Matriz de Confusión. En esta matriz cada columna representa el número de predicciones de cada clase, y cada fila representa las instancias reales clasificadas por un humano. Cuando un modelo tiene un alto grado de validez, la diagonal de la matriz contiene la mayor parte de los resultados. Sin embargo, para el análisis de puntos claves esta técnica no resulta apropiada, ya que en teoría cualquier parte del animal puede ser erróneamente identificada por el modelo como un punto clave. Es decir que el modelo no confunde la predicción, simplemente no puede hacerla (e.g. Imagen G en Figura 4). De ahí, que esta técnica no fue utilizada en este trabajo. Como alternativa, se determinó la Exactitud en las estimaciones (Sokolova y Lapalme, 2009), lo cual puede ser interpretada como la probabilidad de que cada punto clave sea correctamente estimado por el modelo. Para cada punto clave se calculó la Exactitud mediante la proporción de resultados acertados por especie dividido entre el número de puntos esperados en todas las imágenes de dicha especie. Para la validación del modelo se utilizaron 591 imágenes.

### **3. RESULTADOS Y DISCUSIÓN**

Los resultados obtenidos en este trabajo demuestran que el enfoque de transferencia de conocimiento entre dominios resuelve eficazmente el problema de estimación de la pose de animales de granja de distintas especies (Figura 3). Específicamente, se logró un OKS promedio superior a 0.6 (Tabla 1) en el conjunto de prueba (otros rumiantes y equinos). El cual es un nivel de precisión similar al obtenido en la estimación de pose humana (Wei et al.

2020). Y lo que es más importante, brinda resultados prometedores en la estimación de la pose de otras especies animales sin ningún dato etiquetado.

En la Tabla 1 se puede ver que nuestro enfoque puede generalizarse bien a categorías de animales invisibles, como caballos y búfalos. Sin embargo, y a pesar de la cercanía filogenética con los vacunos, el rendimiento del modelo para yaks no es tan bueno como habría de esperarse. Esto se puede atribuir principalmente al pelaje largo del yak.

**Tabla 1.** Índices de similitud para las especies evaluadas

| <b>Especie</b> | <b>Número de imágenes</b> | <b>OKS promedio (SD)</b> |
|----------------|---------------------------|--------------------------|
| Caballo        | 75                        | 0.86 (0.10)              |
| Búfalo         | 59                        | 0.84 (0.12)              |
| Ciervo         | 71                        | 0.75 (0.16)              |
| Bisonte        | 71                        | 0.74 (0.15)              |
| Cabra          | 88                        | 0.73 (0.17)              |
| Oveja          | 46                        | 0.72 (0.18)              |
| Alpaca         | 63                        | 0.71 (0.15)              |
| Camello        | 51                        | 0.68 (0.16)              |
| Yak            | 67                        | 0.63 (0.22)              |

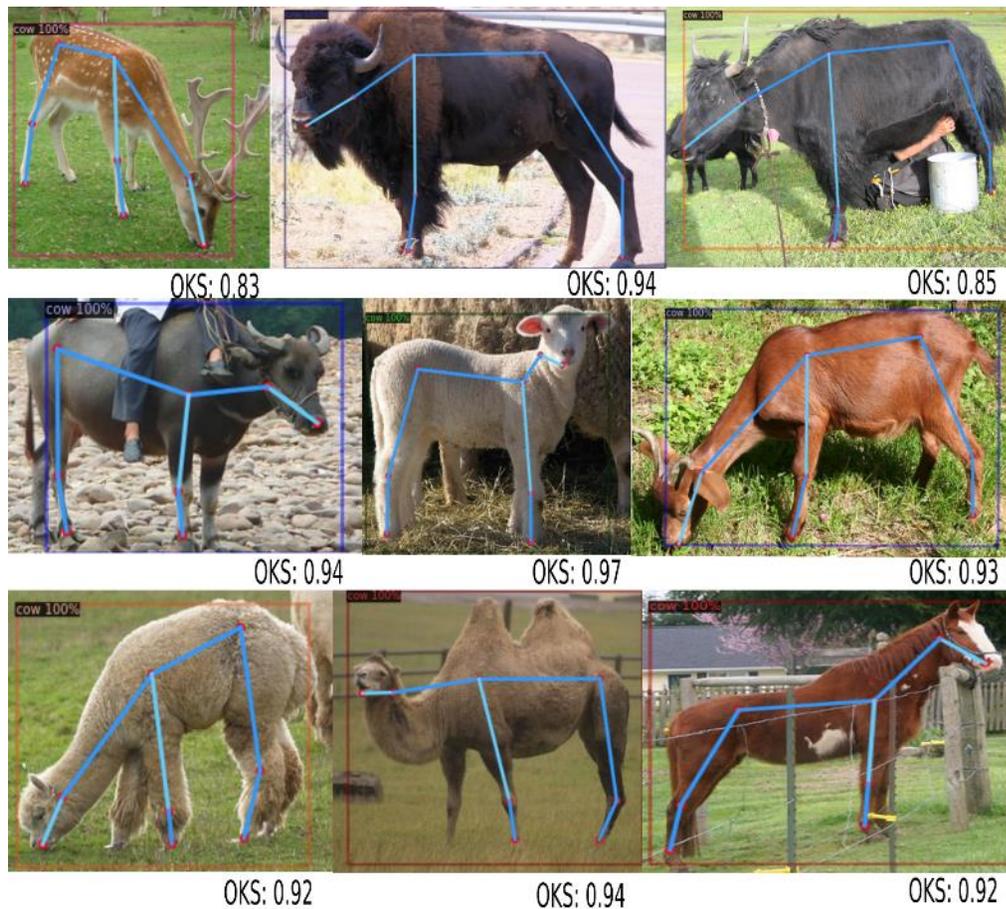
En la Tabla 2 se presentan los porcentajes de Exactitud de cada punto clave por especies. La menor Exactitud en la predicción del punto clave "Boca" para Alpacas y Camellos puede ser atribuida a la diferencia en longitud del cuello entre estas dos especies y la especie utilizada en el entrenamiento del modelo (vacunos). Excepto para Caballos y Búfalos, los puntos clave de las extremidades anteriores y posteriores (codo, mano, rodilla y pata) tienen menores porcentajes de Exactitud debido a que el modelo estima correctamente el punto clave, pero lo hace en la extremidad equivocada (e.g. Imagen A en Figura 4).

**Tabla 2.** Porcentajes de Exactitud de cada punto clave

| <b>Especie</b> | <b>Boca</b> | <b>Cuello</b> | <b>Hombro</b> | <b>Codo</b> | <b>Mano</b> | <b>Anca</b> | <b>Rodilla</b> | <b>Pata</b> |
|----------------|-------------|---------------|---------------|-------------|-------------|-------------|----------------|-------------|
| Caballo        | 97.4        | 92.1          | 90.8          | 92.1        | 89.5        | 89.5        | 90.8           | 86.8        |
| Búfalo         | 96.6        | 84.7          | 88.1          | 88.1        | 78.0        | 84.7        | 84.7           | 81.4        |
| Ciervo         | 74.7        | 59.5          | 53.2          | 57.0        | 53.2        | 49.4        | 53.2           | 46.8        |
| Bisonte        | 84.8        | 75.9          | 73.4          | 70.9        | 68.4        | 67.1        | 73.4           | 68.4        |
| Cabra          | 80.0        | 77.8          | 58.9          | 63.3        | 67.8        | 56.7        | 62.2           | 61.1        |
| Oveja          | 76.2        | 68.3          | 63.5          | 65.1        | 58.7        | 61.9        | 52.4           | 54.0        |
| Alpaca         | 58.6        | 62.9          | 74.3          | 68.6        | 71.4        | 68.6        | 64.3           | 61.4        |
| Camello        | 36.4        | 63.6          | 63.6          | 67.3        | 65.5        | 60.0        | 58.2           | 60.0        |
| Yak            | 81.3        | 76.0          | 57.3          | 56.0        | 61.3        | 53.3        | 52.0           | 53.3        |

La estimación de la pose humana ha sido un campo de investigación activo durante décadas. Recientemente, los enfoques basados en el aprendizaje profundo han logrado un progreso significativo debido a la disponibilidad de datos de entrenamiento a gran escala, como el conjunto de datos MPII (Andriluka et al. 2014), el conjunto de detección de puntos clave de COCO (Lin et al. 2014), conjuntos de escenas de multitudes como CrowdPose (Li et al. 2019) o escenas de oclusión como OCHuman (Zhang et al. 2019). Los modelos entrenados con estos conjuntos de datos han demostrado ser efectivos para la estimación de posturas humanas, debido a la rica diversidad de posturas, iluminación, escala, oclusión y el número de instancias por imagen (Zhang et al. 2021). Pero según Li y Lee (2021), la estimación de la postura de los animales está relativamente poco explorada, principalmente por la falta de conjuntos de datos etiquetados.

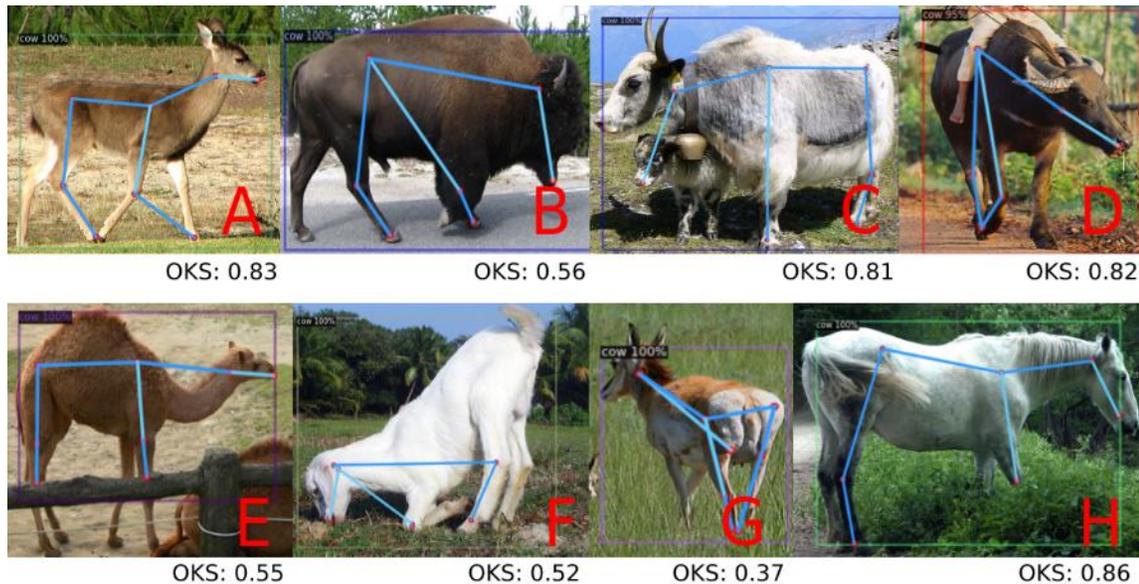
Para resolver este problema, Mu et al. (2020) construyeron una base de datos a partir de datos de animales sintéticos generados a partir de modelos tridimensionales para entrenar su modelo; el cual fue utilizado para generar pseudo etiquetas en las imágenes de animales reales. Cao et al. (2019) propuso un esquema de adaptación de dominio cruzado utilizando un espacio de características compartidas entre imágenes de humanos y animales, de modo que la red pueda aprender de los conjuntos de datos de poses humanas existentes y utilizar este aprendizaje en la identificación de imágenes de animales.



**Figura 3.** Resultados acertados de inferencia.

Basados en el trabajo de Cao et al. (2019), proponemos un método de adaptación de dominio para utilizar el conocimiento de la postura de los animales etiquetados (vacunos) en la identificación de clases de animales no etiquetadas (otros rumiantes y equinos). Utilizando un modesto conjunto de datos de pose de vacunos (2000 animales etiquetados). El esquema de Cao et al. (2019) se basa en que, por lo general, diferentes conjuntos de datos comparten características similares, especialmente cuando son imágenes extraídas de dominios cercanos (especies de animales cuadrúpedos). Para aprovechar esas características compartidas entre dominios, la adaptación de dominio (Hoffman et al. 2017; Tu et al. 2019) ha sido ampliamente estudiada en tareas con imágenes, como la detección (Inoue et al. 2018), clasificación (Ge et al. 2018; Gebru et al. 2017), segmentación (Zou et al. 2018) y estimación de pose (Zhou et al. 2018). En este sentido es importante señalar que el cambio desde el dominio de origen al dominio de destino es mucho más leve entre especies de

animales cuadrúpedos, que cuando se transfieren de un conjunto de datos humanos a animales o entre especies animales con marcadas diferencias anatómicas.



**Figura 4.** Resultados erróneos de inferencia.

En la Figura 4 se presentan ejemplos de imágenes etiquetadas incorrectamente. Como se puede ver, el error en la predicción se debe principalmente a confusión entre puntos claves simétricos (A), entre partes del animal (B) o entre animales (C), a la presencia de objetos que cubren áreas en el animal (E, H), a posturas no vistas por el modelo durante el entrenamiento (F, G). Esto indica que incluso entre dominios muy similares es posible que haya información adicional que dificulte la transferencia de conocimientos, como la ubicación espacial del animal dentro de la imagen, oclusiones o disimilitudes morfológicas. En un sistema de detección de pose en granja es posible que el principal problema sean los individuos que interactúan estrechamente y por tanto se ocluyen entre sí. Estimar la pose de varios animales es un problema desafiante para la visión por computadora, ya que las interacciones causan oclusiones y complican la asociación de los puntos clave detectados con los individuos correctos. Para hacer que los detectores sean robustos en este tipo de escenarios se pueden anotar imágenes con animales que interactúan, asociando los puntos clave de cada individuo en la imagen.

La medición de fenotipos conductuales individuales requiere herramientas que sean robustas, escalables y fáciles de usar. Nuestro enfoque demuestra la capacidad de las redes neuronales para cuantificar de manera rápida y precisa la posición corporal de distintas especies animales con un nivel de detalle apropiado. Sin embargo, y en contraste con los conjuntos de datos de poses humanas, recopilar y anotar puntos clave de animales es una tarea desafiante ya que existen muchas especies y diferentes familias, y se requiere un conocimiento biológico específico para distinguir los puntos claves entre animales. Si bien es desafiante y costoso, es muy necesario establecer un punto de referencia a gran escala para facilitar la investigación en este campo. La aplicación de estos métodos al estudio de individuos y grupos en la naturaleza puede proporcionar datos de comportamiento de alta resolución que van más allá de las capacidades de las tecnologías actuales basadas en GPS y acelerometría (Nagy et al. 2010; Nagy et al. 2013; Kays et al. 2015; Strandburg-Peshkin et al. 2015; Strandburg-Peshkin et al. 2017; Flack et al. 2018). Cuando se combinan, estas tecnologías podrían permitir a los investigadores abordar preguntas sobre el comportamiento de los animales que antes eran imposibles de responder.

#### **4. CONCLUSIONES**

- Los métodos de aprendizaje profundo de última generación ofrecen importantes avances para el análisis detallado de datos, permitiendo a los investigadores estimar automáticamente la ubicación de las partes del cuerpo de un animal directamente a partir de imágenes o videos.
- Un conjunto de datos relativamente pequeño de la posición en animales permite evaluar la generalización de la inferencia de los modelos dentro (vacunos) y fuera del dominio (otros rumiantes y equinos).
- La distancia filogenética entre especies no representa un obstáculo para que la información de pose de una determinada clase de animales permita estimar la pose de animales de otras clases, incluso si el grado de similitud no es máximo.

## **CONTRIBUCIÓN DE LA AUTORÍA (author's contribution statement)**

**John Fredy Ramírez Agudelo:** Conceptualización, Manejo de datos, Software, Análisis formal, Validación, Investigación, Visualización, Metodología, Redacción del borrador original.

**José Fernando Guarín Montoya:** Administración del proyecto, Redacción, revisión y edición. **Sebastián Bedoya Mazo:** Curación de datos, Supervisión, Administración de proyectos, redacción, revisión y edición, Adquisición de fondos.

## **DECLARACIÓN DE CONFLICTO DE INTERESES**

No existe conflicto de intereses en el desarrollo de este trabajo.

## **ORCIDs**

**John Fredy Ramírez Agudelo:** <https://orcid.org/0000-0002-2169-7260>

**José Fernando Guarín Montoya:** <https://orcid.org/0000-0001-7795-0184>

**Sebastián Bedoya Mazo:** <https://orcid.org/0000-0003-0595-9790>

## **AGRADECIMIENTOS**

Los autores manifiestan su agradecimiento a Minciencias por su apoyo al proyecto "Monitoreo y simulación de la emisión de gases de efecto invernadero (GEI) en sistemas de producción lechera especializada". Convocatoria 836-2019, proyecto 66737.

## **REFERENCIAS**

Albiero, V., Chen, X., Yin, X., Pang, G., & Hassner, T. (2021). img2pose: Face alignment and detection via 6dof, face pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7617-7627). [https://openaccess.thecvf.com/content/CVPR2021/html/Albiero\\_img2pose\\_Face\\_Alignment\\_and\\_Detection\\_via\\_6DoF\\_Face\\_Pose\\_Estimation\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Albiero_img2pose_Face_Alignment_and_Detection_via_6DoF_Face_Pose_Estimation_CVPR_2021_paper.html)

Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition* (pp. 3686-3693). [https://openaccess.thecvf.com/content\\_cvpr\\_2014/html/Andriluka\\_2D\\_Human\\_Pose\\_2014\\_CVPR\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2014/html/Andriluka_2D_Human_Pose_2014_CVPR_paper.html)

Bala, P. C., Eisenreich, B. R., Yoo, S. B. M., Hayden, B. Y., Park, H. S., & Zimmermann, J. (2020). Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nature communications*, *11*(1), 1-12. <https://doi.org/10.1038/s41467-020-18441-5>

Biggs, B., Boyne, O., Charles, J., Fitzgibbon, A., & Cipolla, R. (2020). Who left the dogs out? 3d animal reconstruction with expectation maximization in the loop. In *European Conference on Computer Vision* (pp. 195-211). Springer, Cham. [https://doi.org/10.1007/978-3-030-58621-8\\_12](https://doi.org/10.1007/978-3-030-58621-8_12)

Boualia, S. N., & Amara, N. E. B. (2019). Pose-based human activity recognition: a review. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 1468-1475). IEEE. <https://doi.org/10.1109/IWCMC.2019.8766694>

Brooks, 2019. Cocoannotator. <https://github.com/jsbroks/coco-annotator/>

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299). [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Cao\\_Realtime\\_Multi-Person\\_2D\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.html)

Cao, J., Tang, H., Fang, H. S., Shen, X., Lu, C., & Tai, Y. W. (2019). Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9498-9507). [https://openaccess.thecvf.com/content\\_ICCV\\_2019/html/Cao\\_Cross-Domain\\_Adaptation\\_for\\_Animal\\_Pose\\_Estimation\\_ICCV\\_2019\\_paper.html](https://openaccess.thecvf.com/content_ICCV_2019/html/Cao_Cross-Domain_Adaptation_for_Animal_Pose_Estimation_ICCV_2019_paper.html)

Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7103-7112). [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Chen\\_Cascaded\\_Pyramid\\_Network\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Cascaded_Pyramid_Network_CVPR_2018_paper.html)

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.. <https://doi.org/10.1109/CVPR.2009.5206848>

Fang, H. S., Xie, S., Tai, Y. W., & Lu, C. (2017). Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2334-2343). [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Fang\\_RMPE\\_Regional\\_Multi-Person\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Fang_RMPE_Regional_Multi-Person_ICCV_2017_paper.html)

Ge, W., Yang, S., & Yu, Y. (2018). Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1277-1286). [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Ge\\_Multi-Evidence\\_Filtering\\_and\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Ge_Multi-Evidence_Filtering_and_CVPR_2018_paper.html)

Geburu, T., Hoffman, J., & Fei-Fei, L. (2017). Fine-grained recognition in the wild: A multi-task domain adaptation approach. In *Proceedings of the IEEE international conference on computer vision* (pp. 1349-1358). [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Geburu\\_Fine-Grained\\_Recognition\\_in\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Geburu_Fine-Grained_Recognition_in_ICCV_2017_paper.html)

Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., & Couzin, I. D. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife*, 8, e47994. <https://doi.org/10.7554/eLife.47994>

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.

[https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/He\\_Deep\\_Residual\\_Learning\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html)

Hoffman, J., Tzeng, E., Darrell, T., & Saenko, K. (2017). Simultaneous deep transfer across domains and tasks. In *Domain Adaptation in Computer Vision Applications* (pp. 173-187). Springer, Cham. [https://doi.org/10.1007/978-3-319-58347-1\\_9](https://doi.org/10.1007/978-3-319-58347-1_9)

Inoue, N., Furuta, R., Yamasaki, T., & Aizawa, K. (2018). Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5001-5009). [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Inoue\\_Cross-Domain\\_Weakly-Supervised\\_Object\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Inoue_Cross-Domain_Weakly-Supervised_Object_CVPR_2018_paper.html)

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In *European Conference on Computer Vision* (pp. 34-50). Springer, Cham. [https://doi.org/10.1007/978-3-319-46466-4\\_3](https://doi.org/10.1007/978-3-319-46466-4_3)

Khan, M. H., McDonagh, J., Khan, S., Shahabuddin, M., Arora, A., Khan, F. S., ... & Tzimiropoulos, G. (2020). Animalweb: A large-scale hierarchical dataset of annotated animal faces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6939-6948). [https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Khan\\_AnimalWeb\\_A\\_Large-Scale\\_Hierarchical\\_Dataset\\_of\\_Annotated\\_Animal\\_Faces\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Khan_AnimalWeb_A_Large-Scale_Hierarchical_Dataset_of_Annotated_Animal_Faces_CVPR_2020_paper.html)

Khosla, A., Jayadevaprakash, N., Yao, B., & Li, F. F. (2011). Novel dataset for fine-grained image categorization: Stanford dogs. In *Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)* (Vol. 2, No. 1). <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.255.6394&rep=rep1&type=pdf>

Kim, D. U., Kim, K. I., & Baek, S. (2021). End-to-End Detection and Pose Estimation of Two Interacting Hands. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11189-11198).

[https://openaccess.thecvf.com/content/ICCV2021/html/Kim\\_End-to-End\\_Detection\\_and\\_Pose\\_Estimation\\_of\\_Two\\_Interacting\\_Hands\\_ICCV\\_2021\\_paper.html](https://openaccess.thecvf.com/content/ICCV2021/html/Kim_End-to-End_Detection_and_Pose_Estimation_of_Two_Interacting_Hands_ICCV_2021_paper.html)

Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., ... & Shibata, T. (2021). MacaquePose: A novel "in the wild" macaque monkey pose dataset for markerless motion capture. *Frontiers in behavioral neuroscience*, 268. <https://doi.org/10.3389/fnbeh.2020.581154>

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>

Li, C., & Lee, G. H. (2021). From Synthetic to Real: Unsupervised Domain Adaptation for Animal Pose Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1482-1491). [https://openaccess.thecvf.com/content/CVPR2021/html/Li\\_From\\_Synthetic\\_to\\_Real\\_Unsupervised\\_Domain\\_Adaptation\\_for\\_Animal\\_Pose\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Li_From_Synthetic_to_Real_Unsupervised_Domain_Adaptation_for_Animal_Pose_CVPR_2021_paper.html)

Li, S., Li, J., Tang, H., Qian, R., & Lin, W. (2019). ATRW: a benchmark for Amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*. <https://doi.org/10.1145/3394171.3413569>

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer, Cham. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)

Li, J., Wang, C., Zhu, H., Mao, Y., Fang, H., & Lu, C. C. (2019). Efficient Crowded Scenes Pose Estimation and A New Benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA* (pp. 16-20). [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Li\\_CrowdPose\\_Efficient\\_Crow](https://openaccess.thecvf.com/content_CVPR_2019/html/Li_CrowdPose_Efficient_Crow)

Marginean, A., Brehar, R., & Negru, M. (2019). Understanding pedestrian behaviour with pose estimation and recurrent networks. In *2019 6th International Symposium on Electrical and Electronics Engineering (ISEEE)* (pp. 1-6). IEEE. <https://doi.org/0.1109/ISEEE48094.2019.9136126>

Mathis, A., Biasi, T., Schneider, S., Yuksekgonul, M., Rogers, B., Bethge, M., & Mathis, M. W. (2021). Pretraining boosts out-of-domain robustness for pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1859-1868). [https://openaccess.thecvf.com/content/WACV2021/html/Mathis\\_Pretraining\\_Boosts\\_Out-of-Domain\\_Robustness\\_for\\_Pose\\_Estimation\\_WACV\\_2021\\_paper.html](https://openaccess.thecvf.com/content/WACV2021/html/Mathis_Pretraining_Boosts_Out-of-Domain_Robustness_for_Pose_Estimation_WACV_2021_paper.html)

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*(9), 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*. 2020, vol. 108, no 1, p. 44-65. <https://doi.org/10.1016/j.neuron.2020.09.017>

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., ... & Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *science*, *334*(6055), 521-524. <https://doi.org/10.1126/science.1211028>

Mu, J., Qiu, W., Hager, G. D., & Yuille, A. L. (2020). Learning from synthetic animals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12386-12395). [https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Mu\\_Learning\\_From\\_Synthetic\\_Animals\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Mu_Learning_From_Synthetic_Animals_CVPR_2020_paper.html)

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S. H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature methods*, *16*(1), 117-125. <https://doi.org/10.1038/s41592-018-0234-5>

Pratt, L. Y. (1993). Discriminability-based transfer between neural networks. *Advances in neural information processing systems*, 204-204. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.979.7402&rep=rep1&type=pdf>

Ruggero Ronchi, M., & Perona, P. (2017). Benchmarking and error diagnosis in multi-instance pose estimation. In *Proceedings of the IEEE international conference on computer vision* (pp. 369-378). [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Ronchi\\_Benchmarking\\_and\\_Error\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Ronchi_Benchmarking_and_Error_ICCV_2017_paper.html)

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4), 427-437. <https://doi.org/10.1016/j.ipm.2009.03.002>

Tu, X., Zhang, H., Xie, M., Luo, Y., Zhang, Y., & Ma, Z. (2019). Deep transfer across domains for face antispoofing. *Journal of Electronic Imaging*, 28(4), 043001. <https://doi.org/10.1117/1.JEI.28.4.043001>

Yang, L., & Yao, A. (2019). Disentangling latent hands for image synthesis and pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9877-9886). [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Yang\\_Disentangling\\_Latent\\_Hands\\_for\\_Image\\_Synthesis\\_and\\_Pose\\_Estimation\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Yang_Disentangling_Latent_Hands_for_Image_Synthesis_and_Pose_Estimation_CVPR_2019_paper.html)

Wei, F., Sun, X., Li, H., Wang, J., & Lin, S. (2020). Point-set anchors for object detection, instance segmentation and pose estimation. In *European Conference on Computer Vision* (pp. 527-544). Springer, Cham. [https://doi.org/10.1007/978-3-030-58607-2\\_31](https://doi.org/10.1007/978-3-030-58607-2_31)

Wu, M. Y., Ting, P. W., Tang, Y. H., Chou, E. T., & Fu, L. C. (2020). Hand pose estimation in object-interaction based on deep learning for virtual reality applications. *Journal of Visual Communication and Image Representation*, 70, 102802. <https://doi.org/10.1016/j.jvcir.2020.102802>

Zhang, J., Chen, Z., & Tao, D. (2021). Towards high performance human keypoint detection. *International Journal of Computer Vision*. 2021, vol. 129, no 9, p. 2639-2662. <https://doi.org/10.1007/s11263-021-01482-8>

Zhang, S. H., Li, R., Dong, X., Rosin, P., Cai, Z., Han, X., ... & Hu, S. M. (2019). Pose2seg: Detection free human instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. p. 889-898. [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Zhang\\_Pose2Seg\\_Detection\\_Free\\_Human\\_Instance\\_Segmentation\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_Pose2Seg_Detection_Free_Human_Instance_Segmentation_CVPR_2019_paper.html)

Zhou, X., Karpur, A., Gan, C., Luo, L., & Huang, Q. (2018). Unsupervised domain adaptation for 3d keypoint estimation via view consistency. In *Proceedings of the European conference on computer vision (ECCV)*. p. 137-153. [https://openaccess.thecvf.com/content\\_ECCV\\_2018/html/Xingyi\\_Zhou\\_Unsupervised\\_Domain\\_Adaptation\\_ECCV\\_2018\\_paper.html](https://openaccess.thecvf.com/content_ECCV_2018/html/Xingyi_Zhou_Unsupervised_Domain_Adaptation_ECCV_2018_paper.html)

Zou, Y., Yu, Z., Kumar, B. V. K., & Wang, J. (2018). Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 289-305). [https://openaccess.thecvf.com/content\\_ECCV\\_2018/html/Yang\\_Zou\\_Unsupervised\\_Domain\\_Adaptation\\_ECCV\\_2018\\_paper.html](https://openaccess.thecvf.com/content_ECCV_2018/html/Yang_Zou_Unsupervised_Domain_Adaptation_ECCV_2018_paper.html)

## Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.