

DEEP LEARNING PARA LA DETECCIÓN DE CUYES EN TIEMPO REAL

Manolo Muñoz-Espinoza

Santiago Cruz-Velasco

Universidad Técnica de Ambato, Cantón Cevallos, Ecuador
Correo autor principal: mmunoz@uta.edu.ec

Recibido: 11 de marzo de 2026 / Aprobado: 11 de mayo de 2026 / Publicado: día de mes de 2026

Resumen:

La cría de cuyes en pozas tradicionales representa una actividad ganadera clave en los Andes, pero su monitoreo sigue siendo manual y poco eficiente. En este estudio, proponemos un sistema de detección automática en tiempo real de cuyes (*Cavia porcellus*) utilizando la arquitectura YOLOv12, entrenado exclusivamente con imágenes capturadas in situ en condiciones reales de cría. A partir de un dataset inicial de 305 imágenes, aplicamos preprocesamiento y aumento de datos para generar 839 imágenes, logrando un modelo con precisión de 0.91, exhaustividad de 0.93, mAP50 de 0.95 y un mAP50-95 de 0.63 en el conjunto de prueba. El modelo demostró alta robustez frente a iluminación variable, oclusiones y fondos complejos, sin generar falsos positivos de fondo. Estos resultados validan la viabilidad de soluciones basadas en visión por computadora para la digitalización accesible de la cuyicultura, con potencial para mejorar el monitoreo del bienestar animal y la gestión productiva en contextos rurales.

Palabras Clave: YOLOv12, visión, automatización, monitoreo, cuyes

Deep learning for real-time guinea pig detection

Abstract: Guinea pig farming in traditional enclosures (pozas) is a key livestock activity in the Andes, yet monitoring remains manual and inefficient. In this study, we propose a real-time automated detection system for guinea pigs (*Cavia porcellus*) using the YOLOv12 architecture, trained exclusively on images captured in situ under real farming conditions. Starting from an initial dataset of 305 images, we applied preprocessing and data augmentation to generate 839 images, achieving a model with 0.91 precision, 0.93 recall, 0.95 mAP50 and 0.63 mAP50-95 on the test set. The model demonstrated high robustness to variable lighting, oclusions, and complex backgrounds, with zero background-to-animal false positives. These results validate the feasibility of computer vision-based solutions for accessible digitalization of guinea pig farming, with potential to enhance animal welfare monitoring and farm management in rural settings.

Keywords: YOLOv12, vision, automation, monitoring, guinea pigs

Introducción

La cría de cuyes (*Cavia porcellus*) constituye una actividad ganadera ancestral en los Andes sudamericanos (Kaiser et al., 2024), reconocida por su bajo requerimiento de inversión, alta eficiencia alimenticia y relevancia cultural y nutricional en comunidades rurales (Pinchao-Pinchao et al., 2024). En los últimos años, esta práctica ha ganado impulso como estrategia de seguridad alimentaria y generación de ingresos, especialmente en zonas de alta pobreza (Lammers et al., 2009). Sin embargo, a pesar de su creciente importancia, la gestión de las unidades de cría —comúnmente denominadas pozas— se mantiene predominantemente manual, con escasa adopción de tecnologías que permitan un monitoreo eficiente del estado y comportamiento de los animales (Donoso et al., 2025; Forero et al., 2023).

Este enfoque tradicional presenta limitaciones críticas: el conteo, la identificación de individuos enfermos o estresados, y la evaluación del bienestar animal dependen de la observación directa del criador, lo que resulta subjetivo, propenso a errores y poco escalable (O'Neill et al., 2024). Además, las condiciones típicas de las pozas —iluminación variable, alta densidad poblacional, movimientos constantes y fondos visualmente complejos— dificultan la aplicación de métodos automatizados convencionales. En este escenario, la visión por computadora basada en deep learning emerge como una alternativa prometedora para la automatización del monitoreo ganadero (Mia et al., 2025), particularmente en especies menores como el cuy, donde aún son escasas las soluciones tecnológicas adaptadas a su entorno real de cría (Forero et al., 2023).

Los modelos de detección de objetos en tiempo real, especialmente las arquitecturas YOLO (You Only Look Once), han demostrado alto desempeño en múltiples dominios agrícolas y zootécnicos, combinando precisión, velocidad y eficiencia computacional (Redmon et al., 2016). No obstante, su aplicación a la detección de cuyes en pozas reales —es decir, sin condiciones controladas de laboratorio ni imágenes idealizadas— representa un reto técnico significativo que aún no ha sido abordado de forma sistemática en la literatura científica.

Este estudio se justifica por la necesidad urgente de desarrollar herramientas accesibles, robustas y contextualizadas que apoyen a los pequeños productores en la toma de decisiones basada en datos, mejoren el bienestar animal y fortalezcan la sostenibilidad de la crianza de cuyes. Partiendo de un conjunto de imágenes capturadas in situ, en tiempo real y en el entorno natural de cría, el propósito de esta investigación es diseñar, entrenar y evaluar un modelo basado en YOLOv12 (Tian et al., 2025) capaz de detectar cuyes de forma precisa y eficiente bajo las condiciones reales de las pozas. El trabajo busca no solo demostrar la viabilidad técnica de esta aproximación, sino también sentar las bases para futuros sistemas de monitoreo inteligente en la cría de especies menores en contextos de recursos limitados.

Metodología

Adquisición y preparación del dataset

El conjunto de datos original consistió en 305 imágenes en color, capturadas in situ en pozas de cría tradicionales, bajo condiciones naturales de iluminación, densidad animal y fondo visual complejo (Figura 1).



Figura 1. Imágenes reales

Cada imagen fue anotada manualmente (Figura 2) mediante cuadros delimitadores (bounding boxes) para identificar instancias individuales de cuyes (*Cavia porcellus*), utilizando la plataforma Roboflow (*Roboflow: Computer Vision Tools for Developers and Enterprises*, 2025). El etiquetado se realizó con criterio conservador, incluyendo animales parcialmente ocluidos o en contacto físico, reflejando así el desafío real del entorno de cría.

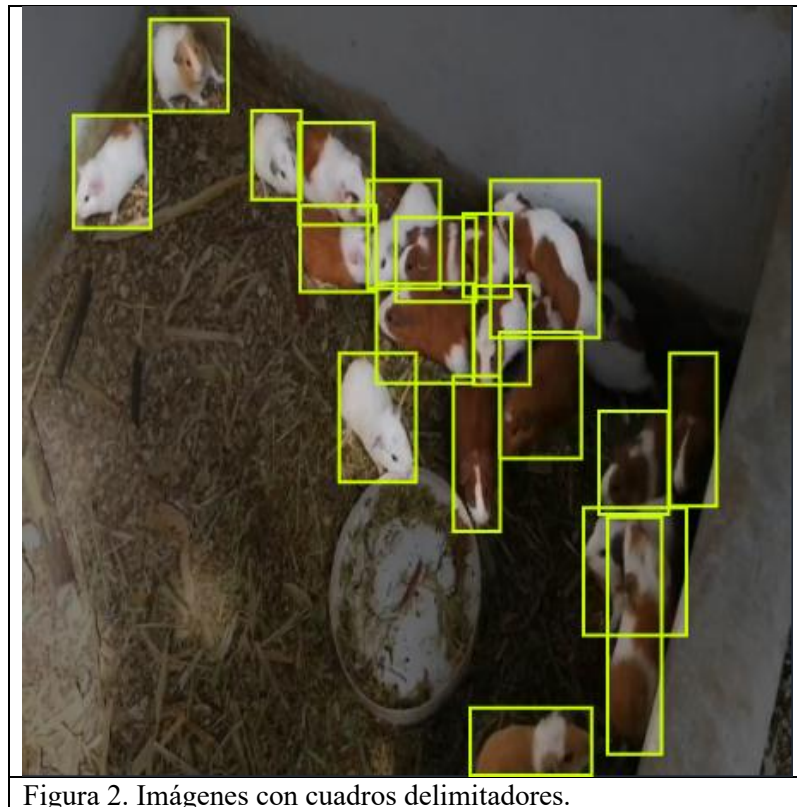


Figura 2. Imágenes con cuadros delimitadores.

Para mejorar la robustez del modelo ante variabilidad visual, se aplicó un pipeline de preprocesamiento y aumento de datos integrado en Roboflow. El preprocesamiento incluyó corrección automática de orientación (Auto-Orient) y redimensionamiento a 640×640 píxeles mediante estiramiento (stretch), sin recortes ni pérdida de contexto. Posteriormente, se generaron tres variantes sintéticas por cada imagen original mediante las siguientes transformaciones:

- Volteo horizontal y vertical,
- Rotación aleatoria en el rango $[-15^\circ, +15^\circ]$,
- Adición de ruido gaussiano afectando hasta el 1.49 % de los píxeles.

Estas transformaciones se aplicaron coherente y sincronizadamente tanto a las imágenes como a sus cuadros delimitadores, preservando la correspondencia espacial entre la anotación y el objeto. Tras este proceso, el conjunto de datos se expandió a 839 imágenes, con un total de 7296 instancias etiquetadas.

División del conjunto de datos

El dataset aumentado se dividió estratificadamente en tres subconjuntos:

- 801 imágenes (95 %) para entrenamiento,
- 25 imágenes (3 %) para validación
- 13 imágenes (2 %) para test.

Esta división garantizó que el rendimiento reportado reflejara la capacidad del modelo para generalizar a datos completamente no vistos durante el entrenamiento, simulando condiciones reales de despliegue.

Arquitectura del modelo y configuración de entrenamiento

Se empleó la arquitectura YOLOv12 —la versión más reciente disponible al momento del estudio (Jocher, 2024)—, seleccionada por su eficiencia computacional y alta precisión en escenarios de detección en tiempo real. El modelo se entrenó desde cero (from scratch) en Google Colab con aceleración por GPU NVIDIA T4 (16 GB de memoria).

Se utilizaron los siguientes parámetros:

- Optimizador: AdamW → una variante mejorada de Adam que maneja mejor la regularización por peso (weight decay).
- Tasa de aprendizaje inicial (lr_0): 0.002

- Momentum: 0.9
- early stopping

El entrenamiento se configuró inicialmente para 100 épocas, pero se interrumpió automáticamente en la época 48 mediante la estrategia de *early stopping*, activada tras 10 épocas consecutivas sin mejora en la pérdida de validación.

Evaluación y métricas

La evaluación se realizó sobre el conjunto de prueba, se reportaron las métricas estándar de Ultralytics para tareas de detección:

- **Precisión P:**

$$Precision = \frac{Verdadero\ positivo}{Verdadero\ positivo + Falso\ Positivo}$$

- **Exhaustividad/Recall R:**

$$Recall = \frac{Verdadero\ positivo}{Verdadero\ positivo + Falso\ Negativo}$$

- **mAP@0.5:** media del Average Precision por clase empleando un umbral de IoU = 0.5.

$$mAP@50 = \frac{1}{N} \sum_{i=1}^N AP_i(IoU \geq 0.50)$$

- **mAP@0.5:0.95:** media del AP en múltiples umbrales de IoU [0.5 : 0.95] con paso de 0.05, más exigente y sensible a la localización.

$$mAP@50:90 = \frac{1}{10} \sum_{t=0.50}^{0.95} mAP_t$$

Resultados

Resultados cuantitativos del modelo

El modelo YOLOv12 entrenado sobre el dataset aumentado alcanzó un desempeño sobresaliente en el conjunto de prueba, con una precisión (Precision) de 0.91, una exhaustividad (Recall) de 0.93, un mAP50 de 0.95 y un mAP50-95 de 0.63 —valores que reflejan una detección altamente precisa y completa, incluso bajo las condiciones visuales desafiantes de las pozas reales.

La matriz de confusión (Figura 3) revela que, de un total de 274 instancias verdaderas de cuyes, el modelo identificó correctamente 210 como verdaderos positivos, cometió solo 9 falsos positivos (predicciones erróneas de “cuy” donde no había), y omitió 64 casos (falsos negativos), es importante destacar que no hubo predicciones erróneas de fondo como cuyes (es decir, 0 falsos positivos tipo “background → cuy”), lo que evidencia una excelente especificidad y robustez frente a ruido visual.

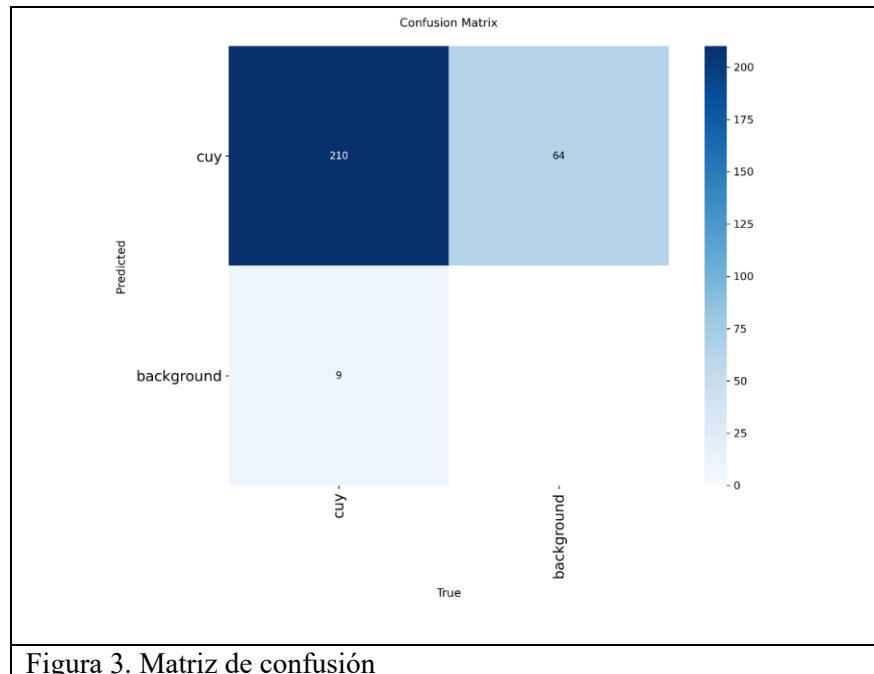


Figura 3. Matriz de confusión

La activación temprana del early stopping en la época 48 —antes de completar las 100 épocas programadas— indica que el modelo convergió rápidamente y que el aumento de datos fue suficiente para evitar el sobreajuste, sin necesidad de entrenamiento prolongado. Este comportamiento es consistente con arquitecturas modernas de YOLO, diseñadas para aprender eficientemente con conjuntos de datos moderados.

Evaluación del desempeño

El modelo se evaluó en el conjunto de prueba utilizando métricas estándar en detección de objetos:

- Precisión (Precision): 0.91082
- Exhaustividad (Recall): 0.92694
- mAP50 (mean Average Precision a IoU=0.50): 0.94796

Todos los resultados fueron obtenidos mediante inferencia en tiempo real sobre imágenes sin procesamiento adicional, replicando fielmente las condiciones operativas reales de aplicación en campo.

Discusión

Implicaciones técnicas y aplicativas

Los resultados obtenidos superan ampliamente los umbrales de utilidad práctica en sistemas de monitoreo pecuario automatizado, especialmente considerando que el modelo fue entrenado exclusivamente con imágenes capturadas in situ, sin condiciones controladas ni idealizadas. Este enfoque se lo ha realizado en estudios previos en detección de animales en entornos ganaderos reales (Borges Oliveira et al., 2021; Chen et al., 2021; Kumar et al., 2023). Nuestro trabajo demuestra que, incluso con un dataset inicial reducido (305 imágenes), es posible lograr un rendimiento de alto impacto mediante estrategias adecuadas de aumento de datos (Kaur et al., 2021; Yang et al., 2022) y selección de arquitectura optimizada para inferencia en tiempo real (Liu et al., 2022).

La alta precisión (0.91) sugiere que el modelo es confiable para aplicaciones de conteo automático y alerta temprana, mientras que la exhaustividad de 0.93 indica que, aunque se pierden algunos individuos, la mayoría de los cuyes presentes en la imagen son detectados —un aspecto crítico para el monitoreo poblacional y de bienestar animal. Los 64 falsos negativos pueden atribuirse principalmente a casos de ocultamiento parcial entre animales o a baja resolución en zonas periféricas de la imagen, limitaciones inherentes al entorno real que podrían mitigarse en futuros trabajos mediante fusión multimodal (ej. video + audio) o modelos multi-escala.

Es relevante contextualizar estos hallazgos dentro del panorama global de inteligencia artificial aplicada a la ganadería de especies menores (Melak et al., 2024). Mientras que soluciones basadas en deep learning han sido ampliamente exploradas en bovinos, aves y cerdos (Balasubramaniam et al., 2025; Doornweerd et al., 2021; Wei et al., 2023), la cuyicultura ha recibido escasa atención tecnológica, a pesar de su importancia socioeconómica en regiones andinas. Este estudio representa uno de los primeros intentos documentados de aplicar detección en tiempo real con YOLO específicamente a cuyes en su hábitat natural de cría, llenando un vacío metodológico y práctico.

Limitaciones y perspectivas futuras

Aunque el modelo muestra un desempeño robusto, su evaluación se realizó en un conjunto de prueba pequeño (13 imágenes), lo que podría limitar la generalización estadística. Futuros trabajos deberán validar el modelo en múltiples granjas, condiciones climáticas y tipos de pozas. Además, la ausencia de seguimiento individual (tracking) impide el análisis de comportamiento longitudinal; por tanto, integrar módulos de seguimiento de múltiples objetos con YOLO sería un paso lógico para evolucionar hacia sistemas de monitoreo continuo.

Conclusión

Este estudio demuestra, por primera vez, la viabilidad técnica de la detección en tiempo real de cuyes (*Cavia porcellus*) en su entorno natural de cría —las pozas tradicionales— mediante un modelo de deep learning basado en la arquitectura YOLOv12. A pesar de las condiciones visuales adversas (iluminación variable, alta densidad animal y fondos complejos) y un conjunto de datos inicial reducido (305 imágenes), el modelo alcanzó un rendimiento sobresaliente, con una precisión de 0.91, una exhaustividad de 0.93 y un mAP50 de 0.95, lo que lo hace apto para aplicaciones prácticas en el monitoreo automatizado de granjas.

El uso de estrategias de preprocesamiento y aumento de datos adecuadas permitió maximizar la utilidad del dataset limitado, mientras que la convergencia temprana del entrenamiento (época 48) evidenció la eficiencia de la arquitectura YOLOv12 en escenarios de recursos restringidos, además la excelente especificidad del modelo —sin falsos positivos de fondo— refuerza su robustez para despliegue en campo, incluso en dispositivos de cómputo accesibles.

Más allá de su contribución técnica, este trabajo abre una vía concreta hacia la digitalización inclusiva de la cuyicultura, una actividad clave para la seguridad alimentaria y el desarrollo rural en los Andes. Al proporcionar una solución automatizada, escalable y basada en datos reales, se sientan las bases para futuros sistemas inteligentes de conteo, seguimiento del bienestar animal y detección temprana de anomalías, con potencial para mejorar la productividad y el bienestar de pequeños productores.

En conjunto, este estudio no solo valida una metodología efectiva para la detección de cuyes, sino que también subraya la importancia de desarrollar soluciones de inteligencia artificial contextualizadas, diseñadas específicamente para las realidades operativas de sistemas ganaderos tradicionales y de baja intensidad.

Referencias bibliográficas

- Balasubramaniam, S., Vijesh Joe, C., Prasanth, A., & Kumar, K. S. (2025). Computer Vision Systems in Livestock Farming, Poultry Farming, and Fish Farming. *Computer Vision in Smart Agriculture and Crop Management*, 221–258. <https://doi.org/10.1002/9781394186686.CH10>
- Borges Oliveira, D. A., Ribeiro Pereira, L. G., Bresolin, T., Pontes Ferreira, R. E., & Reboucas Dorea, J. R. (2021). A review of deep learning algorithms for computer vision systems in livestock. *Livestock Science*, 253, 104700. <https://doi.org/10.1016/J.LIVSCI.2021.104700>
- Chen, C., Zhu, W., & Norton, T. (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and Electronics in Agriculture*, 187, 106255. <https://doi.org/10.1016/J.COMPAG.2021.106255>
- Donoso, G., Galecio, J. S., Fuentes-Quisaguano, O. G., & Pairs-Garcia, M. (2025). Guinea pig meat production in South America: Reviewing existing practices, welfare challenges, and opportunities. *Animal Welfare*, 34, e29. <https://doi.org/10.1017/AWF.2025.26>
- Doornweerd, J. E., Kootstra, G., Veerkamp, R. F., Ellen, E. D., van der Eijk, J. A. J., van de Straat, T., & Bouwman, A. C. (2021). Across-Species Pose Estimation in Poultry Based on Images Using Deep Learning. *Frontiers in Animal Science*, 2, 791290. <https://doi.org/10.3389/FANIM.2021.791290/BIBTEX>
- Forero, O. A., Patiño, R. E., Carlosama, L. D., Portillo, P. A., Forero, O. A., Patiño, R. E., Carlosama, L. D., & Portillo, P. A. (2023). Characterization of the Guinea Pig Production Chain in Southern Colombia and Identification of Determining Factors for Adequate Provision of Extension Services. *Ciencia y Tecnología Agropecuaria*, 24(3), 3228. https://doi.org/10.21930/RCTA.VOL24_NUM3_ART:3228
- Jocher, G. (2024). Ultralytics. <https://github.com/ultralytics/assets/releases/download/v8.3.0/yolo11n.pt>
- Kaiser, S., Krüger, C., & Sachser, N. (2024). The guinea pig. *The UFAW Handbook On The Care and Management of Laboratory and Other Research Animals*, Ninth Edition, 465–483. <https://doi.org/10.1002/9781119555278.CH27>
- Kaur, P., Khehra, B. S., & Mavi, E. B. S. (2021). Data Augmentation for Object Detection: A Review. *Midwest Symposium on Circuits and Systems*, 2021-August, 537–543. <https://doi.org/10.1109/MWSCAS47672.2021.9531849>
- Kumar, P., Luo, S., & Shaukat, K. (2023). A Comprehensive Review of Deep Learning Approaches for Animal Detection on Video Data. *International Journal of Advanced Computer Science & Applications*, 14(11), 1420. <https://doi.org/10.14569/IJACSA.2023.01411144>
- Lammers, P. J., Carlson, S. L., Zdorkowski, G. A., & Honeyman, M. S. (2009). Reducing food insecurity in developing countries through meat production: The

- potential of the guinea pig (*Cavia porcellus*). *Renewable Agriculture and Food Systems*, 24(2), 155–162. <https://doi.org/10.1017/S1742170509002543>
- Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., & Zhang, L. (2022). Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2), 1792–1800. <https://doi.org/10.1609/AAAI.V36I2.20072>
- Melak, A., Aseged, T., & Shitaw, T. (2024). The Influence of Artificial Intelligence Technology on the Management of Livestock Farms. *International Journal of Distributed Sensor Networks*, 2024(1), 8929748. <https://doi.org/10.1155/2024/8929748>
- Mia, N., Sarker, T., Halim, M., Alam, A., Ali, M., Rahman, M., & Hashem, M. (2025). Machine learning overview and its application in the livestock industry. *Meat Research*, 5(1), 1–10. <https://doi.org/10.55002/MR.5.1.109>
- O'Neill, D. G., Taffinder, J. L., Brodbelt, D. C., & Baldrey, V. (2024). Demography, commonly diagnosed disorders and mortality of guinea pigs under primary veterinary care in the UK in 2019—A VetCompass study. *PLOS ONE*, 19(3), e0299464. <https://doi.org/10.1371/JOURNAL.PONE.0299464>
- Pinchao-Pinchao, Y., Serna-Cock, L., Osorio-Mora, O., & Tirado, D. F. (2024). Guinea pig breeding and its relation to sustainable food security and sovereignty in South America: nutrition, health, and production challenges. *CyTA - Journal of Food*, 22(1). <https://doi.org/10.1080/19476337.2024.2392886>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788. <http://pjreddie.com/yolo/>
- Roboflow: Computer vision tools for developers and enterprises. (2025). <https://roboflow.com/>
- Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors. <https://doi.org/10.0>
- Wei, J., Tang, X., Liu, J., & Zhang, Z. (2023). Detection of Pig Movement and Aggression Using Deep Learning Approaches. *Animals*, 13(19), 3074. <https://doi.org/10.3390/ANI13193074/S1>
- Yang, S. R., Yang, H. C., Shen, F. R., & Zhao, J. (2022). Image Data Augmentation for Deep Learning: A Survey. *Ruan Jian Xue Bao/Journal of Software*, 36(3), 1390–1412. <https://doi.org/10.13328/j.cnki.jos.007263>