

REIVINDICACIONES:

1. Un método implementado por computador en el cual el procesador ejecuta etapas para monitorear y detectar pausas en vuelo de colibríes, caracterizado por las etapas de:
 - A. Una primera etapa de captura de videos en condiciones naturales a traves de un sensor CCD:
 - Grabar videos de Colibries a una velocidad de 500 fps en una estación de video.
 - Extraer los fotogramas del video para identificar las pausas en los vuelos de colibríes. Enviar una primera notificación en un cuadro de texto solicitando ingresar los objetos de interés a detectar por el usuario, que incluyen “hummingbird” y “flower”.
 - Envíar una segunda notificación en la cual se determina la eliminación de los fotogramas ya procesados para optimizar recursos computacionales.
 - Envíar a la nube los fotogramas procesados para alimentar la red de entrenamiento.
 - B. Una segunda etapa de detección de objetos en el video a partir de un algoritmo de detección ejecutado por un procesador que incluye redes convolucionales con pesos pre-entrenados en imágenes con 80 clases diferentes, donde los las imágenes de colibríes:
 - Inicializar el algoritmo se carga en memoria el modelo pre entrenado de detección.
 - Definir parámetros de optimización como umbrales de confianza y de intersección para determinar la selección de las cajas de delimitación finales de los objetos en 0.7.
 - Procesar cada uno de los frames de video para localizar objetos de interés en la imagen, que serán clasificados en una siguiente etapa.
 - Las cajas de delimitación (bounding boxes) encontradas se utilizan como entrada tanto en la etapa de clasificación como en la de seguimiento de objetos.
 - C. Una tercera etapa de clasificación de los colibríes de acuerdo a la similitud de características utilizando la distancia del coseno a partir del conjunto entre pares texto - imagen y corrigiendo las detecciones obtenidas en la etapa anterior (redes convolucionales), en la que el procesador se configura para que:
 - Al inicializar el algoritmo se carga el modelo pre entrenado de clasificación.

- Se extraen las características del texto ingresado por el usuario como objetos de interés.
- Una vez realizada la detección de objetos, se extrae la región de interés para obtener la sección de la imagen donde se obtendrán las características del objeto.
- Se comparan las características de texto con las características de la imagen, utilizando la distancia del coseno.
- En esta etapa se realiza también una sub-etapa, en la cual se realiza una fase de rechazo a la identificación de objetos diferentes a los colibríes y cuya definición inicial se realiza a través de entradas de texto personalizables por el usuario utilizando ingeniería de texto (prompt engineering);

D. Una cuarta etapa de seguimiento de cada colibrí durante el vuelo asignando un código de identificación individual a partir de la comparación entre características (embeddings) de frames secuenciales, en la que el procesador esta configurado para:

- obtener el seguimiento del objeto(s) de interés, a partir de las cajas de detección y las características de la imagen definidas en la primera notificación.
- Crear identificadores para los objetos encontrados en el primer fotograma de video.
- A partir del siguiente fotograma, el fotograma actual se convierte en la identificación anterior para comparar características de imagen entre el cuadro anterior, el actual y el siguiente, con la técnica de la frecuencia del coseno.
- Si la distancia del coseno supera un umbral dado, por defecto, 0.4, entonces, se asume que el objeto identificado es el mismo que se había almacenado en un principio. De lo contrario, se procesa como un nuevo objeto.

E. Una quinta etapa de monitoreo de la frecuencia de aleteo y las pausas aéreas realizadas por cada colibrí, el procesador se configura para:

- Seleccionar los colibríes encontrados y calcular la distancia de Jaccard entre las cajas limitantes (bounding boxes) que tienen el mismo id de identificación dado por el módulo de seguimiento de fotogramas subsiguientes. Cada caja se compara con la del fotograma anterior, con el fin de medir el movimiento realizado por el colibrí a lo largo del vídeo.
- A partir de la señal generada por el coeficiente se calculan:
 - i. La frecuencia de aleteo, obtenida a partir de:
 - La transformada de Fourier de tiempo corto (STFT).
 - El suavizado de la frecuencia principal aplicando un filtro savgol.

- El conteo de picos de la frecuencia. Que corresponden a la frecuencia de aleteo.
- ii. Las pausas en vuelo del colibrí, obtenida a partir de:
 - Cálculo de la derivada de la señal con respecto al tiempo.
 - Detección de regiones constantes a partir de la derivada, tomando valores localizados entre cero y la media de la señal

2. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, en el cual la función objetivo del problema de optimización para la detección de objetos, se divide en tres partes:

$$L_{loss} = L_{bbox} + L_{bbox_{score}} + L_{class_{score}}$$

Dónde:

- L_{bbox} : Función objetivo para encontrar las coordenadas de la caja (bounding box) que rodea el objeto de interés.

$$L_{bbox} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] +$$

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

- $L_{bbox_{score}}$ Función objetivo para determinar la confianza de la detección.

$$L_{bbox_{score}} = \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (C_i - \hat{C}_i) + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} (C_i - \hat{C}_i)$$

- $L_{class_{score}}$: Función objetivo para determinar la confianza de la clasificación del objeto.

$$L_{class_{score}} = \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{C \in classes} (\rho_i(C) - \hat{\rho}_i(C))^2$$

Dónde:

- 1_{ij}^{obj} : es 1 si la caja j se encuentra en la celda i ; 0 en otro caso.
- (x, y) y (w, h) : son las coordenadas, la altura y el ancho de la caja predicha.
- $(\hat{x}, \hat{y})$ y $(\hat{w}, \hat{h})$: son las coordenadas, la altura y el ancho de la caja etiquetada como verdadera.
- λ_{coord} y λ_{noobj} : son constantes de regularización.
- B : es el número de cajas encontradas.
- S : Es el número de celdas en la grilla sobre la que se divide la imagen.
- C : Es el puntaje de confianza modulado por la presencia del objeto.
- $\hat{C}$: Es el coeficiente IOU de la caja predicha con la caja etiquetada.
- $\rho_i(C)$: corresponde a la clase predicha.
- $\hat{\rho}_i(C)$: corresponde a la clase etiquetada como verdades.

3. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, en donde la distancia del coseno de la clasificación se calcula con la siguiente fórmula:

$$similarity(A, B) = \cos(\theta) = \frac{A \cdot B}{||A|| ||B||}$$

Dónde:

- θ es el ángulo entre los vectores
- $A \cdot B$ es el producto punto entre los vectores A y B

$$A \cdot B = \sum_{i=1}^n A_i B_i = A_1 B_1 + A_2 B_2 + \dots + A_n B_n$$

- $||A||$ representa la norma L2:

$$||A|| = \sqrt{\sum_{i=0}^n |A_i|^2} = \sqrt{A_1^2 + A_2^2 + \dots + A_n^2}$$

En esta patente utilizamos la distancia del coseno en dos situaciones:

- Clasificación: Cuando los vectores A y B representan los embeddings de texto y los embeddings de imagen que se busca comparar para la clasificación.

- Seguimiento: Cuando los vectores A y B representan los embeddings de imagen almacenados en el diccionario de seguimiento de frames anteriores y los embeddings de imagen obtenidos en el frame actual.
4. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, en donde la distancia de Jaccard, de la etapa de monitoreo así como su derivada, miden la similitud entre dos conjuntos de muestras, y su desplazamiento en el tiempo, y se calculan con la siguiente ecuación:

$$sim_j(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

y determinan el movimiento bounding box de la detección de un mismo objeto a través del tiempo, arrojando un conjunto de datos en los cuales se determina si la unión es cercana a uno (1.0) para los objetos que no tuvieron movimiento y están detenidos en una misma posición.

Finalmente, para medir esta detención se realiza el cálculo de la derivada entre la distancia de Jaccard y el tiempo $\frac{d(sim_j(A, b))}{d(t)}$. Calculando en estado constante la tendencia a cero de la derivada.

5. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, en donde la base de datos consiste en 15 vídeos tomados en condiciones naturales a diferentes horas del día y tipos de iluminación variables de acuerdo con las condiciones climáticas, grabados a una velocidad de 500 fps y una velocidad de reproducción de 30 fps. Cada video genera una etiqueta corroborada por el usuario en donde se detectan las pausas generadas por el (los) colibrí(es) presente(s) y clasificar aquellos en los que hay más de un colibrí en el video.
6. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, que además comprende una estación base caracterizada por una unidad gráfica de procesamiento (GPU) Nvidia T4 con 16 Gb de RAM, una CPU Intel(R) Xeon(R) con 2 núcleos, 12 Gb de RAM y 2 Tb de espacio en disco y cuyo proceso corresponde a:

Recibir la señal enviada por una cámara de punto fijo, con una velocidad de grabación de 30 fps.

Obtener y almacenar el video en la carpeta de destino seleccionado.

Monitorear las pausas en vuelo de forma offline a través del método aquí propuesto.

Almacenar un archivo de texto, comma-separated values file (.csv), que contenga 7 columnas, definidas como: i) frame: el número de frame dónde se genera una detección, ii) bbox: el bounding box con la detección del objeto, iii) class: la clase asignada por el método de clasificación, iv) track: el número de identificación dado por el módulo de seguimiento, v) iou: la distancia de Jaccard entre el frame actual y el siguiente, vi) dy/dx: la derivada de la distancia de Jaccard con respecto al tiempo y, en caso de monitorear una pausa en ese frame, vii) pause: una variable booleana definida como "True" identificando una posible pausa.

7. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, que comprende un módulo de procesamiento de video el cual ejecuta un conjunto de instrucciones implementadas por computador para la captura de un intervalo de video; descomposición en frames de video, validación del tamaño (alto y ancho) de frame así como de la velocidad de grabación medida en frames por segundo mínimas necesarias para el análisis.
8. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, que comprende un método para la detección de pausas en vuelos de colibríes caracterizado por:

Un módulo de detección de objetos a partir de un modelo de redes neuronales convolucionales, con extracción de características de capas intermedias, que incluye:

Un primer entrenamiento realizado en la base de datos benchmark, con 330.000 imágenes, 1.5 millones de objetos identificados y 80 categorías, que contiene entre sus clases (identificada como 14) la detección de colibríes y

Una optimización de la detección mediante un re-entrenamiento realizando la transferencia de aprendizaje y sintonización de pesos priorizando la detección de colibríes vs no colibríes.

9. Un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, que comprende etapas para la clasificación de objetos basados en la obtención de características a partir de un modelo transformers multimodal (texto - imagen). Este módulo se caracteriza por:

Tener una clase base de clasificación: colibrí;

Recibir clases adicionales como entrada de texto por parte del usuario, por ejemplo: flores, hojas, alimentadores de aves.

Recortar las imágenes en regiones de interés obtenidas por las cajas de detección;

Obtener características de texto de las clases de interés;

Obtener características de imagen de las regiones de interés y

Comparar a través de la distancia del coseno las características texto - imagen, minimizando la distancia para determinar la clase a la que pertenece cada objeto.

10. Un método de cálculo de las frecuencias de vuelo de colibríes, de acuerdo con la reivindicación 1, que contempla la etapa de seguimiento del vuelo de los colibríes. Este análisis de frecuencia-tiempo se calcula como una colección de múltiples transformadas discretas de fourier a través de ventanas de intervalos de la señal, representadas por la función dada a continuación [23].

$$X_{STFT}[m,n] = \sum_{k=0}^{L-1} x[k]g[k-m]e^{-j2\pi nk/L}$$

Dónde:

$x[k]$: Corresponds to the signal

$g[k]$: denota la ventana de análisis.

L : denota la longitud de $g[k]$

Y un método de monitoreo de pausas en vuelo de colibríes, de acuerdo con la reivindicación 1, que comprende etapas para realizar el seguimiento del vuelo de los colibríes durante un intervalo de tiempo a través de los frames de video utilizando extracción de embeddings. Estas características (embeddings) se retroalimentan de los obtenidos en la etapa de clasificación, algoritmo transformers bimodal (texto-image) tomando únicamente las características de imagen, las cuales son almacenadas en un diccionario teniendo como llave (key) el track de identificación y la clase del elemento. Seguidamente se comparan entre frames subsiguientes a través de la distancia del coseno de los embeddings almacenados y los embeddings del frame actual.