

Benestar Animal. Monitorització del Comportament d'aus de granja amb Visió per Computador



CIDAI-POC-2026-03

Drets reservats. Aquest treball està disponible sota la llicència Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International ([CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)).

Segons els termes d'aquesta llicència, podeu copiar, redistribuir i adaptar l'obra amb fins no comercials, sempre que l'obra sigui citada adequadament, tal com s'indica a continuació.

En qualsevol ús d'aquest treball, no s'ha de suggerir que el CIDAI doni suport a cap organització, producte o servei específic. No es permet l'ús del logotip CIDAI.

Si adapteu l'obra, heu de llicenciar-la amb la mateixa llicència Creative Commons o equivalent.

Si creeu una traducció d'aquest treball, heu d'afegir la següent exempció de responsabilitat juntament amb la cita suggerida: "Aquesta traducció no la va crear el Centre of Innovation for Data tech and Artificial Intelligence (CIDAI). CIDAI no es fa responsable del contingut ni de l'exactitud d'aquesta traducció. L'edició original en català serà l'edició autèntica i vinculant".

Qualsevol mediació relacionada amb disputes derivades de la llicència es durà a terme d'acord amb les normes de mediació de la [World Intellectual Property Organization](https://www.wipo.int/mediation/).

Cita suggerida. CIDAI-POC-2026-03 // Benestar Animal - Monitorització del Comportament de Pollastres amb Visió per Computador per Biel Ramon Polo i Coen Antens, CVC, CIDAI, 2026.Llicència: [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1.	ABSTRACT	4
2.	DESCRIPCIÓ DEL PROBLEMA QUE RESOL	9
3.	IMPLEMENTACIÓ DE LA SOLUCIÓ	14
4.	POTENCIAL IMPACTE DE LA SOLUCIÓ	18
5.	BIBLIOGRAFIA	19



1. Abstract

1.1. Objectius

En algun moment de la dècada del 2010, el consum de carn de pollastre va superar el porc, per tant la carn més produïda del món. Des de llavors, el seu consum no ha deixat de créixer.

Cada any, la indústria avícola sacrifica uns 75.000 milions d'aus arreu del món per a la producció de carn, dels quals 9,5 mil milions només als Estats Units. Dit d'una altra manera, això equival a aproximadament 206 milions d'aus cada 24 hores [1]. Aquestes xifres reflecteixen la magnitud d'un sector clau en l'alimentació global, però també posen sobre la taula una qüestió cada vegada més present: en quines condicions viuen aquests animals i fins a quin punt es pot millorar el seu benestar. Davant d'aquesta realitat, es vol cercar noves maneres d'entendre i millorar les condicions en què viuen aquests animals.

Figura 1. L'etiqueta de benestar animal a la cantonada superior dreta.



El projecte pretén aportar una eina que permeti analitzar el comportament de les aus de manera objectiva, amb la finalitat de detectar patrons que ajudin a identificar si les condicions de vida afavoreixen o perjudiquen el seu benestar. La millora d'aquestes condicions té un impacte directe en la salut i dignitat dels animals, repercuteix en una alimentació més responsable i de major qualitat, i optimitza la producció en gallines ponedores. Actualment, el segell de benestar animal ja s'integra en productes comercials com a factor diferenciador. S'observa, per exemple, en l'etiquetatge de Bon Àrea (Figura 1), on es mostra el certificat de benestar animal d'AENOR. A partir d'aquesta motivació, es defineixen uns objectius que orienten el desenvolupament tecnològic cap a una finalitat ètica i pràctica: entendre millor el comportament animal per avançar cap a una producció alimentària més conscient i sostenible.

1.2. Visió

L'objectiu principal del projecte és el disseny i desenvolupament d'un sistema automatitzat de monitorització del comportament d'aus mitjançant tècniques de Visió per Computador. El sistema ha de ser capaç de captar i analitzar dades de seqüències de vídeo, identificar patrons de moviment i extreure mètriques quantificables per a una avaluació objectiva.

Objectius específics:

- Revisar l'estat de l'art en Visió per Computador aplicada a la detecció d'objectes i definir el conjunt de dades necessari, incloent criteris d'anotació i adequació al projecte.
- Implementar tècniques bàsiques per detectar i seguir individus en vídeo (OpenCV i models d'IA) i analitzar-ne els patrons de moviment (desplaçaments, agrupacions i zones d'activitat).
- Generar mètriques i estadístiques per avaluar objectivament el comportament i el benestar dels animals, i validar els resultats discutint limitacions i possibles millores del sistema.

1.3. Resum del problema

La producció avícola actual es caracteritza per la presència d'un gran nombre d'animals en espais compartits, fet que dificulta la monitorització contínua i objectiva del seu comportament. Tot i que el benestar animal és un aspecte cada vegada més rellevant tant des del punt de vista ètic com productiu, els mètodes tradicionals d'observació sovint depenen de la supervisió humana o de sistemes limitats que no permeten analitzar de manera sistemàtica els patrons d'activitat dels animals. Aquesta situació fa difícil detectar de manera precoç possibles anomalies en el comportament, com canvis en la mobilitat, concentracions excessives d'individus o períodes prolongats d'inactivitat, que poden ser indicadors de problemes de salut o de condicions ambientals inadequades (Figura 2).

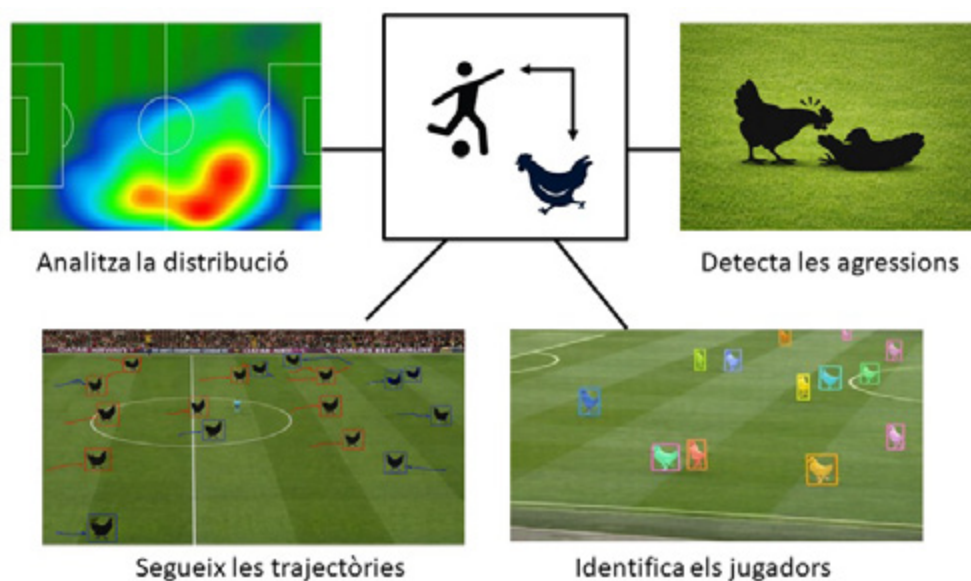
Figura 2. Exemple d'un fotograma



En aquest context, l'anàlisi del comportament animal pot establir-se, de manera anàloga, amb l'estudi del rendiment en un partit de futbol, on es registren de forma sistemàtica els moviments, les interaccions i les accions de cada jugador al llarg del temps. De manera similar, en una granja avícola és possible interpretar l'activitat dels individus a partir de patrons com el desplaçament, l'accés als recursos o les interaccions socials, incloent-hi conductes agressives associades a la competència o a situacions d'estrès. Així com en l'àmbit esportiu l'anàlisi detallada del comportament permet identificar desequilibris tàctics o físics, en el context avícola aquests patrons poden actuar com a indicadors clau de l'estat de benestar dels animals (Figura 3). En aquest sentit s'ha volgut estendre la metàfora que es va utilitzar també al PAI d'exploració de la detecció de dades biomètriques en partits de futbol que es va fer amb HUB23, CVC, EURECAT, Kognia Sports i CIDAI.

En aquest sentit, sorgeix la necessitat de desenvolupar eines tecnològiques capaces d'analitzar el comportament de les aus de manera automatitzada i objectiva. L'ús de tècniques de Visió per Computador permet aprofitar les dades de vídeo capturades en les instal·lacions per identificar els animals, seguir-ne el moviment al llarg del temps i extreure mètriques quantitatives sobre la seva activitat. Aquest tipus de sistemes pot facilitar una millor comprensió dels patrons de comportament i contribuir al desenvolupament de metodologies més eficients per avaluar i millorar el benestar animal dins de les explotacions avícoles.

Figura 3. La metàfora del futbol.



1.4. Solució

Aquest treball proposa el desenvolupament d'un sistema automatitzat de monitorització del comportament d'aus basat en tècniques de Visió per Computador. La proposta se centra en la construcció d'un pipeline de processament de vídeo capaç d'identificar individus, seguir-ne el moviment al llarg del temps i extreure mètriques quantitatives que permetin analitzar patrons de comportament relacionats amb el benestar animal.

El sistema plantejat parteix de la captura de vídeo mitjançant càmeres fixes instal·lades a l'interior de la granja. Aquestes càmeres proporcionen una seqüència contínua d'imatges que reflecteixen l'activitat dels animals dins de l'espai de cria. Tècniques clàssiques de processament d'imatge com el [Background Subtraction](#) es poden utilitzar com a mecanisme preliminar per detectar zones en moviment i obtenir una primera estimació de l'activitat global del grup.

Tanmateix per millorar la detecció d'objectes es proposen models basats en [Deep Learning](#) amb la finalitat d'identificar la posició dels individus dins de la imatge. Models de detecció en temps real, com els basats en l'arquitectura YOLO, permeten obtenir de manera eficient les regions que corresponen als animals presents en cada instant. Aquesta etapa proporciona informació espacial sobre la localització dels individus, però de manera estrictament instantània, ja que cada detecció és independent de les anteriors.

Per incorporar la dimensió temporal necessària per analitzar el comportament, el sistema integra posteriorment un mètode de Mul-

ti-Object Tracking. Aquesta capa de seguiment té com a objectiu associar les deteccions corresponents a un mateix individu al llarg de diferents fotogrames, permetent així reconstruir trajectòries de moviment. La proposta contempla l'ús de models de llibreries de seguiment com Norfair, que basen el procés d'associació principalment en criteris de proximitat espacial entre deteccions consecutives. Aquest tipus de metodologia resulta especialment adequat en escenaris amb recursos computacionals limitats, ja que permet implementar sistemes de seguiment robustos sense necessitat de models d'aparença complexos.

Malgrat això, en entorns on diversos individus estan molt propers o entorns amb oclusions parcials, la informació purament espacial sovint és insuficient per mantenir correctament les identitats. Per aquest motiu, s'ha proposat la incorporació de descriptors visuals d'aparença (embeddings). En particular, descriptors com el Histogram of Oriented Gradients permeten capturar informació sobre la forma i la textura dels objectes detectats, proporcionant una representació compacta que pot utilitzar-se per comparar deteccions entre fotogrames consecutius. La combinació d'informació espacial i visual permet reforçar el procés d'associació i millorar la robustesa del seguiment en situacions amb ambigüitat.

Un cop reconstruïdes les trajectòries individuals, el sistema permet obtenir diverses mètriques quantitatives relacionades amb el comportament dels animals. A partir de la posició i el moviment dels individus al llarg del temps es poden estimar indicadors com el nivell d'activitat global del grup, la velocitat mitjana de desplaçament o la distribució espacial dins de la instal·lació. De la mateixa manera, l'anàlisi temporal de les trajectòries pot facilitar la detecció de si-

tuacions potencialment problemàtiques, com períodes prolongats d'immobilitat o concentracions excessives d'animals en determinades zones.

Finalment, les dades generades pel sistema poden ser agregades i representades mitjançant eines de visualització que permetin interpretar els patrons d'activitat observats. Aquest tipus de representacions, com ara mapes de calor de moviment o gràfiques d'evolució temporal de l'activitat, faciliten la comprensió dels resultats per part d'investigadors, veterinaris o gestors d'explotacions. D'aquesta manera, la informació obtinguda mitjançant tècniques de visió per computador pot transformar-se en coneixement útil per a la presa de decisions relacionades amb la gestió i el benestar dels animals.

2. Descripció del problema que resol

El seguiment i l'anàlisi del comportament animal són dues eines importants per avaluar el benestar dels animals en entorns de producció ramadera. En el cas de la producció avícola, la detecció precoç de canvis en els patrons d'activitat, la distribució espacial o la mobilitat dels individus pot proporcionar informació rellevant sobre possibles problemes de salut, estrès ambiental o condicions inadequades dins de la instal·lació. Tradicionalment, aquesta tasca s'ha dut a terme mitjançant observació directa per part de personal especialitzat. Tanmateix, aquests mètodes presenten diverses limitacions, ja que sovint requereixen una elevada dedicació de temps, poden introduir biaixos subjectius i resulten difícils d'escalar en explotacions de gran dimensió.

En els darrers anys, les tècniques de Visió per Computador han emergit com una alternativa per automatitzar aquest tipus d'anàlisi. L'ús de càmeres permet capturar de manera contínua l'activitat dels animals sense interferir en el seu comportament natural, mentre que els algoritmes d'anàlisi d'imatge poden processar aquestes dades de manera automàtica. Tot i això, la simple detecció d'animals en una imatge no és suficient per obtenir una comprensió completa del seu comportament. Molts sistemes existents es limiten a identificar la presència d'individus en fotogrames aïllats, proporcionant informació essencialment instantània que no permet estudiar l'evolució del moviment ni els patrons temporals d'activitat.

Una altra dificultat rellevant són les característiques pròpies dels entorns avícoles. En aquest tipus d'escenaris és habitual trobar un gran nombre d'individus visualment molt similars que es mouen dins d'un espai compartit. Aquesta situació genera problemes d'oclusions, proximitat entre animals i canvis constants en la confi-

guració espacial del grup. Com a conseqüència, mantenir la identitat dels individus al llarg del temps resulta ser una tasca complexa per als sistemes automàtics de seguiment.

En aquest context, el problema que aborda aquest treball consisteix a desenvolupar un sistema capaç d'identificar i seguir automàticament els animals dins d'un entorn de granja a partir de dades de vídeo, amb l'objectiu d'obtenir informació temporal sobre el seu comportament. Aquest sistema ha de ser prou robust per operar en escenaris amb múltiples individus visualment similars, mantenint una identificació coherent al llarg del temps i permetent la reconstrucció de trajectòries de moviment.

2.1. Visió per Computador Clàssica

En una fase inicial del projecte es van explorar tècniques clàssiques de Visió per Computador amb l'objectiu d'obtenir una primera aproximació a la detecció dels animals. En primer lloc, es va intentar extreure el fons dels vídeos mitjançant mètodes estadístics, com la mitjana i la mediana de frames consecutius. No obstant això, els resultats no van ser satisfactoris a causa de la curta durada dels vídeos i del moviment limitat dels animals, que provocava que part de les aus quedessin incorporats en el fons estimat. Com a alternativa, es va generar un fons aproximat utilitzant tècniques d'intel·ligència artificial generativa, que van permetre obtenir una representació més neta de l'escena (Figura 4).

A partir d'aquest fons de referència, es va aplicar [Background Subtraction](#) per obtenir les siluetes de les aus. Posteriorment, es va experimentar amb la segmentació [Watershed](#), utilitzant prèviament el [Distance Transform](#) per identificar els centres dels objectes i definir marcadors interns. Tot i que aquesta aproximació permetia separar parcialment els individus, els resultats es veien limitats per les freqüents aglomeracions entre animals. Aquesta exploració va servir com a prova de concepte i va permetre identificar les limitacions dels mètodes clàssics en escenaris amb alta densitat d'animals, fet que va motivar l'adopció posterior de tècniques basades en aprenentatge profund.

Figura 4. Background.



2.2. Detecció de les aus amb Deep Learning

A partir de les limitacions observades en els enfocaments de visió clàssica, es va iniciar una nova fase del projecte centrada en la preparació de dades i l'entrenament d'un model de Deep Learning per a la detecció i segmentació d'aus. Per entrenar el model, es van extreure frames individuals dels vídeos enregistrats, seleccionats de manera que quedessin distribuïts de forma equitativa al llarg de la durada dels vídeos amb l'objectiu d'obtenir imatges el més diverses possible.

Les imatges resultants es van etiquetar manualment mitjançant el programari web [CVAT](#) (Computer Vision Annotation Tool). En lloc d'utilitzar únicament caixes delimitadores (bounding boxes), es van definir contorns o màscares de segmentació per representar amb més precisió la forma dels animals. Aquest procés d'anotació es va dur a terme manualment per garantir la qualitat del conjunt de dades. Finalment, es van crear dos subconjunts diferenciats: un destinat a l'entrenament del model i un altre reservat per a la validació.

Figura 5. Detecció d'objectes



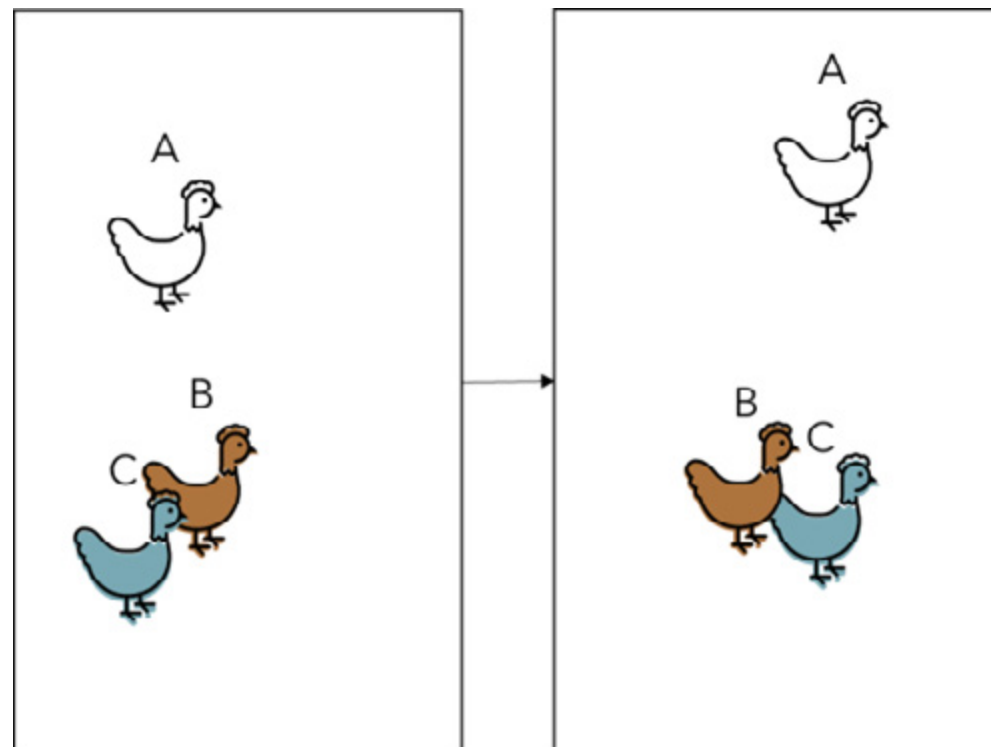
Un cop preparades les dades, es va procedir a l'entrenament d'un model de detecció utilitzant YOLOv8 (You Only Look Once [3]), una arquitectura àmpliament utilitzada en tasques de detecció d'objectes en temps real. En concret, es va partir del model pre-entrenat [yolov8n-seg.pt](#) amb l'objectiu d'aprofitar el coneixement previ del model i millorar la convergència del procés d'entrenament, especialment tenint en compte la mida limitada del conjunt de dades disponible. Després de diverses iteracions d'entrenament i ajust dels híper-paràmetres, es va obtenir un model capaç de detectar els animals amb un nivell de precisió satisfactori en el conjunt de validació. Les deteccions produïdes per aquest model constitueixen la base per a l'etapa posterior de seguiment (tracking) dels animals al llarg del temps (Figura 5).

2.3. Conceptes bàsics de tracking i correspondència d'objectes

Abans d'integrar mètodes de tracking més avançats, es va dur a terme una prova prèvia basada en descriptors clàssics i mètodes d'assignació òptima. L'objectiu era validar si era possible establir correspondències robustes entre deteccions d'aus en frames consecutius combinant informació espacial i visual.

Tot i que alguns models de detecció proporcionaven resultats fiables, no sempre incorporaven un sistema intern de Re-Identification. Això feia que, en situacions d'alta densitat o quan els animals s'ocultaven parcialment, dues deteccions consecutives poguessin intercanviar la seva identitat o detectar-se com un nou individu. Per mitigar aquest problema, es va experimentar amb descriptors HOG (Histogram of Oriented Gradients) per capturar informació visual de les aus, i amb l'algoritme hongarès (Hungarian Algorithm) per resoldre l'assignació òptima entre deteccions de frames consecutius. Aquests dos criteris es combinaven en una matriu de costos ponderada que integrava la distància espacial i la similitud visual (Figura 6).

Figura 6. Explicar la necessitat del Reld.



Aquesta prova va permetre validar el potencial del sistema d'associació i va servir com a base conceptual per a versions híbrides de tracking desenvolupades posteriorment. Tot i que el mètode no era prou robust per utilitzar-lo directament en producció, va resultar clau per comprendre la importància de combinar informació espacial i visual en el seguiment d'animals en escenaris amb alta densitat.

2.4. Algorismes de seguiment d'objectes

Un cop es disposa d'un model de segmentació [yolov8-seg](#) entrenat i validat, aquesta fase del projecte se centra en l'avaluació del seu rendiment sobre imatges i vídeos reals, així com en la integració de diferents mètodes de tracking per al seguiment temporal de les aus.

Amb l'objectiu de donar continuïtat temporal a les deteccions i poder seguir cada individu al llarg del temps, es van integrar diversos algorismes de tracking-by-detection. Aquests mètodes permeten associar deteccions entre frames consecutius a partir de la sortida del model YOLO. Concretament, s'han avaluat quatre enfocaments recents: Norfair, DeepSORT, BoT-SORT i ByteTrack. Aquests sistemes permeten identificar i mantenir la identitat de cada au dins del vídeo, generant trajectòries individuals que poden ser utilitzades posteriorment per a l'anàlisi de comportament, moviment o interaccions.

A continuació es descriuen breument les característiques principals de cada mètode:

1. Norfair

- Basat en l'associació per distància (principalment distància Euclidiana) entre deteccions.
- Pot integrar informació de moviment mitjançant Optical Flow.
- Destaca per la seva simplicitat i eficiència computacional, fet que el fa adequat per a prototipatge ràpid.
- Tanmateix, pot presentar limitacions en escenaris amb oclusions prolongades o alta densitat d'objectes.

2. DeepSORT [5]

- Extensió de SORT que incorpora embeddings d'aparença (Re-Identification o ReID).
- Permet recuperar la identitat d'objectes després d'occlusions parcials o totals.
- Ofereix trajectòries més consistents en entorns complexos.
- Com a inconvenient, incrementa el cost computacional i depèn de la qualitat del model de ReID.

3. BoT-SORT [6]

- Millora de DeepSORT que incorpora estratègies més robustes d'associació.
- Utilitza un filtre de Kalman refinat i mecanismes de compensació de moviment de càmera.
- Mostra un millor rendiment en escenaris amb oclusions freqüents i moviment complex.
- Proporciona trajectòries més estables i consistents.

4. ByteTrack [7]

- Algorisme recent que introdueix l'ús de deteccions de baixa confiança en el procés d'associació.
- Permet recuperar objectes que podrien perdre's en mètodes tradicionals.
- Destaca especialment en entorns amb deteccions sorolloses o inconsistents.
- Ofereix un bon equilibri entre rendiment i eficiència computacional.

3. Implementació de la solució

3.1. Arquitectura, tecnologia i dades utilitzades

Algorisme 1: Millora del seguiment basada en Norfair, HOG i Algoritme Hongarès

Tot i els bons resultats obtinguts amb els mètodes de tracking estàndard, es va desenvolupar una versió millorada del sistema de seguiment basada en Norfair, incorporant un mecanisme propi de Re-Identification (ReID) basat en descriptors Histogram of Oriented Gradients (HOG) i redefinint el procés d'associació mitjançant l'Algoritme Hongarès.

Aquest enfocament híbrid combina dos criteris principals per a l'associació d'objectes entre frames consecutius:

Proximitat espacial, basada en la distància entre posicions (centres de les deteccions).

Similitud visual, mitjançant la comparació de descriptors HOG.

La combinació d'aquests dos factors permet reduir els canvis d'identitat (ID switches) i millorar la coherència temporal de les trajectòries, especialment en situacions complexes com oclusions parcials o interaccions entre múltiples aus. En la pràctica, aquest mètode ha estat el que ha proporcionat els millors resultats globals, aconseguint un bon equilibri entre eficiència computacional i estabilitat del seguiment. Els errors observats es deuen principalment a fallades puntuals en el model de detecció (per exemple, màscares incorrectes o absents en alguns frames), més que no pas a la lògica del sistema de tracking.

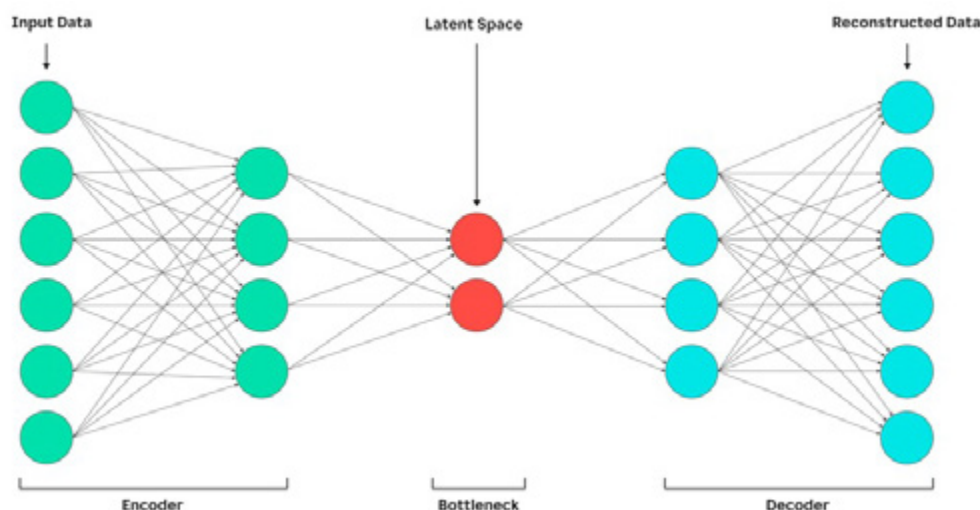
Algorisme 2: Millora del seguiment basada en Norfair i Auto-Encoder

Com a alternativa als descriptors manuals, es va implementar un segon mètode de seguiment que combina Norfair amb un sistema de Re-Identification basat en Deep Learning, utilitzant un Auto-Encoder convolucional entrenat específicament amb retalls d'imatges d'aus.

Per a l'entrenament del model, es van extreure automàticament regions d'interès a partir de les deteccions generades per YOLO-v8-seg. L'Auto-Encoder aprèn una representació latent compacta (embedding) que captura les característiques visuals més rellevants dels animals de manera no supervisada.

Durant el tracking, cada detecció es descriu mitjançant la seva posició espacial centre de la bounding box) i un embedding visual obtingut amb l'Auto-Encoder. L'estructura del qual es pot observar a la Figura 7. Els resultats experimentals indiquen que aquest mètode obté un rendiment molt similar al sistema basat en HOG i Algoritme Hongarès. Tot i no superar-lo clarament, valida l'ús de descriptors apresos automàticament com una alternativa viable als descriptors manuals, especialment en escenaris on es disposa de grans volums de dades.

Figura 7. AutoEncoder



3.2. Reptes resolts i resultats obtinguts

En aquesta secció es presenten els resultats obtinguts a partir del sistema de detecció i seguiment desenvolupat. En particular, s'ha dut a terme una comparació quantitativa dels diferents mètodes de tracking utilitzats al llarg del projecte: Norfair combinat amb HOG i Algoritme Hongarès, DeepSORT, BoT-SORT, ByteTrack i Norfair combinat amb Auto-Encoder.

La comparació s'ha realitzat utilitzant mètriques estàndard en problemes de Multi-Object Tracking (MOT):

- **MOTA (Multiple Object Tracking Accuracy):** avalua la precisió global del sistema, tenint en compte els falsos positius (FP), els falsos negatius (FN) i els canvis d'identitat (ID Switch).
- **MOTP (Multiple Object Tracking Precision):** mesura la precisió espacial del seguiment, indicant com d'ajustades són les deteccions respecte a la veritat de referència (Ground Truth).
- **IDF1:** avalua la consistència en l'assignació d'identitats al llarg del temps.

Els falsos positius (FP) corresponen a deteccions incorrectes d'objectes inexistents, mentre que els falsos negatius (FN) representen objectes reals que no han estat detectats o correctament seguits pel sistema en comparació amb el Ground Truth.

Els resultats quantitatius obtinguts es mostren a la Taula 1, on es pot observar el comportament de cada mètode en termes de precisió i consistència del seguiment.

Taula 1. Resultats

Mètode	MOTA	MOTP	IDF1	Switches	FP	FN
Norfair (HoG)	94,6%	0,015	96,1%	5	130	309
Norfair (AutoEncoder)	94,6%	0,015	9,61%	5	129	308
BoT-Sort	99,7%	0,020	96,8%	11	0	13
ByteTrack	99,7%	0,036	97,3%	10	2	15
DeepSort	93,4%	0,117	92,6%	11	439	87

Més enllà de l'anàlisi global, el sistema desenvolupat permet obtenir mètriques detallades a nivell individual per a cada individu seguit. Entre aquestes destaquen el moviment total, el moviment mitjà per frame i la velocitat màxima associada a cada identificador.

La Figura 8 mostra, a mode d'exemple, la trajectòria seguida per l'au amb ID número 22. S'observa com, després d'un període de repòs, l'animal es desplaça per alimentar-se i posteriorment torna a una posició de descans diferent. La forma de la trajectòria reflecteix el patró de moviment característic de les aus, amb desplaçaments irregulars i pauses freqüents. Aquest tipus d'anàlisi individual resulta especialment rellevant en el context del benestar animal, ja que permet identificar patrons anòmals de comportament, com ara reduccions significatives de mobilitat o canvis sobtats en l'activitat d'un individu concret.

Figura 8. Seguiment d'un individu



3.3. Limitacions actuals de la tecnologia

Una de les principals limitacions del sistema desenvolupat és la forta dependència de la qualitat de la detecció visual. Tot i la integració de mètodes avançats de tracking i mecanismes de Re-Identification (ReID), un sistema basat exclusivament en visió per computador no és suficient per garantir un seguiment robust en totes les situacions. En particular, la gran similitud visual entre les aus dificulta significativament la tasca de diferenciació entre individus. Malgrat l'ús de descriptors com HOG o embeddings generats amb Deep Learning, aquests no sempre són capaços de capturar característiques prou discriminatives, fet que pot derivar en canvis d'identitat (ID switches), especialment en escenaris amb oclusions o alta densitat d'animals.

A més, el rendiment del sistema es veu fortament afectat per les condicions d'il·luminació. En entorns nocturns, la qualitat de les imatges disminueix de manera significativa, comprometent tant la detecció com el tracking. Tot i que l'ús d'il·luminació infraroja pot permetre mantenir un cert nivell de funcionament, aquesta solució introdueix noves dificultats, ja que redueix encara més la informació visual disponible per distingir entre individus. En conjunt, aquestes limitacions evidencien que, per assolir un seguiment realment robust en entorns reals, seria necessari complementar la visió per computador amb altres fonts d'informació o sensors addicionals.



4. Potencial impacte de la solució

Els resultats obtinguts al llarg d'aquest treball demostren que és possible desenvolupar un sistema eficaç de detecció i seguiment del comportament de les aus mitjançant tècniques de Visió per Computador i Deep Learning, fins i tot en escenaris amb alta densitat d'animals i oclusions freqüents. Els resultats mostren que BoT-SORT i ByteTrack obtenen les millors puntuacions globals en termes de MOTA (99.7%), amb un nombre molt reduït de falsos positius i falsos negatius. En particular, BoT-SORT presenta una excel·lent capacitat de filtratge, assolint zero falsos positius i només 13 falsos negatius, mentre que ByteTrack destaca lleugerament en IDF1 (97.3%), indicant una millor consistència en la preservació de les identitats al llarg del temps.

En contraposició, DeepSORT mostra un rendiment inferior en aquest escenari, amb una MOTA del 93.4% i un nombre elevat de falsos positius (439), fet que indica una menor capacitat davant deteccions errònies i oclusions freqüents. Aquests resultats posen de manifest que l'ús de Reidentification embeddings no garanteix necessàriament un millor seguiment quan les condicions visuals són adverses o quan el model de detecció presenta errors puntuals.

D'altra banda, la combinació basada en Norfair amb descriptors HOG i l'Algoritme Hongarès ofereix un equilibri molt competitiu entre precisió, estabilitat d'identitats i cost computacional. Tot i no assolir les puntuacions màximes de MOTA, aquest enfocament presenta un IDF1 elevat (96.1%) amb un nombre reduït de canvis d'identitat, demostrant que la integració d'informació visual clàssica amb criteris espacials pot millorar significativament la consistència del tracking sense necessitat de models complexos o costosos.

Els resultats obtinguts amb Norfair + Auto-Encoder són pràcticament idèntics als del sistema Norfair + HOG + Hungarian Algorithm, tant en MOTA (94.6%) com en IDF1 (96.1%) i nombre de canvis d'identitat. Això indica que, en aquest escenari, l'ús de representacions latents apreses no aporta una millora clara respecte als descriptors visuals clàssics, però sí que valida l'Auto-Encoder com una alternativa viable i flexible, amb un comportament estable i un cost computacional similar. Aquest resultat reforça la idea que, quan el model de detecció és consistent, la combinació d'informació espacial amb descriptors visuals, ja siguin manuals o apresos, pot oferir un seguiment robust sense necessitat de recórrer a mètodes més complexos o exigents en recursos com BoT-SORT o ByteTrack.

A nivell d'aplicació, el sistema desenvolupat ha permès no només seguir individus de manera fiable, sinó també extreure mètriques de comportament rellevants, com ara el moviment total, la velocitat màxima o la distribució espacial de l'activitat. El desequilibri de moviment observat entre el costat esquerre i dret de l'escena evidencia el potencial del sistema per detectar patrons col·lectius de comportament que podrien estar relacionats amb factors ambientals, de disseny de l'espai o de benestar animal.

Aquest treball valida que una solució híbrida, combinant models de detecció basats en Deep Learning amb estratègies de tracking eficients, pot oferir resultats d'alta qualitat amb un cost computacional assumible. Això obre la porta a futures línies de recerca orientades a l'anàlisi automàtica del benestar animal, la detecció primerenca de comportaments anòmals i la integració del sistema en entorns reals de producció avícola.



5. Bibliografía

- [1] S. Millstein, “¿Cuántos pollos se matan cada año?”, Sentient Media, Feb. 2025. [Online]. Available: <https://sentientmedia.org/es/cuantos-pollos-se-matan-cada-ano/>
- [2] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.
- [3] You Only Look Once: Unified, Real-Time Object Detection Redmon, J., et al. (2016)
- [4] Simple Online and Realtime Tracking with a Deep Association Metric Wojke, N., Bewley, A., & Paulus, D. (2017).
- [5] Simple Online and Realtime Tracking Bewley, A., et al. (2016).
- [6] ByteTrack: Multi-Object Tracking by Associating Every Detection Box Zhang, Y., et al. (2022).
- [7] BoT-SORT: Robust Associations Multi-Pedestrian Tracking Aharon, N., et al. (2022).

Socis



Membres

