

Monitorització del Comportament de Pollastres amb Visió per Computador

Biel Ramon Polo

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

A vegades, no ens aturem a pensar d'on prové el menjar que consumim ni en les condicions en què viuen els animals que el fan possible. Aquest treball de fi de grau sorgeix de la necessitat de posar èmfasi en el benestar animal i en la qualitat dels aliments que ingerim. Els pollastres representen una part essencial de la nostra alimentació, i cada dia es sacrifiquen aproximadament 206 milions d'exemplars en tot el món per a la producció d'aliments."[12]" Davant d'aquesta realitat, és molt important cercar noves maneres d'entendre i millorar les condicions en què viuen aquests animals.

El projecte pretén aportar una eina que permeti analitzar el comportament dels pollastres de manera objectiva, amb la finalitat de detectar patrons que ajudin a identificar si les condicions en què viuen afavoreixen o perjudiquen el seu benestar. Millorar aquestes condicions no només té un impacte directe en la salut i la dignitat dels animals, sinó que també repercuteix en una alimentació més responsable i de major qualitat per a les persones, a més de millorar la qualitat i quantitat dels ous de les gallines ponedores.



Figura 1 Pollastre del bonArea mostrant l'etiqueta de benestar animal a la cantonada superior dreta.

El segell de benestar animal ja consta en alguns productes per diferenciar-los dels altres, com es veu a la Figura 1. 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.

2 OBJECTIUS

El projecte té com a objectiu principal el disseny i desenvolupament d'un sistema automatitzat de monitorització del comportament de pollastres mitjançant tècniques de Visió per Computador. Aquest sistema ha de ser capaç de captar i analitzar dades provinents de seqüències de vídeo, identificar patrons de moviment i extreure mètriques quantificables que permetin una avaluació objectiva del benestar animal. L'enfocament tecnològic proposat combina processos de detecció i seguiment d'individus amb l'anàlisi computacional de la seva activitat, amb la finalitat d'obtenir indicadors fiables que puguin contribuir a millorar la gestió i les condicions de les instal·lacions avícoles.

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 pollastres 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.

3 METODOLOGIA

Per tal d'assolir els objectius establerts dins del temps disponible, el projecte es desenvoluparà seguint la filosofia Scrum adaptada a una persona, una metodologia àgil àmpliament utilitzada en projectes de recerca i desenvolupament tecnològic. Aquesta elecció permet dividir el projecte en iteracions curtes (Sprints) que faciliten una planificació flexible, l'avaluació contínua dels progressos i l'adaptació ràpida davant possibles dificultats tècniques o canvis d'enfocament.

- E-mail de contacte: bielrp23@gmail.com
- Menció realitzada: Computació
- Treball tutoritzat per: Coen Antens (Àrea de Ciències de la Computació i Intel·ligència Artificial)
- Curs 2025/26

Per gestionar aquesta dinàmica es farà ús de l'eina Jira, que permet definir i visualitzar les tasques en taulers Kanban o Scrum, establir terminis i fer un seguiment dels avenços de forma estructurada i transparent.

4 PLANIFICACIÓ

Sprint 1: Preparació bibliogràfica i de dades (15 setembre – 5 octubre)

- Tasques:
 - Revisió bibliogràfica sobre Visió per Computador i detecció de moviment.
 - Anàlisi de projectes similars i comparació de metodologies utilitzades.
 - Establir una estratègia clara per definir com s'anotaran les dades i assegurar que el conjunt de dades sigui representatiu i útil per als objectius del projecte.
- Temps estimat: 3 setmanes.

Sprint 2: Programació 1 (6 octubre – 16 novembre)

- Tasques:
 - Implementació inicial de mètodes de detecció d'objectes utilitzant OpenCV.
 - Avaluació de models preentrenats per comparar precisió i eficiència.
 - Test inicial sobre un subconjunt del conjunt de dades per comprovar resultats preliminars.
- Temps estimat: 6 setmanes.

Sprint 3: Programació 2 (17 novembre – 14 desembre)

- Tasques:
 - Millora de la detecció i el seguiment (ajust de paràmetres i entrenament amb deep learning si cal).
 - Identificació de patrons bàsics de moviment (zones d'activitat, densitat de desplaçaments, agrupacions).
- Temps estimat: 4 setmanes.

Sprint 4: Programació 3 i Generació d'estadístiques (15 desembre – 18 gener)

- Tasques:
 - Refinament del processament de detecció.
 - Generació d'estadístiques i mètriques sobre activitat i benestar.
 - Representació visual en gràfics i taules.
- Temps estimat: 5 setmanes.

Sprint 5: Validació i redacció informe (19 gener – 3 febrer)

- Tasques:
 - Validació dels resultats i comprovació amb criteris objectius.
 - Discussió de limitacions i propostes de millora.
 - Redacció final de l'informe.
- Temps estimat: 2 setmanes.

La Figura 2 mostra una representació visual de la duració dels sprints.

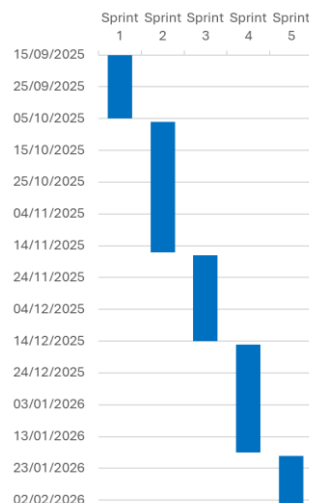


Figura 2 Esquema temporal dels sprints de la planificació

5 ESTAT DE L'ART

L'estudi del comportament animal mitjançant tècniques de Visió per Computador ha esdevingut un camp emergent en la intersecció entre la intel·ligència artificial, l'agricultura i l'ètica del benestar animal. Durant els darrers anys, la disponibilitat de models de Deep Learning i la millora en la capacitat de càlcul han permès automatitzar tasques que abans requerien observació manual o sensors específics.

En el nivell més elemental de complexitat es troben les tècniques clàssiques de processament d'imatge, com el Background Subtraction. Aquestes metodologies permeten identificar regions en moviment mitjançant la comparació entre el fons estàtic i els canvis successius en la imatge. Tot i que no proporcionen informació individualitzada, continuen essent útils com a mecanismes preliminars de segmentació, especialment en entorns controlats amb càmeres fixes i condicions d'il·luminació estables.

Aquest tipus d'enfocament resulta adequat per obtenir indicadors globals, com el nivell general d'activitat o la detecció de canvis bruscos en el comportament del grup, i sovint s'utilitza com a pas previ a sistemes més avançats.

En l'àmbit de la ramaderia avícola, diversos treballs han explorat l'ús de càmeres per monitoritzar els moviments i patrons d'activitat dels pollastres. Sistemes basats en tècniques de detecció d'objectes com YOLO (You Only Look Once), Faster R-CNN o SSD han demostrat ser capaços d'identificar individus en temps real, fins i tot en entorns complexos amb oclusions i il·luminació

variable.“[9], [10]” Aquestes metodologies permeten obtenir mètriques quantitatives com la densitat de moviment, la distribució espacial o la freqüència d'activitat, totes elles indicadors indirectes de benestar. Tot i això, la informació obtinguda és essencialment instantània, ja que no es manté una correspondència directa entre deteccions en fotogrames consecutius.

Per superar aquesta limitació, s'han incorporat mètodes de seguiment multiobjecte (Multi-Object Tracking, MOT), que associen deteccions successives per estimar trajectòries individuals al llarg del temps. “[14]” Aquests mètodes introdueixen una dimensió temporal clau per a l'anàlisi del comportament.

Entre els algoritmes més rellevants destaquen:

- Norfair: “[13]” és una llibreria de seguiment lleugera i flexible que permet implementar sistemes MOT sense necessitat de models d'aparença complexos. Es basa principalment en la distància entre deteccions (per exemple, punts clau o centres d'objectes) i en regles d'associació simples, fet que la fa especialment atractiva per a sistemes amb recursos limitats o quan es vol una solució interpretable i fàcilment adaptable.
- DeepSORT: “[17]” amplia l'algoritme SORT clàssic incorporant descriptors d'aparença extrets mitjançant xarxes neuronals convolucionals. Aquesta combinació de moviment (filtre de Kalman) i aparença permet una associació més robusta d'identitats, especialment en situacions d'oclusió parcial.
- ByteTrack: “[19]” proposa una estratègia eficient que aprofita tant les deteccions d'alta confiança com les de baixa confiança. En lloc de descartar aquestes darreres, les integra en el procés d'associació, millorant el seguiment en escenaris amb deteccions imperfectes o sorolloses.
- BotSort: “[18]” evoluciona a partir de DeepSORT incorporant millores en els descriptors d'aparença i en les estratègies d'associació, amb l'objectiu d'augmentar la robustesa davant canvis d'escala, oclusions prolongades i variabilitat visual entre individus similars.

En els sistemes de seguiment multiobjecte més lleugers, l'associació entre deteccions en fotogrames consecutius es fonamenta sovint en criteris simples però efectius. Un dels més habituals és la proximitat espacial, que assumeix que la posició d'un individu no varia de manera abrupta entre frames consecutius. Així, la distància euclidiana entre els centres de les deteccions esdevé un indicador clau per mantenir la identitat al llarg del temps.

De manera complementària, es pot incorporar la similitud visual mitjançant descriptors d'aparença, com ara el

Histogram of Oriented Gradients (HoG). Aquest descriptor captura la distribució de gradients d'intensitat de la imatge, proporcionant una representació compacta de la forma i la textura de l'objecte. La comparació entre descriptors HoG de deteccions successives permet reforçar l'associació quan la informació espacial és ambigua, per exemple en situacions de proximitat entre individus o lleugeres oclusions.

La combinació de proximitat espacial i similitud visual ofereix un compromís adequat entre simplicitat computacional i robustesa, especialment en escenaris amb individus visualment similars i recursos de càlcul limitats. En conjunt, aquests algoritmes permeten reconstruir trajectòries individuals i analitzar patrons de desplaçament, interaccions entre animals o situacions anòmales, com immobilitat prolongada o aglomeracions excessives, aspectes directament relacionats amb el benestar animal.

En un nivell de major complexitat conceptual, s'han desenvolupat projectes que combinen Visió per Computador i Machine Learning per avaluar condicions ambientals o emocionals dels animals de granja. Alguns estudis utilitzen càmeres RGB, tèrmiques o infraroges per identificar signes d'estrès, coïxesa o canvis en la postura corporal.“[11]” Aquests sistemes no només aporten informació objectiva, sinó que permeten reduir la necessitat d'intervenció humana directa i millorar la traçabilitat de les dades.

Tot i aquests avenços, encara hi ha reptes oberts. Entre ells, la manca de conjunts de dades específics i ben anotats per a cada espècie, la dificultat d'obtenir models efectius a diferents condicions d'entorn, i la necessitat de metodologies interpretables que permetin traduir els resultats tècnics en conclusions útils per a veterinaris o gestors d'explotacions.

En aquest context, el present treball s'emmarca dins d'una línia de recerca actual que busca desenvolupar eines automatitzades i accessibles per analitzar el comportament dels pollastres de manera objectiva. L'objectiu no és només assolir un rendiment tècnic elevat, sinó contribuir a una producció alimentària més ètica i sostenible, basada en evidències quantitatives sobre el benestar animal.

6 DESENVOLUPAMENT

En aquesta secció es descriu el procés complet de creació del sistema de detecció i seguiment del comportament dels pollastres.

El desenvolupament s'ha estructurat en quatre blocs principals que corresponen a l'evolució metodològica del projecte: des de tècniques clàssiques de Visió per

Computador fins a models avançats d'IA i perspectives futures amb models generatius.

Preparació de l'entorn de treball

Per tal de garantir un entorn flexible i amb suficients recursos computacionals, s'ha treballat principalment amb Google Colab, aprofitant la seva capacitat de processament i integració amb Google Drive.

Aquesta connexió ha permès utilitzar la memòria RAM de Colab de manera eficient i accedir directament als fitxers del projecte (vídeos, imatges i models), mantenint una estructura clara i compatible.

6.1 Background Subtraction i Watershed: Tècniques de la vella escola

La primera fase del projecte va consistir en explorar tècniques clàssiques de Visió per Computador per obtenir una base inicial de detecció sense recórrer a models profunds.

Extracció del fons del vídeo

Com a primer pas per a l'anàlisi del moviment, es va intentar extreure el fons dels vídeos mitjançant tècniques estadístiques com la mitjana i la mediana de frames consecutius.

Aquest enfocament no va oferir resultats satisfactoris degut a la curta durada del vídeo i a la manca de moviment significatiu dels animals, que feien que els valors mitjans incorporessin parts del primer pla.

Com a alternativa, es va provar un enfocament basat en Intel·ligència Artificial Generativa, utilitzant Stable Diffusion per generar un fons aproximat sense els elements mòbils. Aquesta tècnica va permetre obtenir resultats qualitativament millors.

Background Subtraction

Un cop obtingut un fons de referència, es va aplicar una tècnica de Background Subtraction sobre imatges de mostra per aïllar els pollastres del fons, obtenint les seves siluetes. Els resultats s'observen a la Figura 3.

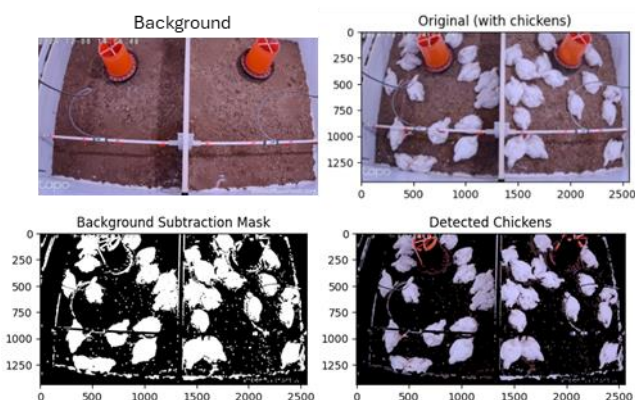


Figura 3 Detecció de pollastres amb background subtraction.

Aplicació de Watershed

A continuació, es va provar la segmentació Watershed "[16]", una tècnica clàssica útil per separar objectes que es toquen, habitual en problemes com la segmentació de cèl·lules o d'animals.

Tanmateix, els resultats no van ser del tot satisfactoris degut a les constants aglomeracions dels pollastres com es mostra a la Figura 4.

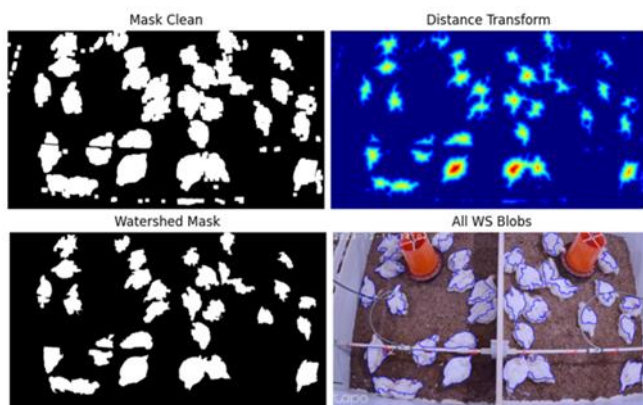


Figura 4 Procés de Watershed

Aquesta fase va servir com a prova de concepte per validar que era possible extreure les siluetes dels animals i obtenir una primera versió funcional de la detecció. Els resultats d'aquesta etapa, encara que preliminars, van permetre definir millor els requisits per a les etapes següents.

6.2 Preparació del conjunt de dades i entrenament del model YOLO

A partir de les limitacions dels enfocaments clàssics, es va iniciar una fase centrada en la preparació de dades i l'entrenament d'un model de Deep Learning per a la detecció i segmentació de pollastres.

Preparació del conjunt de dades i etiquetatge

Per entrenar el model de detecció, es van extreure frames individuals dels vídeos enregistrats, seleccionant-los de forma que estiguin distribuïts equitativament en la duració del vídeo, fent que fossin el més diferents possible.

Les imatges resultants es van etiquetar manualment mitjançant el programari web CVAT (Computer Vision Annotation Tool), utilitzant contorns o màscares de segmentació en lloc de simples caixes delimitadores, per tal d'obtenir una anotació més precisa de la forma dels animals.

Aquesta tasca es va dur a terme manualment per garantir la qualitat del conjunt d'entrenament i validació. Es van crear així dos subconjunts diferenciats: un per a l'entrenament i un altre per a la validació del model.

Entrenament del model YOLO

Un cop preparades les dades, es va procedir a entrenar un model de detecció utilitzant YOLOv8, aprofitant el model preentrenat “yolov8n-seg.pt” per accelerar la convergència.

El procés d’entrenament es va dur a terme a Colab, configurant els paràmetres principals a un fitxer de configuració.

Després de diverses iteracions d’entrenament i ajust de paràmetres, es va aconseguir un model capaç de detectar els pollastres amb un nivell de precisió satisfactori sobre imatges de validació.

6.3 Proves amb Histogram of Oriented Gradients (HoG) i l’Hungarian Algorithm

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

Tot i que alguns models proporcionen deteccions fiables, alguns no inclouen un sistema intern de ReIdentification. Això fa que, en situacions d’alta densitat o quan els animals s’oculten parcialment, dues deteccions consecutives puguin intercanviar-se la identitat o es detectin com un nou individu. Per mitigar-ho, es va experimentar amb:

- Descriptors HoG per capturar informació visual dels pollastres (contorn, textura, orientació dels gradients). “[15]”
- Algoritme Hongarès per resoldre l’assignació òptima de deteccions entre frames.
- Combinació d’ambdós criteris –distància espacial + similitud HoG– en una matriu de costos ponderada.

Exemple simplificat de l’algoritme Hongarès en tracking d’objectes

Per facilitar la comprensió del procés d’assignació, es presenta un exemple mínim amb dos objectes detectats en dos frames consecutius (frame A i frame B).

- Frame A: deteccions A_1 i A_2
- Frame B: deteccions B_1 i B_2

Per a cada parell A_i-B_j es calcula un cost que combina:

- la distància espacial entre centres
- la distància entre descriptors visuals (per exemple HoG)

Suposant una combinació ponderada d’ambdós criteris, s’obté la següent matriu de costos:

	B_1	B_2
A_1	0.2	0.8
A_2	0.6	0.3

Cada valor representa el cost d’associar una detecció del frame A amb una del frame B.

L’algoritme Hongarès resol el problema d’assignació minimitzant el cost total i obtenint la correspondència òptima:

- $A_1 \rightarrow B_1$
- $A_2 \rightarrow B_2$

Amb un cost total de: $0.2 + 0.3 = 0.5$

Tot i que A_2 també podria associar-se amb B_1 , aquesta assignació produiria un cost global superior. L’algoritme Hongarès evita aquest tipus de decisions locals i garanteix una assignació globalment òptima.

Metodologia

En primer lloc, es van seleccionar dos frames separats del vídeo (frame A i frame B) per tal d’avaluar l’associació d’objectes en un escenari controlat. Per a cada frame:

- S’executa el model YOLO entrenat per obtenir les caixes delimitadores i les màscares de cada pollastre.
- Per a cada pollastre detectat es calcula el seu vector HoG. El descriptor proporciona una representació compacta però informativa del contorn i l’estructura de cada animal.
- Es construeix la matriu de costos. Per cada parell de deteccions entre A i B es calcula la distància espacial entre centres i la distància HoG, mesurant la diferència visual entre pollastres.
- Es fa una combinació ponderada de 30% distància espacial i 70% distància HoG. Es va arribar a aquests percentatges mitjançant experimentació.
- Un cop construïda la matriu de costos, l’algoritme Hongarès calcula l’assignació òptima minimitzant el cost total. Això produeix un mapa d’associacions del tipus: $A[i] \rightarrow B[j]$
- Per facilitar la visualització del resultat, es genera una imatge comparativa on es mostren els dos frames en paral·lel, s’hi dibuixen les caixes de cada detecció amb la seva ID i es tracen línies que connecten cada pollastre del frame A amb la seva correspondència al frame B. El resultat obtingut es mostra a la Figura 5.

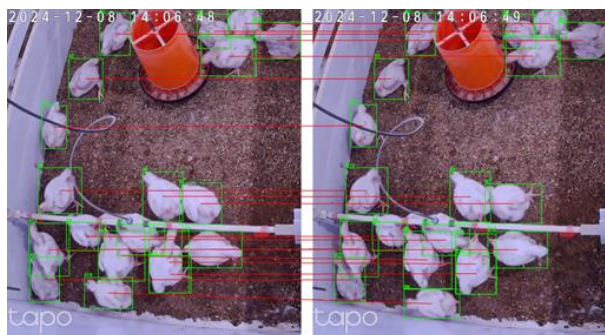


Figura 5 Correspondència de pollastres amb HoG i Hungarian Algorithm

Aquesta prova va permetre validar el potencial del sistema d'associació i va servir com a base conceptual per a una versió híbrida de tracking desenvolupada posteriorment. Tot i no ser un mètode prou robust per utilitzar-lo directament en producció, va resultar clau per comprendre la importància de combinar informació espacial i visual en el seguiment de pollastres.

6.4 Model YOLO entrenat i mètodes de tracking

Un cop disposant 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 dels pollastres.

Predicció i seguiment en vídeo

Un cop entrenat el model, es van generar prediccions sobre imatges i vídeos de prova, observant un reconeixement consistent dels pollastres en diferents escenes. Les prediccions es troben representades a la Figura 6.



Figura 6 Mostra de la detecció de pollastres amb el model entrenat. Color verd: 70% o més de certesa. Color vermell: menys de 70% de certesa.

Per tal de donar continuïtat al seguiment en moviment, es van integrar quatre mètodes que permeten fer tracking d'objectes entre frames consecutius a partir de les deteccions del model YOLO. Els mètodes utilitzats són Norfair [13], DeepSORT, BoT-SORT i ByteTrack. Aquests sistemes de seguiment aconsegueixen identificar i mantenir la trajectòria de cada pollastre dins del

vídeo, permetent futurs càlculs per al projecte.

Dels mètodes de BoT-SORT i ByteTrack només s'han realitzat proves de 10 segons degut al seu alt cost computacional, el qual utilitza més memòria RAM de la que proveeix Google Colab de forma gratuïta.

1. Norfair
 - a. Basat en distàncies euclidianes i optical flow.
 - b. Molt eficient i adequat per a prototipatge ràpid.
 - c. Pot mostrar limitacions en situacions d'oclusió prolongada.
2. DeepSORT
 - a. Afegeix ReIdentification Embeddings per reconèixer animals després d'una oclusió.
 - b. Millora les trajectòries, però depèn del bon funcionament del model ReID.
 - c. Més cost computacional.
3. BoT-SORT (Prova de 10s)
 - a. Millora DeepSORT afegint filtratge per Kalman filter avançat i estratègies anti-déficit.
 - b. Sol ser més robust quan hi ha animals que es tapen uns als altres.
 - c. Bona estabilitat de trajectòries.
4. ByteTrack (Prova de 10s)
 - a. Algorisme recent que utilitza deteccions de baixa confiança per millorar l'associació.
 - b. Funciona millor en entorns amb oclusions i deteccions inconsistents.

Millora del seguiment: Norfair + HoG + Hungarian Algorithm

Tot i els bons resultats dels mètodes anteriors, es va desenvolupar una versió millorada del seguiment basada en Norfair, afegint-hi un sistema propi de Reidentification basat en Histogram of Oriented Gradients (HoG) i reorganitzant el procés d'associació d'objectes amb l'algorisme Hongarès.

L'objectiu d'aquesta versió híbrida és integrar dos criteris:

- Proximitat espacial (posició entre frames consecutius)
- Similitud visual (comparació del descriptor HoG)

Això permet reduir canvis d'identitat i millorar la consistència temporal, sobretot en moments on dos o més pollastres es travessen o s'oculten parcialment.

En la pràctica, aquest enfocament ha estat el que ha proporcionat els millors resultats globals, equilibrant eficiència computacional i estabilitat del seguiment. Els petits errors observats són deguts principalment a falles

puntuals del model de detecció (màscare incorrectes o absents en algun frame), més que no pas a la lògica de tracking.

Millora del seguiment: Norfair + Auto-Encoder

Com a alternativa al sistema basat en descriptors manuals, es va implementar un nou mètode de seguiment que combina Norfair amb un sistema de ReIdentification basat en Deep Learning, utilitzant un Auto-Encoder convolucional entrenat específicament amb retalls d'imatges de pollastres.

Per entrenar l'Auto-Encoder, es van extreure automàticament retalls dels animals a partir de les deteccions del model YOLOv8-seg. El model aprèn una representació latent compacta (embedding) que captura les característiques visuals més rellevants de cada pollastre 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.

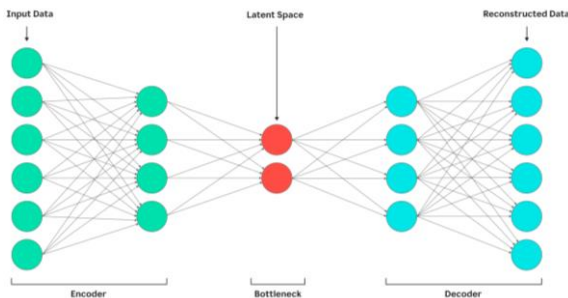


Figura 7 Diagrama d'un Auto-Encoder

En les proves realitzades, aquest mètode ha obtingut resultats molt semblants als del sistema Norfair + HoG + Hungarian Algorithm. Tot i no mostrar una millora clara respecte a aquest enfocament, valida l'ús de descriptors apresos automàticament com una alternativa viable als descriptors manuals.

7 RESULTATS

En aquesta secció es presenten els resultats obtinguts a partir del sistema de detecció i seguiment desenvolupat.

7.1 Comparació dels mètodes de tracking

S'ha realitzat una comparació quantitativa dels diferents mètodes de seguiment utilitzats en el projecte: Norfair combinat amb HoG i Hungarian Algorithm, DeepSORT, BoT-SORT, ByteTrack i Norfair combinat amb Auto-Encoder.

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

MOTA (Multiple Object Tracking Accuracy) que avalua la precisió global del sistema, tenint en compte els falsos positius (FP), els falsos negatius (FN) i els canvis d'identitat (ID Sw.), MOTP (Multiple Object Tracking Precision) que mesura la precisió espacial del seguiment, indicant com d'ajustades són les deteccions respecte a la veritat de referència i IDF1 que 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 seguits pel sistema respecte les deteccions del Ground Truth.

	MOTA	MOTP	IDF1	Switches	FP	FN
Norfair (HoG)	94.6%	0.015	96.1%	5	130	309
Norfair (AutoEncoder)	94.6%	0.015	96.1%	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

Figura 8 Taula dels resultats obtinguts

7.2 Resultats de moviment individual dels pollastres

A més de l'anàlisi global, el sistema permet obtenir mètriques detallades per a cada pollastre seguit individualment, com el moviment total, el moviment mitjà per frame i la velocitat màxima per a cada ID.

La Figura 9 presenta com a exemple el camí seguit pel pollastre amb ID número 22. Es pot observar com després de estar en repòs, s'ha desplaçat per a menjar per a després tornar a tombar-se a un lloc diferent. La forma de la línia mostra el caminar característic dels pollastres.



Figura 9 Mostra del camí seguit per un pollastre

Aquest tipus d'anàlisi individual és especialment rellevant per a futures aplicacions relacionades amb el benestar animal, ja que permet detectar patrons anòmals de moviment o canvis sobtats en el comportament d'un individu concret.

7.3 Anàlisi del moviment segons la posició a l'escena (costat esquerre vs dret)

Finalment, es va calcular la quantitat total de moviment dels pollastres en funció de la seva posició dins del recinte, dividint l'escena en dues regions: costat esquerre i costat dret.

El moviment s'ha quantificat com la suma de les distàncies recorregudes pels pollastres en cada frame. Aquesta mesura permet obtenir una aproximació robusta de l'activitat global dels animals al llarg del vídeo.

La Figura 10 mostra el gràfic comparatiu del moviment acumulat entre ambdós costats. Com es pot observar, existeix un desequilibri molt significatiu cap al costat esquerre, que concentra la major part del moviment total detectat.

Aquest resultat indica una distribució espacial clarament no homogènia del comportament dels pollastres durant el període analitzat. Tot i que en aquesta secció no s'analitzen les causes d'aquest fenomen, el sistema desenvolupat demostra la seva capacitat per detectar i quantificar patrons de comportament col·lectiu de manera automàtica.

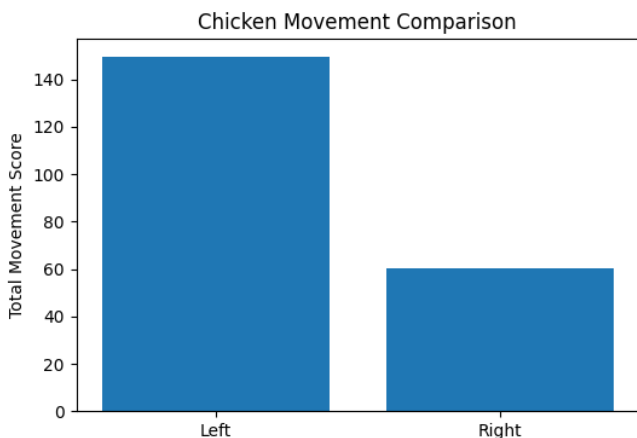


Figura 10 Comparació del moviment total dels pollastres entre el costat esquerre i el costat dret de l'escena

8 CONCLUSIONS

Els resultats obtinguts al llarg d'aquest treball demostren que és possible desenvolupar un sistema eficaç de detecció i seguiment del comportament dels pollastres mitjançant tècniques de Visió per Computador i Deep Learning, fins i tot en escenaris amb alta densitat d'animals i oclusions freqüents.

Els resultats de la Figura 8 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. Tot i aquests bons resultats, cal destacar que aquestes proves s'han realitzat sobre fragments de vídeo més curts (10 segons) a causa de les limitacions computacionals imposades per l'entorn de treball.

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.

En conclusió, 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.

9 TREBALL A FUTUR

Tot i els bons resultats obtinguts amb YOLO i Norfair, existeixen línies de millora basades en models generatius de nova generació.

9.1 Stable Diffusion

Stable Diffusion ja es va utilitzar en fases preliminars per obtenir el fons del vídeo. En el futur podria servir per:

- Generar escenaris sintètics d'explotacions avícoles.
- Simular diferents densitats d'animals.
- Millorar la robustesa del model amb Data Augmentation sintètic.

Això reduiria dràsticament la necessitat d'obtenir dades reals, sovint costoses i difícils de registrar.

A més a més, també es capaç de fer deteccions correctes dels pollastres i altres elements de les imatges com es mostra a la Figura 11.



Figura 11 Detecció i recoloració de pollastres/menjadors fet per Stable Diffusion

9.2 Segment Anything Model (SAM v3)

La tercera generació de SAM proporciona:

- Segmentació instantània i molt precisa. Fins i tot amb només el nom dels objectes que es volen detectar.
- Capacitat de refinar màscares amb un clic o exemples visuals.
- Funcionament robust en condicions d'il·luminació i ocultació adverses.

La integració d'un model com SAM v3 permetria:

- Anotar dades amb molta més rapidesa.
- Refinar deteccions automàtiques per entrenar models més precisos.
- Millorar el tracking quan les màscares són irregulars.

Una prova realitzada amb el model mostrat a la Figura 12 ha sigut capaç de detectar tots els pollastres perfectament només indicant-li que segmenti "chickens".

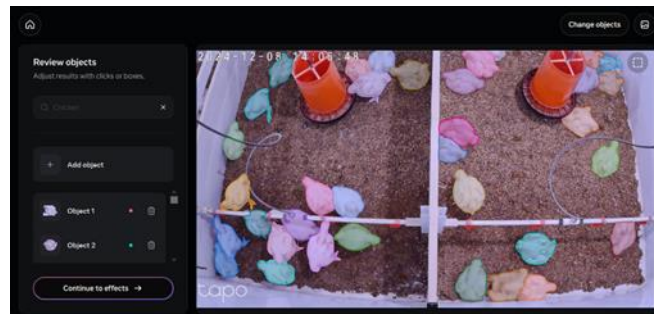


Figura 12 Detecció de pollastres utilitzant el model SAM v3

9.3 Detecció automàtica del comportament de pecking

Una de les línies de futur més rellevants d'aquest projecte és la detecció automàtica del comportament de pecking (picoteig) o bullying, ja que es tracta d'un indicador clau tant de l'activitat alimentària com de possibles problemes de benestar animal, com l'estrès o el picoteig de plomes.

La detecció de pecking requereix un nivell de precisió superior al de la simple detecció i seguiment del cos del pollastre, ja que implica identificar moviments ràpids, localitzats i repetitius del cap i, especialment, del bec. Això suposa un repte important en escenaris amb alta densitat d'animals i freqüents oclusions.

En aquest context, els models generatius i de segmentació avançada obren noves possibilitats. En proves exploratòries realitzades amb Stable Diffusion, s'ha observat que el model és capaç de diferenciar correctament parts específiques de l'anatomia del pollastre, com ara el cos i el bec, tal com es mostra a la Figura 13. Aquesta capacitat de comprensió semàntica de l'escena suggereix que aquests models podrien utilitzar-se com a suport per a la detecció fina de comportaments.



Figura 13 Mostra de com Stable Diffusion diferencia entre cos i bec

10 AGRAÏMENTS

Vull expressar el meu profund agraïment a la meua mare, que ha estat un pilar constant de suport moral al llarg de tot aquest projecte. La seva confiança, paciència i ànim en els moments més exigents han estat essencials per mantenir-me motivat i avançar amb determinació.

També desitjo agrair especialment al meu professor tutor la seva dedicació i la qualitat de la seva guia. La seva orientació clara, la seva disponibilitat i la seva capacitat per encaminar-me en els moments de dubte han estat claus per al bon desenvolupament d'aquest treball.

11 ÚS DE LA IA GENERATIVA EN EL DESENVOLUPAMENT DE L'INFORME I DEL TREBALL

Per a la realització de l'informe només s'ha fet ús de ChatGPT per a la revisió tipogràfica i gramàtica del text.

No s'ha fet cap ús de IA generativa en la creació de la metodologia del projecte ni en el disseny de l'informe.

S'ha fet ús de IA generativa per a fer la generació del contingut de les proves de l'apartat 9 per a poder comprovar les millores que podrien aportar al tracking d'objectes.

BIBLIOGRAFIA

- [1] P. Kadam, G. Fang and J. J. Zou, "Object Tracking Using Computer Vision," *Computers*, vol. 13, no. 6, Art. no. 136, pp. 1-18, May. 2024. [Online]. Available: <https://www.mdpi.com/2073-431X/13/6/136>
- [2] V. K. Sharma and R. N. Mir, "A comprehensive and systematic look up into deep learning based object detection techniques," *Engineering Applications of Artificial Intelligence*, vol. 38, Art. no. 100301, Nov. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1574013720304019>
- [3] M. Hu, S. Ali and M. Shah, "Detecting Global Motion Patterns in Complex Videos," *Computer Vision Lab, University of Central Florida, Orlando, FL, USA*. [Online]. Available: <https://www.crcv.ucf.edu/papers/Detecting%20Global%20Motion%20Patterns.pdf>
- [4] Y. Xu, X. Zhou, S. Chen and F. Li, "Deep learning for multiple object tracking," *IET Computer Vision*, May 2019. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-cvi.2018.5598>
- [5] A. Amini, "MIT Introduction to Deep Learning," YouTube, 2025. [Online Video]. Available: https://www.youtube.com/watch?v=alfdI7S6wCY&list=PLtBw6njQRU-rwp5_7C0oIVt26ZgiG9NI
- [6] Kaggle, "Intro to Deep Learning," [Online Course]. Available: <https://www.kaggle.com/learn/intro-to-deep-learning>
- [7] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri and F. Herrera, "Deep learning in video multi-object tracking," *Neurocomputing*, vol. 381, pp. 61-88, Mar. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231219315966>
- [8] D. Meimetis, I. Daramouskas, I. Perikos and I. Hatzilygeroudis, "Real-time multiple object tracking using deep learning methods," *Neural Computing and Applications*, vol. 35, pp. 89-118, 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s00521-021-06391-y>
- [9] X. Hui, D. Zhang, W. Jin, Y. Ma, G. Li, "Fine-tuning faster region-based convolution neural networks for detecting poultry feeding behaviors," *International Journal of Agricultural and Biological Engineering*, vol. 18, issue 1, pp. 64-73, Feb. 2025. [Online]. Available: <https://www.sciopen.com/article/10.25165/j.ijabe.20251801.6344>
- [10] X. Li, M. Cai, Z. Liu, C. Yin, X. Tan, J. Wen, Y. Han, "A lightweight model based on knowledge distillation for free-range chickens detection in complex commercial farming environments," *Artificial Intelligence in Agriculture*, vol. 16, issue 1, pp. 266-283, Mar. 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S258972172500090X>
- [11] D. Baltzer, S. Douglas, J. Haunert, Y. Dehbi, I. Tiemann, "Smart glasses in the chicken barn: Enhancing animal welfare through mixed reality," *Smart Agricultural Technology*, vol. 10, Mar. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772375525000206>
- [12] S. Millstein, "¿Cuántos pollos se matan cada año?", *Sentient Media*, Feb. 2025. [Online]. Available: <https://sentientmedia.org/es/cuantos-pollos-se-matan-cada-ano/>
- [13] Tryolabs, "Norfair: Lightweight Python library for object tracking", GitHub Repository, 2025. [Online]. Available: <https://github.com/tryolabs/norfair>
- [14] W. A. Triyanto, K. Adi, J. E. Suseno, "Chicken tracking for location mapping of lameness chickens using YOLOv8 and deep learning-based tracking algorithm", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, issue 1, pp. 407-418, Apr. 2024. [Online]. Available: <https://ijeecs.iaescore.com/index.php/IJECS/article/view/35560/18138>
- [15] S. Mallick, "Histogram of Oriented Gradients explained using OpenCV", *LearnOpenCV*, Dec 2016. [Online]. Available: <https://learnopencv.com/histogram-of-oriented-gradients/>
- [16] "Image Segmentation with Watershed Algorithm", *OpenCV*. [Online]. Available: https://docs.opencv.org/4.x/d3/db4/tutorial_py_watershed.html
- [17] Nwojke, "Simple Online Realtime Tracking with a Deep Association Metric", GitHub Repository, 2025. [Online]. Available: https://github.com/nwojke/deep_sort
- [18] NirAharon, "BoT-SORT: Robust Associations Multi-Pedestrian Tracking", GitHub Repository, 2025. [Online]. Available: <https://github.com/NirAharon/BoT-SORT>
- [19] FoundationVision, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box", GitHub Repository, 2025. [Online]. Available: <https://github.com/FoundationVision/ByteTrack>