



This is the **published version** of the bachelor thesis:

Brustenga Garriga, Bernat; Antens, Coen Jacobus, dir. Monitorización del comportamiento de cerdos con Visión por Computador. 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/317553>

under the terms of the  license

Monitorització del comportament de porcs amb Visió per Computador

Bernat Brustenga Garriga

1 de juliol de 2025

Resum– Aquest projecte presenta el desenvolupament d'un sistema de monitorització intel·ligent del comportament de porcs mitjançant tècniques avançades de visió per computador i intel·ligència artificial. El sistema integra detecció d'objectes, segmentació d'instàncies, estimació de postura i seguiment multiobjecte per identificar cada animal, mantenir-ne la identitat al llarg del temps i analitzar-ne el moviment i el comportament alimentari.

El treball inclou la creació de conjunts de dades, l'entrenament de models basats en YOLOv8 i la implementació d'un sistema propi de tracking amb assignació d'identitats, reforçat amb l'algorisme hongarès. A més, es vincula cada identitat amb la seva pose corporal per detectar si el porc menja i calcular la distància que recorre. El sistema desenvolupat representa una base sòlida per a futures millores i aplicacions en l'àmbit de la ramaderia intel·ligent.

Paraules clau– Smart Farming, Benestar Animal, Visió per Computador, Tracking d'Objectes, Intel·ligència Artificial.

Abstract– This project presents the development of an intelligent system for monitoring pig behavior using advanced techniques in computer vision and artificial intelligence. The system combines object detection, instance segmentation, pose estimation, and multi-object tracking to individually identify pigs, maintain their identity over time, and analyze their movement and feeding behavior.

The work includes dataset generation, training of YOLOv8-based models, and the implementation of a custom identity tracking system enhanced by the Hungarian algorithm. Each pig's identity is linked to its estimated posture, enabling the detection of feeding events and the calculation of traveled distance. Although the system shows promising results, further validation is needed to ensure robustness under real farm conditions. The developed system represents a solid foundation for future improvements and applications in the field of smart livestock farming.

Keywords– Smart Farming, Animal Welfare, Computer Vision, Object Tracking, Artificial Intelligence.



el màxim rendiment de les seves terres i els seus cultius.

1 INTRODUCCIÓ

EL sector agroalimentari està experimentant una transformació significativa gràcies a la implementació de tecnologies avançades, donant lloc al concepte de Smart Farming o agricultura intel·ligent. Aquest enfocament consisteix a administrar l'agricultura utilitzant tecnologia moderna capaç d'augmentar la qualitat i la quantitat dels productes agrícoles. Eines basades en GPS, escaneig de l'estat del sòl, anàlisi de dades i internet de les coses ajuden avui a agricultors de tot el món a obtenir

Pel que fa a la producció porcina, Catalunya ocupa una posició destacada a Espanya. Segons l'Informe Anual del Porcí de 2022 [1], Catalunya lidera la producció de carn de porc al país, representant el 41% de la producció total, amb una producció de més de dos milions de tones. Aquesta xifra és significativa, tenint en compte que Espanya és el país amb més producció porcina d'Europa.

La implementació de tecnologies com la Visió per Computador en projectes de vigilància de porcs permet detectar si els animals mengen, beuen, es mouen i estan saludables, contribuint així a millorar el benestar animal i optimitzar la producció. Aquesta monitorització contínua facilita la detecció precoç de problemes de salut, permetent una intervenció ràpida.

- E-mail de contacte: bernat.brustenga@autonoma.cat
- Menció realitzada: Computació
- Treball tutoritzat per: Coen Antens
- Curs 2024/25

2 OBJECTIUS

L'objectiu principal d'aquest projecte és la monitorització del comportament dels porcs en temps real mitjançant tècniques de visió per computador i intel·ligència artificial, amb l'objectiu final de millorar la seva salut i benestar.

Per assolir aquest objectiu general, es plantegen els següents objectius específics:

- Desenvolupar un sistema capaç d'identificar i detectar porcs de manera automàtica en vídeos mitjançant models d'aprenentatge profund.
- Implementar tècniques de segmentació d'instàncies per obtenir una delimitació precisa de cada individu dins l'escena.
- Estimar la postura corporal dels porcs per analitzar-ne l'orientació i identificar accions com menjar o beure.
- Desenvolupar un sistema de tracking robust per fer el seguiment continu i individualitzat de cada porc, mantenint la seva identitat al llarg del temps.
- Detectar anomalies en el comportament alimentari i patrons de moviment que puguin indicar problemes de salut, com ara la immobilitat o inactivitat prolongada.
- Integrar tots els models en un sistema únic capaç d'oferir informació detallada sobre el comportament dels porcs.
- Contribuir a una gestió més eficient i proactiva de les granges, millorant tant el benestar animal com la rendibilitat del sistema productiu.

Aquest conjunt d'objectius constitueix la base per al desenvolupament d'un sistema intel·ligent de monitorització que aprofiti el potencial de la visió per computador per donar suport a la producció ramadera moderna.

3 PLANIFICACIÓ

El desenvolupament del sistema s'ha estructurat en diverses fases, que comprenen des del tractament inicial de les imatges fins a la integració dels models i la validació del sistema final.

Estudi inicial del sistema (3 setmanes)

En aquesta fase inicial es realitzarà un estudi preliminar per entendre el sistema actual i les seves limitacions. S'identificaran les necessitats específiques del projecte i s'exploraran les tecnologies adequades per a la solució proposada.

Entrenament del model de detecció amb bounding boxes (2 setmanes)

Es desenvoluparà i entrenarà un model específic per a la detecció d'objectes utilitzant bounding boxes, aplicant tècniques avançades d'aprenentatge profund amb YOLO.

Entrenament del model de Instance Segmentation (3 setmanes)

A continuació, es desenvoluparà un model avançat de segmentació d'instàncies utilitzant YOLO per identificar i delimitar clarament cada porc present a les imatges, facilitant una anàlisi més precisa i detallada.

Entrenament del model de Pose Estimation (2 setmanes)

A continuació, s'integrarà un model de detecció de la posició del cos a través d'algoritmes de *Pose Estimation*, que permetran analitzar la postura i els moviments dels porcs. L'objectiu és identificar patrons de comportament relacionats amb activitats com menjar o beure. Això també permetrà veure la orientació de l'animal i veure si aquest està menjant o no.

Desenvolupament del sistema de tracking (3 setmanes)

Per al seguiment dels porcs al llarg del temps es preveu utilitzar algoritmes com **SORT** o **DeepSORT**, així com realitzar proves específiques amb l'**algorisme hongarès** aplicat de forma independent. Aquest enfocament permetrà comparar el rendiment de cada mètode per associar deteccions i gestionar trajectòries individuals.

Aquests mètodes permetran implementar un sistema de tracking eficient per a l'anàlisi del comportament dels animals.

Integració dels models per a l'anàlisi del comportament (2 setmanes)

En aquesta etapa s'integraran tots els models desenvolupats (detecció, segmentació, pose i tracking) per realitzar una anàlisi conjunta i detallada del comportament dels porcs. Això permetrà una interpretació més robusta i precisa dels resultats.

Millora conjunta dels models integrats (2 setmanes)

Finalment, es realitzaran ajustos i optimitzacions conjuntes dels models integrats per millorar-ne la precisió global i l'eficiència operativa, assegurant que el sistema sigui robust i efectiu en condicions reals.

Elaboració de l'informe i preparació de la presentació (2 setmanes)

Finalment, es redactarà l'informe tècnic, es prepararà la presentació del projecte, i es realitzarà una revisió exhaustiva del contingut per garantir-ne la coherència i qualitat. També es practicarà la defensa oral per assegurar una comunicació efectiva davant del tribunal.

Aquest enfocament estructurat i escalonat garanteix que cada fase es desenvolupi, provi i optimitzi de manera efectiva abans de la seva integració final. La distribució temporal d'aquestes fases es pot visualitzar al diagrama de Gantt que es mostra a la Figura 13.

4 METODOLOGIA

Per a la gestió del treball, he escollit la metodologia *Scrum*, ja que em permet definir dates de finalització de les tasques amb antelació. Aquest aspecte és clau per poder establir una planificació clara i tenir una previsió detallada de quan s'acabarà cada fase del projecte.

Tot i que Scrum està dissenyat principalment per a equips, en aquest cas l'aplicaré en un entorn individual adaptant-ne alguns elements. La divisió del treball en Sprints i la definició de tasques concretes amb terminis establerts em permetran tenir un millor control del progrés i garantir que el projecte avanci de manera estructurada.

Una de les raons per les quals he optat per aquesta metodologia és que ja l'he utilitzat en altres projectes i sempre m'ha funcionat bé. A més, proporciona una gran flexibilitat a l'hora de fer ajustos si sorgeixen imprevistos.

Per a la gestió i seguiment de les tasques, utilitzaré *Jira*, una eina que facilita l'organització dels Sprints i la visualització de l'estat de cada tasca. Amb *Jira* podré establir històries d'usuari, definir prioritats i fer un seguiment detallat del progrés del projecte.

5 ESTAT DE L'ART

En aquest apartat s'exposen els principals models, algorismes i eines actuals en visió per computador aplicables a la monitorització del comportament animal. Tot i que no tots han estat utilitzats en el desenvolupament, s'han analitzat i valorat per determinar-ne la viabilitat dins del projecte. Les tecnologies seleccionades s'han detallat en els apartats següents.

5.1 YOLO (You Only Look Once)

YOLO [5] és un model de detecció d'objectes en temps real que es destaca per la seva rapidesa i eficiència. La seva arquitectura permet identificar múltiples objectes en una sola passada per la imatge, oferint una gran velocitat de processament. En el context d'aquest projecte, s'utilitzaran diverses variants de YOLO:

- El model de detecció d'objectes basat en *bounding boxes*, que permet localitzar i classificar els porcs en la imatge.
- El model de segmentació d'instàncies, que proporciona una segmentació més precisa per a cada porc, identificant els límits exactes dins l'entorn.
- YOLO també s'utilitzarà per a tasques de *Pose Estimation*, aprofitant les seves últimes variants com YOLO-Pose [6], que permeten estimar les posicions clau del cos dels animals en moviment.

La seva capacitat de processar imatges de forma ràpida i eficient fa de YOLO una opció idònia per a la monitorització de vídeos del comportament dels animals en una granja.

5.2 Mask R-CNN

Mask R-CNN [7] és una extensió de Faster R-CNN que afegeix una capa de segmentació enmascarada als objectes detectats. A més de classificar i localitzar els objectes en la imatge, Mask R-CNN genera una màscara pixel-per-pixel per a cada objecte detectat, permetent una segmentació precisa. Aquest model és útil per a tasques on es necessita una segmentació detallada i precisa de cada objecte, com la identificació de les parts del cos dels animals. Tot i que és més costós computacionalment que YOLO, és una opció valuosa quan la segmentació detallada és necessària per obtenir una millor anàlisi del comportament, especialment en entorns on la precisió és fonamental.

5.3 DensePose

DensePose [8] és un model avançat de segmentació humana que permet mapar els píxels d'una imatge 2D a una superfície 3D del cos humà. Tot i que està dissenyat per a humans, la seva arquitectura podria adaptar-se a altres aplicacions, com l'anàlisi de la postura i moviment d'animals. No obstant això, la seva adaptació per als porcs podria resultar difícil, ja que requeriria entrenar un model específic per a aquesta espècie. A més, la complexitat d'adaptar la representació 3D del cos humà als porcs faria que aquest model no fos el més adequat per al projecte. En canvi, models com YOLO amb segmentació d'instàncies podrien proporcionar una solució més efectiva i eficient per a la detecció i seguiment dels animals.

5.4 Models de seguiment d'objectes

Per tal de fer el seguiment dels porcs al llarg del temps en els vídeos, s'utilitzaran models de **tracking-by-detection**. Els més destacats són:

- **Filtre de Kalman:** Algorisme de predicció-recuperació que permet estimar l'estat (posició i velocitat) d'un objecte mitjançant un model lineal amb soroll gaussià. En seguiment d'objectes, el filtre de Kalman s'utilitza per predir la posició futura d'un objecte, fins i tot quan no ha estat detectat en un fotograma concret.
- **Algorisme Hongarès:** Mètode d'assignació òptima que resol el problema de correspondència entre deteccions noves i trajectòries existents, minimitzant una matriu de costos basada, habitualment, en la distància entre bounding boxes o centroides. És una peça clau en sistemes com SORT o DeepSORT.
- **SORT (Simple Online and Realtime Tracking)**[3], un algorisme lleuger que utilitza un filtre de Kalman i associació de deteccions mitjançant l'algorisme d'assignació Hongarès.
- **DeepSORT** [9], una extensió de SORT que incorpora característiques visuals obtingudes mitjançant xarxes neuronals profundes per millorar l'associació d'identitats entre fotogrames.
- **ByteTrack** [13]: Algorisme de tracking que millora sistemes anteriors tenint en compte també deteccions

amb baixa confiança. Aquesta estratègia permet mantenir trajectòries més estables i contínues, fins i tot quan la detecció és parcial o incerta.

Aquestes tècniques de seguiment permeten identificar i analitzar el comportament individual de cada porc, fins i tot quan es mouen o es solapen parcialment dins de l'escena.

5.5 Segment Anything Model (SAM2)

El *Segment Anything Model* (SAM2) [4] és un model de segmentació desenvolupat per Meta AI que permet generar màscares d'objectes de forma automàtica a partir d'entrades simples com punts, caixes o regions de la imatge. SAM2 es basa en una arquitectura de tipus *Vision Transformer* (ViT), que processa la imatge en forma de seqüència i aplica mecanismes d'atenció espacial per identificar les àrees rellevants.

Una de les característiques més destacades del model és la seva capacitat de generalització, ja que ha estat entrenat amb onze milions d'imatges i mil milions de màscares, la qual cosa li permet oferir resultats sòlids en escenaris diversos, fins i tot en dominis no específicament coberts durant l'entrenament.

Aquest model és especialment útil en aquest projecte per generar màscares de segmentació d'instàncies a partir de bounding boxes prèviament anotades. La seva flexibilitat i precisió permet millorar considerablement la qualitat de les anotacions i reduir el temps requerit per obtenir un dataset complet i precís.

5.6 Eines d'anotació

Per a la generació dels conjunts de dades utilitzats en aquest projecte, s'han considerat diverses plataformes especialitzades en anotació per a visió per computador. Les més destacades són:

- **Roboflow** [10]: Eina basada en el núvol amb una interfície intuïtiva i fàcil d'utilitzar, ideal per a projectes individuals o amb posada en marxa ràpida. Permet augmentació automàtica de dades, control de versions i exportació en formats com YOLO o COCO. Ha estat l'eina principalment utilitzada en aquest projecte.
- **Supervisely** [11]: Plataforma avançada enfocada a entorns col·laboratius, amb eines potents per a l'anotació manual i assistida. Requereix una corba d'aprenentatge més elevada i una configuració més complexa.
- **CVAT** [12]: Eina de codi obert desenvolupada per Intel i OpenCV, orientada a usuaris tècnics que busquen flexibilitat i control total. Tot i ser molt completa, la seva interfície és menys accessible i l'instal·lació pot ser més exigent.

Totes aquestes plataformes permeten fer anotacions per a detecció d'objectes, segmentació d'instàncies i estimació de pose, i integren models automàtics com el *Segment Anything Model* (SAM), que facilita la generació de màscares precises a partir de bounding boxes o punts.

6 DESENVOLUPAMENT

En aquest apartat es descriu detalladament el procés seguit per construir i entrenar els models utilitzats en el sistema de monitorització del comportament porcí. Es presenten les diferents fases de desenvolupament, des de la creació dels datasets fins a l'obtenció de resultats, incloent els experiments realitzats, les decisions preses i les limitacions trobades.

6.1 Object Detection amb Bounding Boxes

En primer lloc, es va treballar la detecció bàsica dels porcs mitjançant caixes delimitadores (bounding boxes), una tècnica fonamental per localitzar els animals dins dels fotogrames dels vídeos. Aquesta secció descriu com es va crear el dataset, com es van entrenar els models i quins resultats es van obtenir.

6.1.1 Creació del dataset

La primera fase va consistir en la creació d'un conjunt de dades d'anotacions mitjançant la plataforma Roboflow [10], on es van marcar manualment les *bounding boxes* corresponents als porcs en les diferents imatges. Per generar aquestes imatges, es van extreure fotogrames (*frames*) dels diferents vídeos disponibles, agafant-ne un cada segon. Aquesta estratègia es va escollir amb l'objectiu d'assegurar una certa variabilitat entre imatges consecutives i evitar que els porcs apareguessin exactament en la mateixa posició. Així, es donava temps als animals per moure's i s'obtenia un conjunt de dades més ric i divers.

En total, es van labelejar manualment 336 imatges. A més, gràcies a les opcions de descàrrega del dataset disponibles a Roboflow [10], es van aplicar tècniques d'augmentació de dades com ara desenfocament (*blur*), escala de grisos (*greyscale*), rotació (*rotation*) i retall (*crop*). Aquestes augmentacions van permetre generar un conjunt final de més de 700 imatges. Aquestes es van dividir en un 60% per entrenament (*train*), un 20% per test i un 20% per validació. Totes les imatges es van redimensionar a una resolució de 640x640 píxels per facilitar un entrenament més ràpid i eficient del model.

6.1.2 Entrenament del model

Un cop completades les anotacions, es va descarregar el dataset en format YOLO, un estàndard àmpliament utilitzat en visió per computador. Aquest conjunt de dades es va utilitzar per entrenar diversos models de detecció d'objectes basats en YOLOv8. L'entrenament es va dur a terme inicialment amb els paràmetres per defecte, amb l'objectiu de tenir una primera aproximació dels resultats.

S'han entrenat tres variants del model YOLOv8:

- **YOLOv8n** (nano): model lleuger i ràpid, ideal per dispositius amb pocs recursos.
- **YOLOv8m** (medium): equilibri entre precisió i eficiència computacional.
- **YOLOv8x** (extra-large): model gran amb més paràmetres i, potencialment, major precisió, però també més exigent computacionalment.

YOLOv8 ha estat escollit per la seva eficiència i bon rendiment en temps real, sent una de les arquitectures més modernes i optimitzades en l'àmbit de la detecció d'objectes. Un dels seus principals avantatges és la documentació clara i completa, així com el suport d'una comunitat molt activa, que facilita enormement la integració, el desplegament i la resolució d'incidències durant el desenvolupament.

6.2 Object Detection amb Instance Segmentation

Amb l'objectiu de millorar la precisió respecte a la localització i contorn dels porcs, es va incorporar la segmentació d'instàncies. Aquesta tècnica permet obtenir màscares específiques per a cada individu, essent útil per a una anàlisi més detallada del comportament. A continuació es detallen les estratègies seguides per generar els datasets, entrenar els models i analitzar-ne els resultats.

6.2.1 Creació del dataset

En veure aquestes limitacions, vaig decidir explorar altres opcions i aprofitar el conjunt de dades ja anotat amb *bounding boxes* per generar segmentacions més precises amb el model SAM2 (*Segment Anything Model*) desenvolupat per Meta AI [4]. Aquest model utilitza una arquitectura basada en transformadors i és capaç de generar màscares d'objectes a partir d'entrades com punts, caixes o màscares prèvies. La màscara final s'obté mitjançant l'atenció espacial sobre les característiques visuals de la imatge, responnent de manera eficient a les instruccions d'entrada.

Entrada mitjançant punts centrals

En una primera prova, vaig optar per passar només un punt central dins de cada bounding box com a entrada al model SAM, però els resultats no van ser els esperats. Les màscares generades sovint no cobrien bé tot el porc o incloïen parts del fons. Això em va fer adonar que, tot i que el punt central indica una ubicació aproximada, no proporciona prou informació perquè el model pugui saber amb claredat quina és la mida i forma de l'objecte que ha de segmentar. A la Figura 1 es mostra la posició d'aquests punts centrals (marcats amb una estrella) dins de cada bounding box, i es pot observar que en alguns casos no estan ben posicionats dins del cos del porc, fet que ajuda a explicar la poca precisió de les màscares generades.

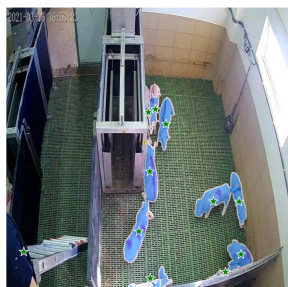


Fig. 1: Resultats amb punts centrals de la bounding box

Entrada mitjançant bounding boxes

A partir d'aquí, vaig decidir fer una segona prova passant directament la bounding box com a entrada. Aquesta estratègia va resultar molt més efectiva. Al proporcionar una caixa que delimita tota la regió on hi ha l'objecte, el model té una idea molt més clara d'on es troba i fins on s'estén. Això li permet generar una màscara molt més precisa, ajustada als contorns reals de l'animal, fins i tot en imatges complicades on els porcs apareixen girats, parcialment tapats o en postures poc habituals. A la Figura 2 es pot veure com els resultats obtinguts amb aquesta estratègia són molt més precisos, aconseguint capturar la silueta del porc amb gran fidelitat en la majoria de casos.



Fig. 2: Resultats passant la bounding box sencera

Dataset autogenerat amb SAM2

A partir d'aquest procés, es va generar un primer dataset de segmentació utilitzant les bounding boxes com a entrada al model SAM. Per tal de tenir el conjunt de dades en un format estàndard per a entrenament, es va implementar un bucle que recorria totes les imatges del dataset original amb bounding boxes i transformava les anotacions en màscares segmentades, guardant finalment el conjunt de dades en format COCO.

Dataset manual amb Roboflow

A més, es va crear un segon dataset independent mitjançant la plataforma Roboflow [10]. Aquest dataset constava inicialment de 150 imatges etiquetades manualment amb màscares de segmentació. Posteriorment, es van aplicar tècniques d'augmentació de dades com la modificació de la saturació, la brillantor i l'exposició. Després d'aquest procés, es va arribar a tenir un total de 359 imatges per a entrenament, 15 per a validació i 15 per a test.

6.2.2 Entrenament del model

Un cop generats els dos datasets, es va procedir a entrenar un model de detecció amb segmentació d'instàncies utilitzant YOLOv8x. Per tal de poder entrenar amb YOLO, el primer dataset generat amb SAM es va convertir del format COCO al format YOLO. En canvi, el segon dataset, creat manualment a Roboflow [10], ja es va descarregar directament en el format adequat per a YOLOv8.

Conversió de formats: COCO a YOLO

El format COCO (Common Objects in Context) utilitza un únic fitxer `.json` que conté totes les anotacions per a totes les imatges del dataset. Aquest fitxer inclou informació

detallada com les categories, les coordenades de les màscares (en forma de polígons), les caixes delimitadores i la relació amb les imatges corresponents. Això permet gestionar grans conjunts de dades de manera centralitzada.

En canvi, el format utilitzat per **YOLO** és diferent. Cada imatge del dataset té associat un fitxer `.txt` amb el mateix nom, on es troba la informació de les anotacions de segmentació (màscara). Cada línia d'aquest fitxer representa una màscara i conté l'identificador de la classe i les coordenades normalitzades dels punts que defineixen el contorn de la màscara. A més, el conjunt de dades inclou un fitxer `data.yaml`, on s'especifica:

- Les rutes a les carpetes d'entrenament, validació i test.
- El nombre de classes.
- Els noms de les classes.

Per tant, per poder entrenar el model YOLOv8 amb el primer dataset generat amb SAM, es va haver de fer una conversió completa de totes les anotacions del format COCO al format YOLO, generant els fitxers `.txt` corresponents per a cada imatge i creant el fitxer `data.yaml` amb la informació general del dataset.

6.3 Pose Estimation

Per tal de comprendre millor el comportament dels porcs, és fonamental saber on es troba el cap i on la part posterior del seu cos en les imatges capturades. Aquesta distinció ens permetrà determinar si el porc està menjant o bevent segons la seva orientació, evitant així confondre aquestes accions amb altres postures o moviments que no indiquen ingestió d'aliments o líquids. Per aconseguir-ho, es va optar per implementar un model de *Pose Estimation* que identifiqui de forma precisa les parts clau del cos del porc, especialment el cap i la cua.

6.3.1 Creació del dataset

Per abordar la problemàtica d'identificar la postura i orientació dels porcs, es va decidir implementar un model de *Pose Estimation*. Inicialment, es va crear manualment un dataset específic a la plataforma Roboflow [10]. Per definir el que s'ha de detectar, es va construir una estructura esquelètica simplificada que reflectís les parts clau del porc visibles en les imatges. Donat que les imatges, majoritàriament, mostraven la part superior dels porcs i rarament es veien les potes, es va optar per definir només 5 *keypoints* representatius: la cua, les dues orelles, el nas i el coll, aquest últim com a punt de connexió central per altres parts.

Amb aquesta estructura, es van etiquetar manualment 150 imatges, extretes de tres vídeos diferents mitjançant la captura d'un fotograma cada 12 segons per garantir diversitat temporal i evitar redundàncies. Posteriorment, es va aplicar augmentació de dades, incloent modificacions d'escala de grisos, brillantor, exposició i desenfocament (*blur*), assolint un total final de 339 imatges per a entrenament, 22 per a validació i 15 per a test.

6.3.2 Entrenament del model

Per a l'entrenament del model d'estimació de posició, es va utilitzar la variant **YOLOv8n-pose**, que combina l'eficiència de YOLOv8 amb una arquitectura específica per a detectar i localitzar *keypoints* en les imatges. Aquest model lleuger és adequat per entorns amb recursos limitats i permet obtenir resultats ràpids i efectius.

6.4 Tracking i Assignació d'Identitats

Un cop entrenats els models de detecció, es va abordar el problema del *tracking*, amb l'objectiu de mantenir la identitat de cada porc al llarg del temps.

Postprocessat de les *Bounding Boxes*

Abans de passar al sistema de tracking, es va implementar un **postprocessat** sobre les deteccions amb l'objectiu de filtrar i estabilitzar les bounding boxes obtingudes. Aquest postprocessat incloïa dues etapes:

1. **Filtrat per regió espacial**: només es consideraven les bounding boxes situades a la part dreta de la imatge, concretament a partir del **68% de l'amplada**, ja que la part esquerra no formava part del recinte on es trobaven els porcs.
2. **Selecció òptima de bounding boxes**: per evitar deteccions duplicades o molt properes, es va aplicar l'algorisme **Non-Maximum Suppression amb distància mínima** modificat. Aquest algorisme selecciona les **10 deteccions amb més confiança** i descarta aquelles que estiguin a una distància inferior a un llindar definit, assegurant així una cobertura uniforme i sense solapaments entre les deteccions seleccionades.

Aquest postprocessat no només millora l'estabilitat del *tracking*, sinó que també permet visualitzar clarament l'efecte del filtratge aplicat. A la Figura 14, es pot observar de manera comparativa el resultat d'aquest procés en tres etapes diferenciades: **a l'esquerra**, totes les deteccions inicials generades pel model YOLO abans de qualsevol filtratge; **al centre**, les deteccions que s'han mantingut després d'aplicar el postprocessat (filtrat per zona i *Non-Maximum Suppression*); i **a la dreta**, les deteccions que han estat descartades durant aquest procés per no complir els criteris definits. Aquesta visualització ajuda a entendre com el sistema elimina deteccions espúries o duplicades, retenint únicament les més fiables per tal de garantir un seguiment més robust i coherent al llarg del temps.

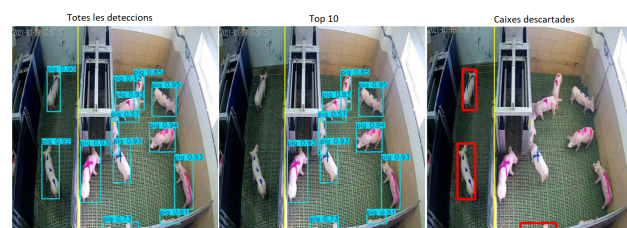


Fig. 3: Comparativa visual del postprocessat de les *bounding boxes*: a l'esquerra, deteccions inicials de YOLO; al centre, deteccions conservades; a la dreta, deteccions descartades pel filtratge.

Aquest filtratge prèviament al tracking va resultar clau per millorar l'estabilitat i fiabilitat de les identitats assignades, especialment en escenes on hi havia múltiples porcs en moviment.

Prova amb DeepSORT

Inicialment es va provar d'utilitzar el sistema **DeepSORT**, aprofitant les *bounding boxes* generades pel model YOLOv8 entrenat prèviament. Aquest enfocament, tot i ser conegut per combinar un filtre de Kalman i un sistema d'aparença visual per a cada objecte, no va oferir bons resultats en aquest cas: les identitats (ID) dels porcs canviaven constantment, especialment quan es creuaven entre ells o desapareixien breument de l'escena. Aquest fet dificultava l'anàlisi de comportament i invalidava part de la informació temporal recollida.

Implementació pròpia amb l'Algoritme Hongarès

Per millorar els resultats, es va implementar un sistema de **tracking propi** basat en l'algoritme hongarès (*Hungarian Algorithm*)[[14]], que permet fer l'assignació òptima entre les deteccions actuals i les identitats anteriors mitjançant una matriu de costos basada en la distància euclidiana entre centroides.

Aquest sistema de tracking inclou:

- Control de distància màxima per a associar deteccions.
- Un mecanisme per a considerar pistes perdudes i intentar la seva reassociació si reapareixen pròximament.
- Limitació del nombre màxim d'identitats actives (en aquest cas, 10).

Tot i que aquest enfocament va millorar substancialment la persistència de les identitats, encara es detectaven problemes en les zones de menjar, on els porcs deixaven de ser detectats temporalment. Aquests errors estaven causats pel fet que les *bounding boxes* generades per YOLO sovint fallaven quan el porc estava en postures menys habituals o parcialment ocult.

Millora amb el model de Instance Segmentation

Per tal de solucionar-ho, es va substituir el model de detecció per *bounding boxes* pel model de **segmentació d'instàncies**, també entrenat amb YOLOv8. Aquest model genera màscares precises per a cada porc, les quals posteriorment es convertien en *bounding boxes* mitjançant una funció pròpia, permetent així la integració amb el sistema de tracking.

Amb aquest canvi, es van obtenir resultats molt més estables: els porcs mantenien la seva identitat de forma contínua i la detecció en situacions complexes era més robusta.

6.5 Integració: Anàlisi de Comportament i Vinculació de Pose al Tracking

Un cop obtingut un sistema estable de *tracking* mitjançant segmentació d'instàncies, es va procedir a integrar-lo amb

el model d'**estimació de pose** per tal d'obtenir una anàlisi més completa del comportament dels porcs.

L'objectiu principal d'aquesta integració era identificar si un porc estava **menjant**, i alhora calcular la **distància recorreguda** per cada individu al llarg del vídeo. Per assolir-ho, es van seguir els passos següents:

1. Per cada fotograma, s'aplica el model de segmentació per obtenir les màscares dels porcs, les quals es transformen en *bounding boxes* i s'utilitzen per fer el *tracking* com descrit a la secció anterior.
2. Paral·lelament, s'aplica el model de *Pose Estimation* per detectar els punts clau (keypoints) del cos dels porcs, especialment el **nas**, ja que és l'indicador principal per saber si el porc està dins la zona de menjar.
3. Un cop obtinguts els *keypoints*, es calcula el **centroide del cap** de cada porc i es filtra per eliminar poses duplicades o errònies mitjançant una distància mínima entre ells.
4. Es calcula una matriu de distàncies entre els **centroids dels tracks** (*bounding boxes*) i els **centroids del cap** detectats, i s'aplica l'algoritme hongarès per fer la millor assignació entre cada identitat i la seva pose.
5. Si la posició del cap del porc (assignada a un ID concret) es troba dins d'una caixa virtual predefinida que representa la zona de menjar, es considera que el porc està menjant, i es comptabilitza el temps que hi passa.
6. Addicionalment, es guarda el desplaçament de cada porc en píxels a partir de la seqüència de centroides dels *bounding boxes*, per tal de calcular la **distància total recorreguda**.

Aquesta integració va permetre passar d'un sistema de detecció per separat a un sistema **cohesionat**, on cada ID manté la seva identitat, s'hi vincula una estimació de postura, i es pot fer un seguiment detallat tant de la seva activitat com del seu comportament alimentari.

Aquest sistema integrat és capaç de proporcionar:

- La identitat de cada porc al llarg del vídeo.
- El temps acumulat que cada porc passa menjant.
- La distància total que cada porc recorre.
- Una visualització enriquida amb màscares, *bounding boxes*, ID i trajectòria.

La integració de la *Pose Estimation* dins del sistema de tracking permet un control molt més precís del comportament i facilita la identificació d'anomalies alimentàries, com per exemple porcs que no mengen o que tenen una activitat reduïda durant períodes prolongats.

7 RESULTATS I EXPERIMENTS

En aquesta secció es presenten els resultats obtinguts a partir de l'entrenament i la validació dels diferents models utilitzats en aquest estudi. S'hi analitzen tant els resultats parcials (detecció amb bounding boxes, segmentació per instàncies i estimació de posicions) com els resultats del sistema integrat de tracking i anàlisi de comportament. L'objectiu és avaluar el rendiment global de cada enfocament, identificar possibles limitacions i establir línies de millora per a futures iteracions del projecte.

7.1 Resultats de detecció amb Bounding Boxes

Tot i la potència del model YOLOv8, els resultats obtinguts van presentar algunes limitacions depenent de la variant utilitzada:

- Els models **YOLOv8n** i **YOLOv8m** oferien bones deteccions generals, però sovint mostraven comportaments erràtics, com ara:
 - Deteccions falses (bounding boxes on no hi havia porcs).
 - Duplicació de deteccions sobre el mateix porc.
 - Invenció de bounding boxes amb confiança baixa.

Per mitigar aquests errors, es va aplicar un filtratge que seleccionava únicament les 10 *bounding boxes* amb major confiança, tenint en compte que en cap vídeo de mostra hi havia més de 10 porcs visibles alhora.

- El model **YOLOv8x**, en canvi, tendia a ser molt més estricte amb les seves prediccions: les deteccions que feia eren molt precises i amb alta confiança, però deixava de detectar alguns porcs visibles, probablement per voler assegurar-se massa abans de considerar una detecció com vàlida.

En general, es pot concloure que els models **n** i **m** ofereixen un millor balanç entre detecció i cobertura, tot i requerir un post-processat per eliminar deteccions espúries. El model **x**, si bé pot ser útil en contextos on es prioritza la precisió sobre la cobertura, no resulta òptim en aquest cas on és essencial detectar tots els porcs presents a la imatge.

Tot i aquestes limitacions, en moltes imatges el rendiment va ser notable, com es mostra a la Figura 4. Amb futurs ajustos i refinaments, es preveu que els resultats puguin millorar encara més.

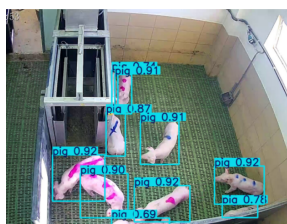


Fig. 4: Resultats model YOLOv8

7.2 Resultats de detecció amb Instance Segmentation

Model entrenat amb el dataset autogenerat amb SAM2

El primer model entrenat amb el dataset generat automàticament mitjançant SAM solia obtenir resultats inconsistents. Sovint, el model agrupava més d'un porc dins de la mateixa màscara o bé generava diverses màscares per un sol porc, cosa que dificultava una segmentació precisa per instàncies. A la Figura 5 es poden observar aquests errors, on les màscares no corresponen adequadament als animals individuals.



Fig. 5: Exemple de resultats erronis del primer model amb dataset generat amb SAM: màscara que agrupa més d'un porc

Model entrenat amb el dataset manual de Roboflow

En canvi, el segon model, entrenat amb el dataset creat manualment a Roboflow [10], acostuma a oferir resultats més precisos i coherents. Tot i això, de tant en tant es detecten errors en la segmentació. A la Figura 6 es mostren exemples de segmentacions encertades, mentre que a la Figura 20 es poden veure algunes de les ocasions en què el model falla.

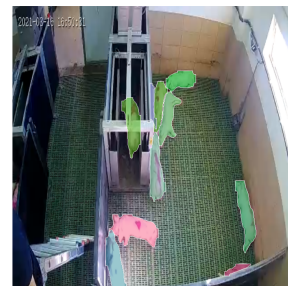


Fig. 6: Resultats correctes del segon model entrenat amb dataset manual de Roboflow [10]

7.2.1 Prova amb Detectron2

Per intentar millorar els resultats, es va provar d'entrenar un model amb **Detectron2**. Tot i això, després de l'entrenament, el model no retornava cap màscara durant la inferència. Aquest problema es treballarà en el futur per determinar si es pot resoldre i si Detectron2 pot oferir millors resultats que YOLOv8.

7.3 Resultats del model de Pose Estimation

Els resultats van superar les expectatives inicials, amb un model capaç de localitzar amb precisió els *keypoints*

escollits en la majoria de les imatges de test. Aquesta capacitat és molt útil per determinar cap a on està mirant el porc, ja que la posició relativa dels punts clau com el nas, les orelles i el coll permet inferir l'orientació de l'animal.

Tot i això, es van observar alguns errors puntuals, especialment duplicats en la detecció de poses quan un mateix porc era detectat més d'una vegada, fet que pot complicar l'anàlisi posterior. Per evitar problemes futurs derivats d'aquest fenomen, es preveu integrar aquest model d'estimació de posició amb els models de segmentació d'instàncies anteriors, de manera que dins d'una mateixa màscara d'instància només aparegui una única pose, assegurant una correspondència unívoca entre segmentació i esquelet.

Com a línia de futur, s'està planificant entrenar també amb versions més grans de YOLOv8-pose, com les variants **YOLOv8m-pose** i **YOLOv8x-pose**, per comprovar si un model més complex pot aportar encara millors resultats, sobretot en situacions on la postura és més difícil d'identificar o quan la qualitat de la imatge no és òptima.

A la Figura 7 es mostra un exemple representatiu dels resultats obtinguts pel model de Pose Estimation, on es visualitzen clarament els cinc *keypoints* sobreposats a la imatge original del porc.



Fig. 7: Resultats del model de Pose Estimation

7.4 Anàlisi conjunta de les mètriques dels models

A continuació, es presenta una explicació conjunta de les principals mètriques avaluades durant l'entrenament dels diferents models. Aquestes mètriques són fonamentals per comprendre el rendiment dels models desenvolupats:

- **Loss (pèrdua):** Indica l'error comès pel model durant l'entrenament. Una gràfica ideal de loss mostra una reducció ràpida al principi i una estabilització posterior en un valor baix.
- **Recall:** Mesura la capacitat del model per identificar correctament tots els objectes rellevants (porcs). La gràfica ideal hauria d'augmentar ràpidament i estabilitzar-se en valors propers a 1.
- **mAP50 (Mean Average Precision a IoU=0.5):** Avalu la precisió mitjana de les deteccions considerant una Intersecció sobre Unió (IoU) mínima de 0,5. La gràfica hauria de mostrar un creixement sostingut fins a estabilitzar-se en valors elevats.

- **mAP50-95 (Mean Average Precision entre IoU=0.5 i IoU=0.95):** Calcula la precisió mitjana considerant diverses llindars d'IoU (des de 0.5 fins a 0.95). Aquesta mètrica, sent més estricta, normalment presenta valors inferiors als del mAP50.

Resultats específics per model

Bounding boxes

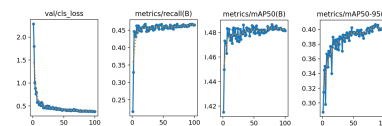


Fig. 8: Gràfiques d'entrenament del model amb bounding boxes

Class	Images	Instances	Box(P	R	mAP50	mAP50-95)
all	61	675	0.98	0.475	0.487	0.42
person	1	2	1	0	0	0
pig	61	673	0.96	0.949	0.975	0.839

Fig. 9: Taula de resultats de l'entrenament del model amb bounding boxes.

La Figura 8 mostra una estabilització ràpida de la pèrdua, amb valors finals moderats. El recall i el mAP50 s'estabilitzen al voltant de 0,50, però cal remarcar que aquest valor global està afectat per l'única imatge amb una persona, que es detecta incorrectament. Això fa baixar la mitjana general, tot i que les deteccions sobre els porcs, que són la prioritat del sistema, presenten uns resultats molt més satisfactoris. Per tant, el comportament global no és tan negatiu com podria semblar inicialment, sinó que reflecteix la penalització d'una única detecció errònia de persona.

Instance Segmentation

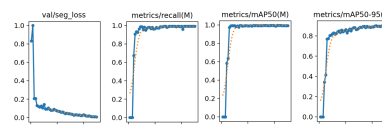


Fig. 10: Gràfiques del entrenament de Instance Segmentation

Aquest model mostra gràfiques amb valors elevats en recall i mAP50, propers a 1, indicant deteccions robustes i precises. El mAP50-95 també presenta valors molt alts (aproximadament 0.90), demostrant que les segmentacions són efectives fins i tot en condicions més exigents.

Pose Estimation

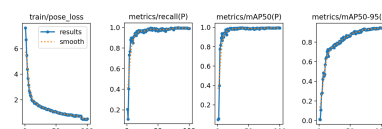


Fig. 11: Gràfiques del entrenament de Pose Estimation

La Figura 11 mostra un excel·lent comportament del model, amb valors molt baixos de pèrdua i resultats sòlids en recall i mAP50, propers a 1. El mAP50-95 també és molt satisfactori, amb valors al voltant de 0.90, indicant que el model gestiona bé fins i tot les prediccions més precises.

En general, les gràfiques confirmen la qualitat satisfactòria dels entrenaments realitzats, amb un rendiment especialment destacat en els models de segmentació i pose, tot i que el model basat en bounding boxes té marge de millora en precisió i robustesa.

7.5 Resultats del Tracking integrat amb Pose Estimation

Els resultats obtinguts pel sistema complet integrant tracking, segmentació d'instàncies i estimació de postura han demostrat una notable robustesa i precisió. La Figura 25 mostra clarament com cada porc manté consistentment la seva identitat (ID) al llarg del vídeo, així com la correcta detecció de la postura corporal i el càlcul precís del temps que cada individu passa dins del menjador.

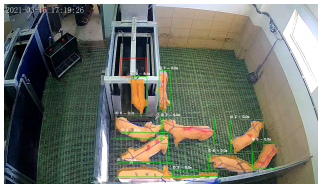


Fig. 12: Resultats del tracking dels porcs integrant detecció, segmentació d'instàncies, estimació de postura, i càlcul del temps (en segons) dins del menjador.

D'altra banda, la Figura 26 presenta un exemple representatiu del recorregut complet fet per un dels porcs durant el període d'observació. Aquest tipus de visualització permet identificar patrons de moviment específics, àrees de preferència i comportaments potencialment rellevants per a una anàlisi més profunda del benestar animal.

Aquestes visualitzacions integrades confirmen la capacitat del sistema per a oferir informació detallada i fiable sobre el comportament individual dels porcs, facilitant així una millor gestió i presa de decisions en entorns ramaders.

8 CONCLUSIONS

En general, els models de detecció i segmentació entrenats han permès obtenir resultats prometedors en la identificació, localització i seguiment de porcs. Tot i algunes limitacions detectades durant les proves amb models específics, les estratègies implementades, com l'ús de segmentació d'instàncies i l'assignació mitjançant l'algoritme hongarès, han millorat notablement l'estabilitat i precisió del sistema.

La integració amb estimació de pose ha proporcionat una eina potent per identificar comportaments específics, com menjar, i ha obert noves possibilitats d'anàlisi més detallades sobre la conducta animal. Futurs ajustos, així com proves amb altres models i conjunts de dades més

amplis i variats, podrien millorar encara més el rendiment i ampliar les capacitats d'aquest sistema.

REFERÈNCIES

- [1] Departament d'Agricultura, Ramaderia, Pesca i Alimentació de la Generalitat de Catalunya, "Observatori del Porcí Anual 2022," 2022.
- [2] F. Farahnakian, F. Farahnakian, S. Björkman, V. Bloch, M. Pastell, and J. Heikkonen, "Pose estimation of sow and piglets during free farrowing using deep learning," *Computers and Electronics in Agriculture*, vol. 216, 2024.
- [3] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Uroic, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 3464–3468.
- [4] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," *arXiv preprint arXiv:2304.02643*, 2023.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *arXiv preprint arXiv:1506.02640*, 2016.
- [6] D. Maji, S. Nagori, M. Mathew, and D. Poddar, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss," *arXiv preprint arXiv:2204.06806*, 2022.
- [7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *arXiv preprint arXiv:1703.06870*, 2017.
- [8] R. A. Güler, N. Neverova, and I. Kokkinos, "DensePose: Dense Human Pose Estimation In The Wild," *arXiv preprint arXiv:1802.00434*, 2018.
- [9] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," *arXiv preprint arXiv:1703.07402*, 2017.
- [10] Roboflow, "Roboflow," [Online]. Available: <https://roboflow.com/>. [Accessed: 30-Jun-2025].
- [11] Supervisely, "Supervisely," [Online]. Available: <https://supervisely.com/>. [Accessed: 30-Jun-2025].
- [12] CVAT, "CVAT," [Online]. Available: <https://www.cvat.ai/>. [Accessed: 30-Jun-2025].
- [13] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," *arXiv preprint arXiv:2110.06864*, 2022.
- [14] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1–2, pp. 83–97, 1955.

APÈNDIX: FIGURES DEL PROJECTE

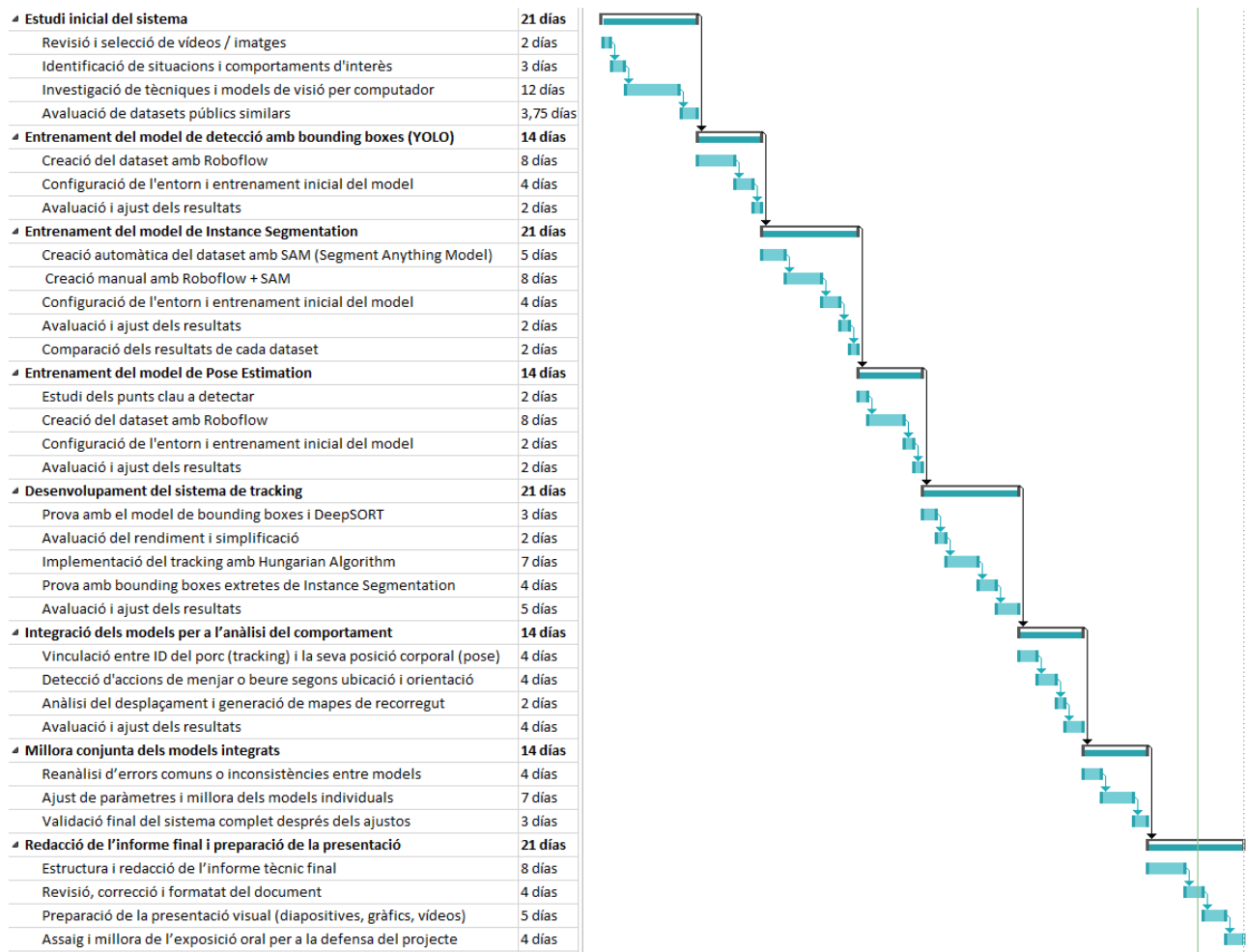


Fig. 13: Diagrama de Gantt del projecte

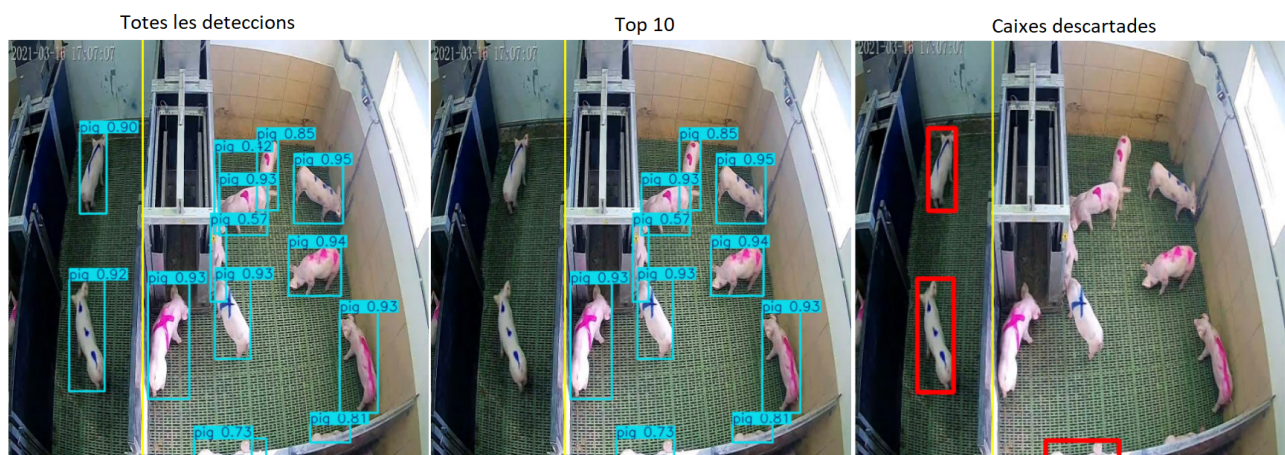


Fig. 14: Comparativa visual del postprocessat de les *bounding boxes*: a l'esquerra, deteccions inicials de YOLO; al centre, deteccions conservades; a la dreta, deteccions descartades pel filtratge.

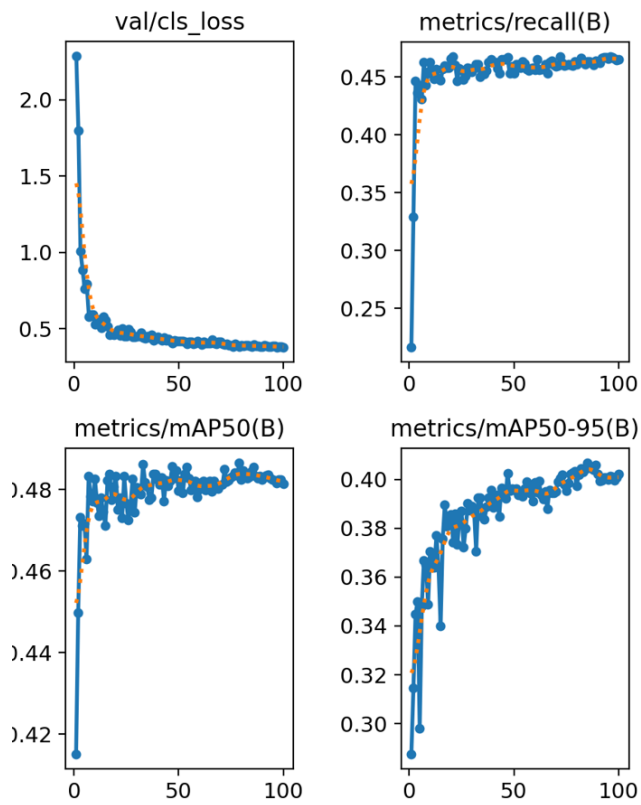


Fig. 15: Gràfiques d'entrenament del model amb bounding boxes

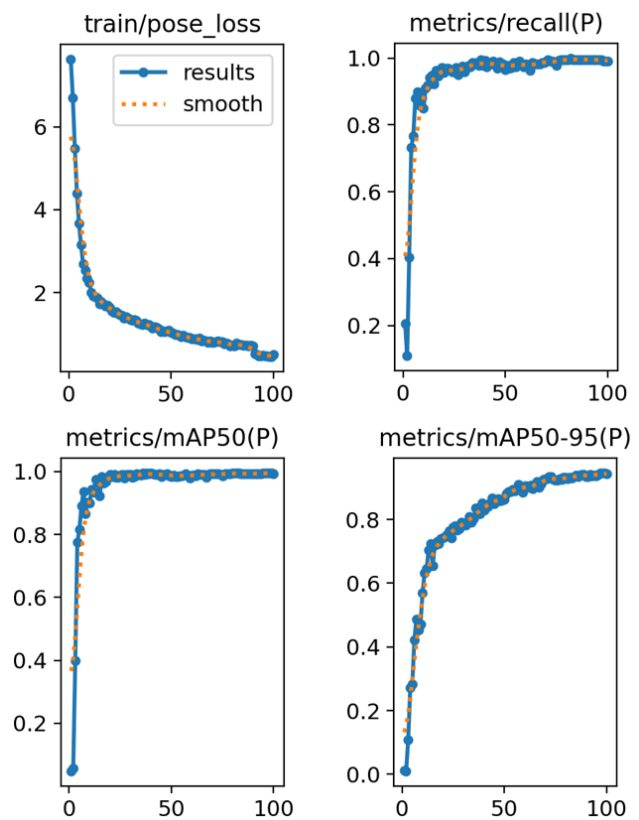


Fig. 17: Gràfiques del entrenament de Pose Estimation

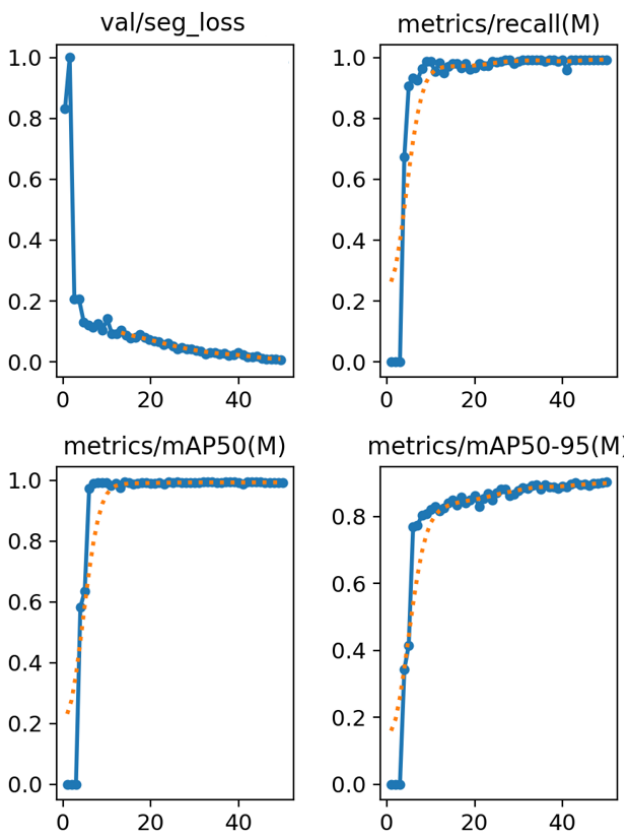


Fig. 16: Gràfiques del entrenament de Instance Segmentation

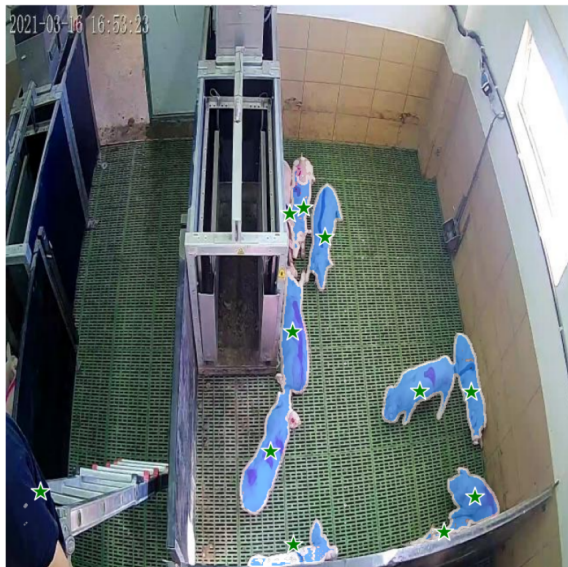


Fig. 18: Resultats amb punts centrals de la bounding box

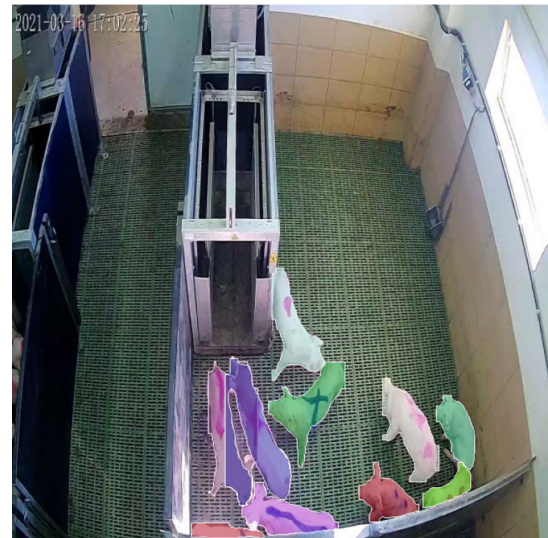


Fig. 21: Error: màscara que agrupa més d'un porc

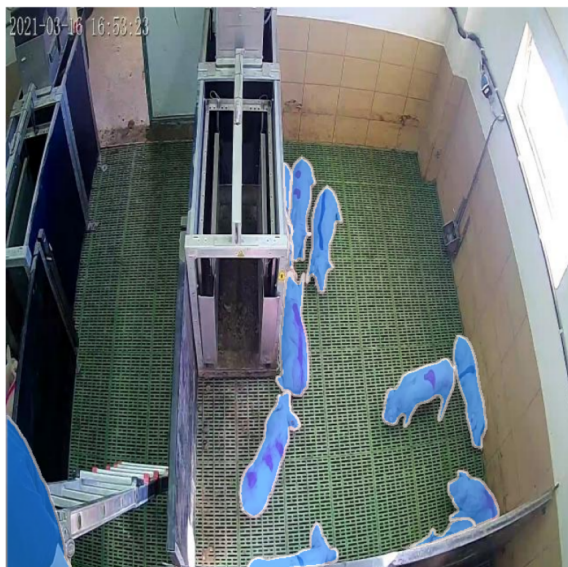


Fig. 19: Resultats passant la bounding box sencera

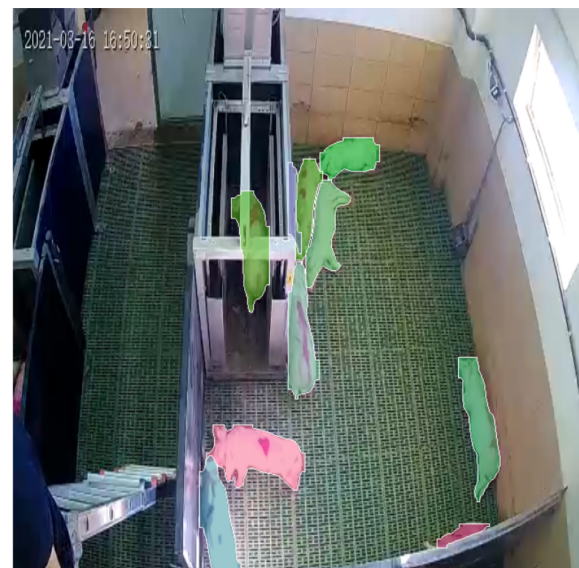


Fig. 22: Segmentacions correctes del segon model

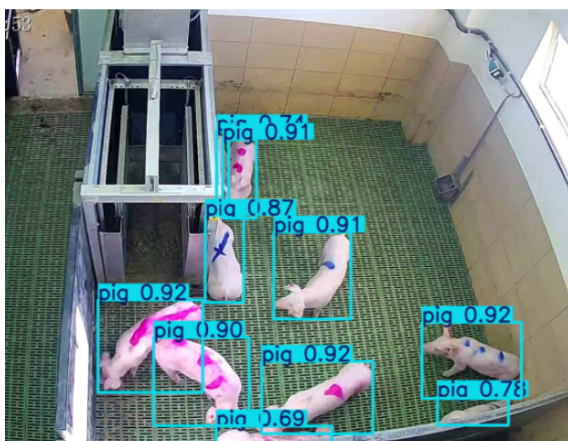


Fig. 20: Resultats model YOLOv8

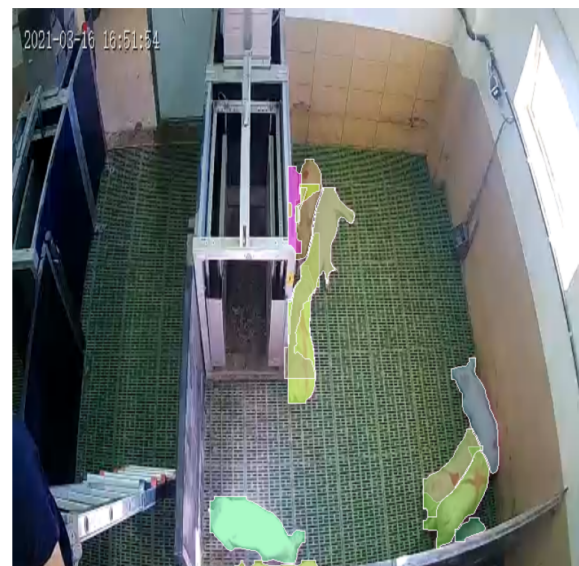


Fig. 23: Errors puntuals en la segmentació



Fig. 24: Resultats del model de Pose Estimation

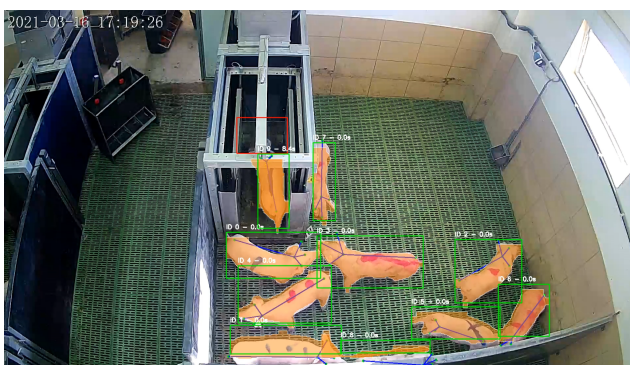


Fig. 25: Resultats del tracking dels porcs integrant detecció, segmentació d'instàncies, estimació de postura, i càlcul del temps (en segons) dins del menjador.

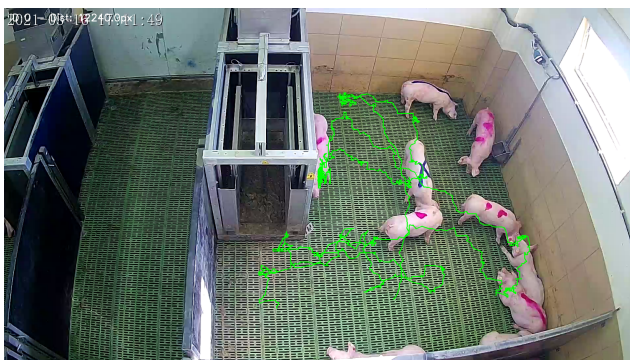


Fig. 26: Trajectòria completa d'un dels porcs dins del corral durant el període monitoritzat.