

Projecte fi de grau

Estudi: Grau en Enginyeria Informàtica

Títol: Desenvolupament d'una xarxa neuronal profunda per la ramaderia de precisió

Document: Resum

Alumne: Virginia Ramón Ferrer

Tutor: Rafael Garcia Campos

Departament: Arquitectura i Tecnologia de Computadors

Àrea: ATC

Convocatòria (mes/any): Febrer 2022

# 1 Introducció

L'arribada de les tecnologies de la informació i la comunicació a la ramaderia permet l'adquisició de dades directament a les granges, i el seu processament per a convertir-les en informació útil per a la presa de decisions sobre el maneig a realitzar. En aquest context, les granges actuals tendeixen a incrementar el nombre d'animals per treballador, al temps que el benestar animal és cada dia més rellevant. Els animals, en funció de la seva edat, estat fisiològic, i altres factors, dediquen una sèrie d'hores més o menys determinades a descansar, menjar, beure, interactuar... És el que s'anomena el seu "pressupost diari del temps". Quan un animal o un grup, es desvia significativament del seu "pressupost horari" pot ser indicatiu d'un problema.

En aquest projecte ens centrarem en la detecció d'animals i persones per a posteriorment veure com afecta la presència de persones al comportament de cada porc, és a dir, l'efecte dels factors externs al comportament dels animals. Es partirà d'una arquitectura CNN (Convolutional Neural Network) pre-entrenada per la detecció d'objectes, la YOLOv4, i es modificaran les darreres capes d'aquesta xarxa neuronal per detectar porcs i persones. Inicialment només es disposava d'una base de dades que contenia exclusivament porcs i una altra que contenia només persones. Posteriorment s'han obtingut imatges on hi ha la presència tant de porcs com a persones per a poder fer proves amb ambdues.

També s'ha desenvolupat un algorisme d'anàlisi d'imatge que interpreta les deteccions de la CNN, de forma que es mesurarà el moviment de cada animal al llarg del temps. L'algorisme utilitzat per a mesurar el moviment serà l'Unscented Kalman Filter.

Tot i que aquest PFG se centra en la detecció i seguiment del moviment dels porcs per qüestions de temps, també es vol entrenar la xarxa per a la detecció de persones per a deixar-la llesta per a utilitzar aquestes mesures i la detecció de presència de persones per a analitzar com afecta aquesta presència als animals. La implementació s'ha realitzat en Python i s'ha fet servir les llibreries Tensorflow i Keras per a la implementació de la xarxa neuronal.

# 2 Metodologia

Aquest projecte va ser una proposta del tutor Rafael Garcia. Es va analitzar el que suposava dissenyar i implementar aquest projecte en la seva totalitat i es va arribar a la conclusió de que en el temps que es tenia per a desenvolupar el Projecte Final de Grau no es podia fer en la seva totalitat, per tant es va arribar al consens de només implementar la part de detecció de porcs i persones i l'algorisme de seguiment d'aquests.

El següent pas va ser fer l'anàlisi de les dades que es tenien i/o es necessitaven per al correcte desenvolupament d'aquest. Inicialment es tenia un dataset de porcs aconseguit per l'Universitat d'Edimburg. En el cas d'aquest projecte, s'analitza com afecta la presència de persones al comportament dels porcs, mesurat en moviments, per tant, és necessari també tenir un da-

---

taset amb persones per a poder entrenar la detecció de persones i porcs. Per a fer això, s'ha utilitzat part del dataset de Google Open Images v4, només fent ús d'imatges amb persones etiquetades. Al principi només s'anava a entrenar la detecció de porcs i persones per separat ja que no es disposava d'un dataset que contingui's ambdós objectes a les seves imatges, però ja entrat en el desenvolupament d'aquest, es va aconseguir un dataset proporcionat per l'IRTA el qual conté seqüències d'imatges de persones interactuant en el mateix espai que els porcs, per tant va permetre poder fer entrenaments i proves amb ambdós junts. El problema amb aquest dataset és que no són imatges etiquetades, per tant l'alumne va haver d'emprar temps en etiquetar i processar algunes d'aquestes imatges per a poder fer-ne ús. A la figura 1 es poden veure algunes mostres d'aquests datasets.

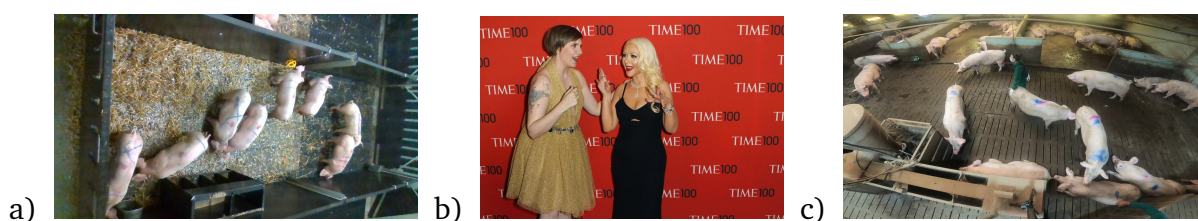


Figura 1: Mostra dels datasets: a) Universitat d'Edimburg, b) Open Images v4, c) IRTA

Per a fer la detecció de persones i porcs es va haver d'eleger amb quin mètode fer-ho. Al analitzar les diferents opcions que es tenien, es va arribar a la conclusió de fer ús d'una Xarxa Neuronal Convolutiva o CNN. Després de fer proves amb algunes xarxes i comparar diferents factors com a resultats i velocitat d'entrenament i detecció, es va arribar al consens de fer ús de la YOLOv4, específicament una implementació en Python creada per l'usuari de GitHub "pythonlessons" anomenada "TensorFlow-2.x-YOLOv3", ja que té un temps molt reduït de detecció. Quan es va tenir la xarxa elegida, es va començar a fer entrenaments amb les dades i, a partir dels resultats obtinguts després de cada entrenament, es va anar refinant el dataset.

Per a fer el seguiment dels animals i les persones, des d'un principi el tutor va expressar la idea de fer ús d'un Filtre de Kalman. Ja que el moviment que es té a les seqüències d'imatges no és lineal, es va eleger fer ús del Unscented Kalman Filter o UKF, que és una variant del filtre per a models no lineals, específicament una modificació de la implementació en Python creada per l'usuari de GitHub "sj23patel" anomenada "Object-Tracking".

Quan ja es van tenir la Xarxa Neuronal i l'algorisme de seguiment elegits, es varen combinar, essent la sortida de la xarxa neuronal l'entrada de l'algorisme de seguiment. A mesura que s'anava entrenant la YOLOv4, s'anaven fent proves amb el UKF per veure com funcionaven en conjunt i anar refinant les dades i els paràmetres en conseqüència.

Pel que fa a la memòria, al llarg del projecte s'ha anat fent un seguiment del procés i, quan ja es tenia implementat tot el sistema i només quedava refinar dades i paràmetres, es va començar a redactar aquest document en paral·lel a aquest refinament per a aprofitar el temps que tardaven els entrenaments i les proves.

### 3 Resultats

Com ja s'ha introduït en aquest document, es tenen dos datasets que contenen porcs, el de la Universitat d'Edimburg i el de l'IRTA, el primer dels quals es té una quantitat decent de dades etiquetades i amb un punt de vista de la càmera bastant beneficiós a l'hora de diferenciar els animals, mentre que el segon es caracteritza per ser dades menys controlades i marcades per la falta d'etiquetes. Aquesta diferència és visible al llarg dels resultats obtinguts.

La xarxa neuronal final ha set entrenada amb un Learning Rate de  $5e-6$ , durant 50 epochs, amb un batch de 4. A la taula 1 es poden observar els resultats obtinguts a les diferents proves. Com es pot observar, es fa ús de tres score thresholds diferents i de quatre datasets: *Universitat d'Edimburg*: (només hi ha les imatges de validació i test del dataset de la Universitat d'Edimburg), *IRTA*: (Dataset on només hi ha les imatges de validació i test del dataset de l'IRTA, *Dataset de validació* i *Dataset de test*). Si es comparen els resultats entre els dos primers datasets en la classe porc, es veu que hi ha una diferència molt alta d'AP per a tots els score thresholds utilitzats, arribant a tenir una diferència de més del 30%. En ambdós cassos l'AP es redueix notablement quan més gran es fa l'score threshold, passant del 92.55% al 26.32% al dataset de la Universitat d'Edimburg i del 59.56% al 14.21% al dataset del IRTA només amb una diferència de 0.2 de valor Si es miren els resultats de les persones de l'IRTA, es veu que són molt millors que els dels porcs, variant menys d'un 10% entre l'score threshold del 0.5 al 0.7, tinguent 96.56% i 86.77% respectivament. Això es dona perquè les persones en aquest dataset duen sempre una indumentària similar, per tant ha pogut generalitzar-les millor que els porcs, els quals tenen marques pintades i tendeixen a estar molt junts. Si es miren els datasets de validació i de test, es veu que els resultats són similars als del porc de la Universitat d'Edimburg, ja que una gran part de les mostres que es tenien venia d'aquest dataset, però l'AP es veu penalitzat per la presència d'imatges del IRTA pel que fa als porcs. A la figura 2 es poden veure exemples de les deteccions.

Si es fa un anàlisi dels problemes més comuns en la detecció són: la persona o un altre porc tapa a l'altre porc el suficient per a que la xarxa no ho detecti, la persona passa pel davant d'un porc i la xarxa detecta un mateix porc com a dos regions d'interès separades, i, com que els porcs són animals que tendeixen a estar molt junts, es dona repetidament la situació en que la xarxa confon dos porcs junts com a un mateix animal.

		Dataset									
		Universitat d'Edimburg	IRTA			Val			Test		
AP (en %)		Porc	Porc	Pers.	mAP	Porc	Pers.	mAP	Porc	Pers.	mAP
Score thres.	0.5	92.55	59.56	96.56	78.06	87.31	96.87	92.09	87.57	96.25	91.91
	0.6	68.76	37.74	93.07	65.40	61.76	93.33	77.54	66.37	92.80	79.58
	0.7	26.32	14.21	86.77	50.49	21.20	86.73	53.96	28.17	86.82	57.49

Taula 1: Resultats proves mAP als datasets en percentatges



Figura 2: Exemples de deteccions: a) dataset d'Universitat d'Edimburg, b) IRTA, c) Comparació de deteccions on blau són deteccions correctes, vermell porcs al ground truth que no han set detectats i verd porcs que han set detectats però no apareixen al ground truth.

Pel que fa al seguiment, si es torna a parlar dels dos datasets, els resultats obtinguts amb el dataset de la Universitat d'Edimburg són remarcables, aconseguint fer el seguiment de la majoria dels porcs durant seqüències de vídeo completes. A la figura 3 es poden observar dos trajectòries de porcs diferents al llarg de la seqüència de vídeo, marcats amb una línia groga i una línia rosa.

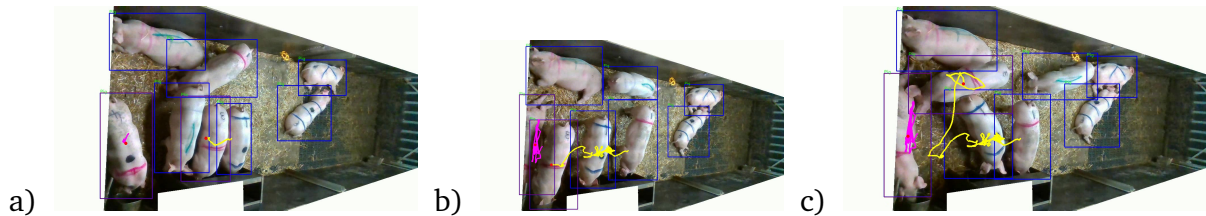


Figura 3: Mostra del seguiment de porcs.

L'error més donat en les proves d'aquest datasets s'ha donat quan un porc està molt aferrat a un altre, a sobre inclòs, impedit la detecció, seguit del intercanvi del seguiment de dos porcs que es creuen. A la figura 4 es pot veure com el porc verd, amb la trajectòria rosa, passa per davant del porc vermell, amb seguiment violeta, i a la imatge c) s'intercanvien els tracks, observant que a la imatge d) el porc verd ha passat a tenir el seguiment lila i el porc vermell el seguiment rosa.



Figura 4: Mostra de l'intercanvi de tracks entre dos porcs.

## 4 Conclusions

S'ha vist que a partir de les xarxes neuronals i d'algorismes de seguiment, en aquest cas el Filtre de Kalman en la seva variant Unscented Kalman Filter, és possible dur a terme la detecció i el seguiment de persones i animals com són els porcs i extreure'n informació en

temps reduït.

Analitzant els resultats, la primera conclusió que es treu és la importància que té un bon dataset per a obtenir bons resultats: tenir una quantitat notable de dades, les dades han d'estar ben etiquetades, ja que és la base de l'entrenament i, en el cas d'aquest projecte, que el punt de vista de les imatges utilitzades sigui l'òptim per a poder extreure'n la màxima quantitat d'informació. Un punt que hi ha que remarcar és la diferència notòria de treballar en situacions ideals en comparació a situacions reals. Com s'ha vist, el dataset de la Universitat d'Edimburg es podria entendre aquí com la situació quasi ideal, on el punt de vista és bo, el nombre d'objectes és reduït i les dades ja estan etiquetades. Moltes vegades s'assumeix que aquestes situacions ideals o quasi ideals es donaran sempre i, al veure el funcionament del mateix sistema en un entorn real, com són les imatges de l'IRTA, es veu com un sistema que es creia que tenia un funcionament correcte ha de ser modificat per a que pugui funcionar en entorns reals. Seguidament, es remarca la importància dels paràmetres usats en el desenvolupament de qualsevol algorisme, ja que, com s'ha pogut veure a la prova del Learning Rate, si algun valor és molt gran o molt petit, pot fer que no s'arribi mai al resultat desitjat.

Tot i que els resultats de la detecció estan lluny de ser perfectes, cosa que beneficiaria al sistema de seguiment, aquesta detecció és suficientment bona com per produir resultats de "tracking" acceptables. Des del principi d'aquest projecte es sabia que, per temps i recursos, arribats a aquest punt de la feina uns resultats com els que es tenen amb un mAP al voltant del 60% per a score thresholds de 0.6 o 0.7 eren els esperats. Com a primer avenç en un projecte d'aquestes magnituds és un bon resultat i obre la porta a poder seguir avançant.

Pel que fa a la viabilitat de continuar amb el projecte, tal i com s'indicava a la introducció, al veure els resultats obtinguts es veu que, tot i que fa falta molt de treball i temps refinament tant dels algorismes com de les dades els quals comportarien la despesa d'una gran quantitat de temps, és una tasca que es podria arribar a dur a terme e implantar en un espai real, tot i que en aquest cas s'haurien d'analitzar factors externs com poden ser la instal·lació del hardware necessari per a dur-ho a terme en l'espai, la il·luminació... Tot i això, en la opinió del alumne, el benefici que aquest sistema podria aportar al benestar dels animals i al benefici econòmic dels ramaders és una bona motivació per a seguir fent feina en aquesta branca de treball.

Pel que fa al treball futur, analitzant els resultats obtinguts i veient on es vol arribar, farà falta dur a terme les següents tasques: augmentar les dades etiquetades per a poder entrenar millor la xarxa neuronal, fer registres dels moviments quantificats per l'algorisme de seguiment i buscar una forma òptima d'emmagatzemar-les i que siguin de fàcil post-processament, afegir una xarxa neuronal intermitja entre la YOLOv4 i l'UKF que agafi les deteccions de persones i detecti les parts de cos per a tenir dades més exactes de la relació entre la posició de la persona i el moviment dels animals i, quan s'hagin dut a terme les tasques anteriors, desenvolupar un algorisme que agafi les dades extretes i faci l'anàlisi de la variació del moviment dels animals quan hi ha persones a prop.