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

Trullo Salas, Gerard; Bernal del Nozal, Jorge, tut. Reconocimiento de cartas por visión por computador del videojuego "Pokémon TCG Pocket". 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/317560>

under the terms of the  license

Reconeixement de cartes per visió per computador del videojoc “Pokémon TCG Pocket”

Gerard Trullo Salas

Resum— Aquest projecte aborda el repte del reconeixement automatitzat de cartes del videojoc Pokémon TCG Pocket a partir de captures de pantalla de les col·leccions dels usuaris, incloent el nombre de còpies de cada carta. El projecte es divideix en tres fases metodològiques principals. En primer lloc, la implementació de la segmentació d'imatges. Aquí s'exploren dues metodologies, una basada en la detecció de vores (Canny) i l'anàlisi de contorns, i l'altra que aprofita l'estructura columnar de l'aplicació i operacions morfològiques horitzontals. En segon lloc, s'exploren dues metodologies per a la classificació de les cartes. El primer és un enfocament basat en Xarxes Neuronals Convolucionals (CNN) amb aprenentatge per transferència i augmentació de dades. El segon és una metodologia de Reconeixement per Comparació de Característiques Locals (Feature Matching). Per últim, s'utilitza un mòdul OCR juntament amb un preprocesament previ de la imatge per detectar el nombre de còpies de cada carta. En conjunt, el projecte constitueix el primer pas cap al desenvolupament d'una futura aplicació per a la gestió d'intercanvis de cartes.

Paraules clau—Visió per Computador, Segmentació d'Imatges, Classificació d'Imatges, Feature Matching, ORB, Aprenentatge per Transferència, Xarxes Neuronals Convolucionals (CNN), OCR, Pokémon TCG Pocket, Python

Abstract— This project addresses the challenge of automated card recognition in the Pokémon TCG Pocket video game from user collection screenshots, including the number of copies for each card. The project is divided into three main methodological phases. Firstly, the implementation of image segmentation. Here, two methodologies are explored: one based on edge detection (Canny) and contour analysis, and the other which leverages the application's columnar structure and horizontal morphological operations. Secondly, two methodologies for card classification are investigated. The first is an approach based on Convolutional Neural Networks (CNN) with transfer learning and data augmentation. The second is a Feature Matching recognition methodology. Lastly, an OCR module, along with prior image preprocessing, is used to detect the number of copies for each card. Collectively, this project constitutes the first step towards the development of a future mobile application for managing card trades.

Index Terms— Computer Vision, Image Segmentation, Image Classification, Feature Matching, ORB, Transfer Learning, Convolutional Neural Networks (CNN), OCR, Pokémon TCG Pocket, Python



1 INTRODUCCIÓ - CONTEXT DEL TREBALL

Pokémon TCG Pocket és un joc de cartes col·leccionables per a mòbil que ha aconseguit una gran popularitat gràcies a la seva dinàmica de joc i a la comunitat activa que el recolza. Tot i que la plataforma oficial ofereix una experiència de joc sòlida, un dels seus punts febles és el sistema d'intercanvi de cartes. Aquest sistema és molt limitat i dificulta notablement la realització d'intercanvis entre jugadors, fent que la interacció en aquest aspecte sigui pràcticament inviable.

El dia 14 de març de 2025, la conta oficial de Pokémon TCG Pocket va anunciar [1] que, a finals d'any, s'implementaran diversos canvis en el sistema d'intercanvi, incloent la possibilitat de compartir amb altres jugadors quines cartes

es busquen per intercanviar. Aquest anunci posa en manifest que el sistema actual presenta problemes significatius i que, tot i els esforços dels desenvolupadors, potser no s'adapta completament a les necessitats de la comunitat.

Amb aquest context, el present treball té com a objectiu principal implementar un sistema de reconeixement de cartes mitjançant tècniques de visió per computador, que permetrà identificar automàticament les cartes i el nombre d'exemplars a partir de captures de pantalla de la col·lecció dels jugadors. Aquest serà el primer pas per desenvolupar una aplicació mòbil que ofereixi als usuaris una eina complementària i centralitzada per a la gestió d'intercanvis. L'eina permetrà als usuaris registrar-se, pujar captures de pantalla de la seva col·lecció i gestionar els intercanvis de manera directa.

La motivació darrere d'aquest projecte neix de la meua experiència personal com a jugador i del desig de millorar un aspecte clau que afecta la interacció dins del joc.

-
- E-mail de contacte: gtstrullo@gmail.com
 - Menció realitzada: Enginyeria de Computació
 - Treball tutoritzat per: Jorge Bernal del Nozal (Ciències de la Computació)
 - Curs 2024/25

2 OBJECTIUS

2.1 Objectius primaris

Reconeixement i identificació de cartes a partir de captures de pantalla

Desenvolupar un mòdul (usant visió per computador) capaç d'extraure i identificar automàticament les cartes mostrades en una captura de pantalla de la col·lecció del jugador.

Detecció del nombre de còpies per carta

Extreure la informació numèrica (present a la part inferior esquerra de cada carta) que indica el nombre de còpies de cada carta dins la mateixa captura.

2.2 Objectius secundaris

Reconeixement de cartes en captures de sobres oberts

Desenvolupar la funcionalitat per identificar les cinc cartes d'un sobre obert, de manera que s'actualitzi la col·lecció sense haver de capturar tota la biblioteca cada cop.

3 ESTAT DEL ART

El reconeixement automàtic de cartes en imatges (físiques o digitals) combina tècniques de segmentació i classificació. En diversos projectes existents s'ha usat un enfocament per fases: primer es segmenten les cartes de la imatge, després es classifiquen i finalment es poden aplicar mètodes addicionals (com OCR per llegir text) segons convingui. Tot i que molts estudis s'han centrat en cartes físiques (p. ex. cartes de pòquer o *Magic*), les tècniques principals s'apliquen igualment a jocs digitals com Pokémon. A continuació es ressenyen algunes de les tècniques més rellevants utilitzades en projectes recents.

3.1 Segmentació de cartes

La segmentació consisteix a localitzar i delimitar cada carta dins la captura d'entrada. En projectes pràctics s'ha optat per combinacions de preprocessament, detecció de vores i operacions morfològiques.

Contorns i umbralització: Molts algorismes detecten cartes aplicant un filtre de thresholding (binarització) o Canny, i després cerquen contorns tancats. Per exemple, un detector de cartes *Magic* primer umbralitza la imatge per millorar el contrast i després usa *findContours* d'OpenCV per trobar possibles cartes [2]. També es recomana convertir a escala de grisos, aplicar desenfocament Gaussià i threshold d'Otsu, després cridar *findContours* i filtrar per àrea per extreure el contorn de la carta. En ambdós casos l'element principal és localitzar el contorn més gran que correspon al rectangle de la carta, per després fer-li un *perspective transform* i obtenir la carta en vista frontal si fos necessari.

Operacions morfològiques lineals: Si la interfície té estructures fixades (per exemple, columnes o files de cartes), es poden usar dilatacions/erosions amb kernels lineals per extreure línies [3]. En pantalles on les cartes s'alineen en files col·locades, aplicar morfologia

horitzontal pot servir per separar files de cartes i segmentar-les més fàcilment.

Detecció d'objectes (NN): Alternativament, s'han usat xarxes de detecció d'objectes per localitzar directament cada carta. Una d'elles, *tiny-YOLO*, s'ha fet servir per trobar la posició de cartes de *Magic*, i posteriorment fer-ne l'identificació amb *perceptual hashing* [4]. En general, detectors basats en CNN com YOLOv3/v8 poden aprendre a reconèixer la forma de carta en diverses condicions i així substituir processos de contorns més manuals.

3.2 Classificació de cartes

Un cop cada carta està segmentada (normalitzada en vista frontal), cal identificar-la. En el context de jocs de cartes, es consideren principalment dues estratègies: aprendre un model amb CNNs o comparació de característiques.

Xarxes neuronals convolucionals (CNN) amb Transfer Learning:

Les CNN són avui la base de molts sistemes de classificació d'imatges. Es pot emprar una arquitectura preentrenada i reentrenar-la amb un conjunt d'imatges de cartes. Un estudi sobre classificació de cartes genèriques (52 cartes de pòquer) va construir una CNN personalitzada i va obtenir 100% d'accuracy en test amb augmentació de dades [5]. Aquestes aproximacions requereixen tenir un dataset anotat de cartes o personatges, i normalment s'incrementa mitjançant *data augmentation* tal com es va fer en els experiments amb cartes de pokér on amb només 10 fotos originals de cada carta es van generar milers d'imatges augmentades [5].

Aprentatge "one-shot" (siamese/triplet nets): Quan hi ha milers de classes (cada carta diferent) però pocs exemples, es poden usar xarxes siamese o triplet. Per exemple, en un projecte de Yu-Gi-Oh amb més de 10.000 cartes, es va usar un siamese/triplet amb ResNet101 que aprèn a comparar similituds entre imatges enlloc de fer classificació directa de classe [6]. Aquest enfocament aprèn una "distància" entre imatges, sent capaç de recuperar la carta més semblant a la consulta. Al final, en casos marginals es pot reforçar la predicció amb un sistema de característiques clàssiques.

Comparació de característiques locals (feature matching):

Es pot descriure cada carta mitjançant punts clau i descriptors i després buscar coincidències amb una base de dades de cartes referència. Per exemple, en el projecte de Yu-Gi-Oh abans esmentat, es va usar l'algorisme ORB per a reorganitzar els resultats de la CNN, això va servir com a suport quan la xarxa neuronal no era concloent [6]. D'altres projectes han usat tècniques similars: un usuari va fer servir ORB+BruteForce per buscar la millor carta de la base de dades coincidint descriptors, i fins i tot **perceptual hashing** per identificar cartes de *Magic* [4]. Aquest mètode resulta eficient quan la col·lecció és fixa: es pre-calculen les característiques (o hashes) de les cartes conegudes, i es compara contra la imatge segmentada.

OCR complementari: Molts jocs de cartes mostren informació textual (nom de la carta, nombre de còpies, valors numèrics) que es poden llegir amb un OCR. Per exemple, alguns sistemes primer segmenten la carta i llavors apliquen Tesseract OCR sobre el text imprès (per exemple, el tipus o unitats de vida de la carta) per ajudar a identificar-la [7].

4 RISCOS

Cas	Risc	Contingència	Probabilitat
Problemes amb la segmentació de la imatge	Alt	Investigar i provar diferents tècniques de segmentació (detecció de contorns, deep learning) i realitzar proves exhaustives	Mitjana
Problemes en el reconeixement automàtic de cartes per visió per computador	Alt	Provar i ajustar diversos models (CNN, template matching) per trobar la solució més efectiva	Mitjana
Error en la detecció del nombre de còpies per carta	Mitjà	Implementar múltiples tècniques OCR i ajustar el pre-processament per millorar la precisió	Baixa
Planificació temporal insuficient per completar totes les tasques	Alt	Revisar el cronograma amb el tutor, replanificar objectius i ajustar les dates si cal	Alta

TAULA 1: Possibles riscos del projecte

5 REQUISITS SOFTWARE / HARDWARE

Els requisits mínims per poder córrer l'aplicació *Pokémon TCG Pocket* son el següents:

Plataforma	Requisits Mínims
iOS	iPhone 8 o superior
	iOS 13 o superior
Android	2GB de RAM o més
	Android 7.0 o superior
	Arquitectura 64-bit
	Suport per Vulkan
	3GB de RAM o més
	GPU amb rendiment similar al Snapdragon 845 o superior

TAULA 2: Requisits mínims de Pokémon TCG Pocket

Tots els experiments i desenvolupaments s'han fet en un ordinador amb les següents característiques:

CPU: AMD Ryzen 7 5700X3D
GPU: NVIDIA GeForce RTX 3070
RAM: 32 GB DDR4
Sistema Operatiu: Windows 11 Pro Versió 10.0.22631

6 STACK DE PROGRAMACIó I EINES DE DESENVOLUPAMENT

IDE: Visual Studio Code
Llenguatge de programació: Python
Biblioteques de visió per computador: Cv2, Numpy, Pyplot, Tensorflow, Pytesseract

7 PLANIFICACIó

A continuació es presenta la planificació del projecte dividida en quatre fases:

Fase 1: Recerca i Definició de Requisits

En aquesta fase es van recollir les fonts d'informació i es va definir amb precisió el problema a resoldre. Es van delimitar els requisits funcionals i no funcionals, es van establir els objectius principals i es va proposar la metodologia a seguir.

Fase 2: Desenvolupament del mòdul de segmentació

Aquesta fase es va centrar en la implementació i optimització dels algoritmes de preprocessament i detecció de regions de cartes de les captures de pantalla. L'objectiu era obtenir retalls individuals de cada carta, que servien com a entrada per a les fases posteriors de classificació i detecció de còpies.

Fase 3: Desenvolupament del mòdul de classificació

Aquesta fase es va centrar en la implementació i l'avaluació de les metodologies per classificar cada carta segmentada. Es va treballar en la preparació dels conjunts de dades per a l'entrenament i avaluació, incloent l'organització de la biblioteca de referència i l'etiquetatge manual de retalls de captures de pantalla reals. L'objectiu era determinar la metodologia més efectiva per a la tasca d'identificació de cartes.

Fase 4: Desenvolupament del mòdul del nombre de còpies de cada carta

Aquesta fase es va dedicar a la implementació del sistema per determinar el nombre d'exemplars de cada carta identificada, procesant l'imatge de forma adequada i integrant un mòdul OCR per extreure el valor numèric corresponent.

La planificació completa, amb la divisió en fases i tasques, es pot consultar en el diagrama de Gantt inclòs a l'Apèndix A1.

8 ESTRATÈGIES DE SEGMENTACIÓ DE CARTES A PARTIR DE CAPTURES DE PANTALLA

8.1 Segmentació Basada en Tècniques Clàssiques de Processament d'Imatges

Objectiu Inicial: La primera estratègia implementada tenia com a finalitat aïllar les regions corresponents a les cartes dins de les captures de pantalla utilitzant una seqüència d'operacions clàssiques de processament d'imatges. Es buscava transformar la imatge original per ressaltar les cartes com a objectes diferenciats del seu entorn.

A continuació es detallarà la pipeline de processament efectuada (també es pot veure a l'Apèndix A2):

Pre-retallada Vertical: Com a pas inicial, es va aplicar una retallada a les zones superior i inferior de la imatge original. Aquesta operació es va basar en proporcions fixes respecte a l'alçada total de la captura, amb l'objectiu d'eliminar elements de la interfície d'usuari que no contenien informació rellevant per a la detecció de cartes.

Conversió a Escala de Grisos: La imatge retallada es va transformar d'una representació en color a una en escala de grisos. Aquest pas redueix la complexitat de la imatge a un sol canal d'intensitat lumínica, simplificant els anàlisis posteriors.

Suavitat de la Imatge (Desenfocament Gaussiana): Per tal de minimitzar l'impacte del soroll i homogeneïtzar lleugerament les textures, es va aplicar un filtre de desenfocament gaussià. Aquesta tècnica consisteix a fer una mitjana ponderada dels píxels veïns, on els píxels més propers tenen més pes, resultant en una imatge més suau i facilitant la detecció consistent de regions o vores significatives.

Binarització (Llindarització): Posteriorment, la imatge suavitzada es va convertir en una imatge binària, és a dir, una imatge composta únicament per píxels blancs i negres. Això es va aconseguir mitjançant una operació de llindarització: es va definir un valor d'intensitat llindar; tots els píxels amb una intensitat superior a aquest llindar es van convertir a blanc i els inferiors o iguals a negre. L'objectiu era crear un contrast màxim entre les cartes i el seu fons.

Consolidació d'Objectes Mitjançant Ompliment de Forats Interns: Després de la binarització, era comú que les regions corresponents a les cartes presentessin "forats", especialment si contenien text o il·lustracions amb colors similars al fons original. Per tal de tractar les cartes com a objectes sòlids i facilitar la detecció dels seus contorns externs, es van explorar tècniques per omplir aquests forats interns.

El primer mètode es va basar en l'anàlisi de la jerarquia de contorns. Aquesta tècnica primer identifica tots els contorns presents a la imatge binaritzada. L'algorisme pot

determinar quins contorns són externs i quins són interns. Un cop identificats els contorns interns, es "pinten" del color de l'objecte principal, efectivament omplint el forat.

El segon mètode va utilitzar **operacions morfològiques**, concretament l'operació de "Tancament". El tancament consisteix a aplicar primer una dilatació i després una erosió a la imatge amb un element estructurant (kernel). La dilatació tendeix a expandir les regions blanques i a omplir petits forats negres interns, mentre que l'erosió posterior ajuda a restaurar la mida original dels objectes sense reobrir els forats que s'han omplert. La mida del kernel influeix en la mida dels forats que es poden omplir.

Detecció de Vores (Algorisme de Canny): Amb l'objectiu d'identificar els límits precisos de les cartes, es va aplicar l'algorisme de detecció de vores de Canny. Aquest és un mètode de múltiples etapes que busca trobar canvis bruscos d'intensitat a la imatge. Inclou passos com el suavitzat per reduir el soroll, el càlcul del gradient d'intensitat, la supressió de no-màxims (per aprimar les vores) i la llindarització per histèresi. El resultat és una imatge binària on només els píxels que formen vores significatives són blancs.

Identificació i Filtratge de Regions Candidates: Finalment, sobre la imatge de les vores detectades per Canny, es va procedir a buscar contorns tancats. Cada contorn trobat es considerava una regió candidata a ser una carta. Per discriminar entre les cartes reals i altres elements o soroll, es van aplicar dos filtres principals:

Filtre per Àrea del Contorn: Es va calcular l'àrea real del contorn. Es van descartar els contorns amb una àrea massa petita o excessivament gran.

Filtre per Relació d'Aspecte del Requadre Delimitador: Per a cada contorn que superava el filtre d'àrea, es calculava el seu requadre delimitador mínim (el rectangle més petit que l'envoltava), la relació entre l'amplada i l'alçada d'aquest requadre i es comparava amb la relació d'aspecte típica d'una carta de Pokémon, amb un cert marge de tolerància.

8.1.1 Resultats Esperats i Limitacions Crítiques Observades

L'objectiu d'aquesta metodologia era obtenir una llista de requadres corresponents a cada carta. No obstant això, es van identificar dues limitacions principals que comprometien la seva efectivitat:

Problemes amb Cartes de Colors Clars: Com es pot observar a la figura 1, les cartes amb una gran proporció de colors clars (com les de Pokémon de tipus Normal, que són majoritàriament blanques) sovint es confonien amb el fons durant l'etapa de binarització. El llindar global aplicat, pensat per separar les cartes del fons general de l'aplicació, feia que aquestes cartes clares perdessin gran part de la seva estructura o es fusionessin amb el fons, impeding la

detecció correcta de les seves vores i contorns.

Interferència d'Elements de la Interfície d'Usuari (UI):

Com es pot veure a la figura 3, petits elements gràfics de la interfície de l'aplicació, com la barra de desplaçament lateral o botons d'opcions situats a la part inferior, sovint se superposaven parcialment a les cartes. Aquests elements interferien amb la detecció de vores o es fusionaven amb els contorns de les cartes, resultant en una segmentació incorrecta (cartes retallades o amb parts de la UI incloses). Això afectava regularment la identificació d'una o dues cartes per captura.

Aquestes limitacions, especialment la dificultat per gestionar de manera robusta la varietat de colors de les cartes i les oclusions parcials per elements de la UI, van evidenciar que aquest enfocament clàssic de segmentació no oferia la fiabilitat necessària per el projecte. Es va concloure que es requeria una metodologia alternativa, menys dependent d'una binarització global precisa i més adaptada a l'estructura específica de l'aplicació.

8.2 Segmentació basada per Columnes i Morfologia Horitzontal

Per superar les limitacions identificades en la primera aproximació, es va desenvolupar una segona estratègia de segmentació. Aquesta nova metodologia es va dissenyar amb l'objectiu de ser més robusta a les variacions internes de les cartes i menys dependent d'una detecció de vores externes perfectes. Es va basar en dues idees principals: l'aprofitament de l'estructura espacial coneguda de l'aplicació, on les cartes apareixen organitzades en columnes verticals, i l'ús d'operacions morfològiques per consolidar les regions de les cartes en blocs sòlids abans de la seva detecció individual.

A continuació es detallarà la pipeline de processament efectuada (es pot consultar l'Apèndix A3):

Pre-retallada Vertical: De manera similar a l'aproximació anterior, la imatge de la captura de pantalla es carrega i se li aplica una retallada inicial a les zones superior i inferior. Aquesta operació, basada en proporcions fixes, elimina elements irrelevants de la interfície, concentrant el processament a l'àrea on es troben les cartes.

Conversió a Escala de Grisos i Binarització: La imatge retallada es transforma a escala de grisos i posteriorment es binaritza. Es va utilitzar un llindar fix i una inversió per intentar que les cartes es representessin com a objectes blancs sobre un fons negre. Aquesta representació és adequada per a les operacions morfològiques posteriors que busquen connectar i omplir regions blanques.

Operacions Morfològiques Horitzontals per a la Consolidació de Cartes: Aquesta és l'etapa clau d'aquesta metodologia. L'objectiu és transformar cada carta, que pot contenir múltiples detalls interns (text, il·lustracions, símbols), en un bloc horitzontal sòlid.

Primer Tancament Morfològic (Closing): S'aplica una operació de tancament a la imatge binària utilitzant un element estructurant (kernel) rectangular que és significativament més ample que alt. L'efecte d'aquest kernel horitzontalment allargat és fusionar components blancs propers en la direcció horitzontal. Això ajuda a connectar els diferents elements interns d'una carta, creant una "barra" o bloc horitzontal que representa a quina altura està la carta. L'alçada petita del kernel evita la fusió de cartes diferents que estiguin separades verticalment.

Inversió Intermitja: Després del primer tancament, la imatge s'inverteix. Ara, les "barres" que representen les cartes són negres sobre un fons blanc.

Segon Tancament Morfològic: S'aplica una segona operació de tancament, aquest cop sobre la imatge invertida. Es torna a utilitzar un kernel horitzontalment ample, però ara amb una alçada mínima (1 píxel). Aquest segon tancament té l'efecte d'omplir petits forats o discontinuïtats que puguin haver quedat dins de les barres negres, fent-les més sòlides i contínues. Al final d'aquesta etapa, s'espera tenir una imatge on les cartes apareguin com a blocs negres horitzontals ben definits sobre un fons blanc.

Anàlisi Estructurada per Columnes: Aprofitant el coneixement que les cartes es mostren en columnes a l'aplicació, la imatge processada morfològicament es divideix virtualment en diverses columnes verticals. Les coordenades horitzontals de cada columna es van definir mitjançant proporcions (ràtios) respecte a l'amplada total de la imatge.

Per a aquest projecte, es va treballar principalment amb la identificació de tres columnes centrals.

Dins de cada tira de columna aïllada, s'analitza fila per fila (verticalment) la presència de píxels negres (que corresponen als blocs de cartes). Es detecten els inicis i finals dels segments verticals continus on hi ha píxels negres. Cada un d'aquests segments es considera un bloc candidat a ser una carta.

Filtrat Final per Relació d'Aspecte: Per a cada bloc candidat detectat, es va calcular la seva relació d'aspecte. Aquest ràtio es va comparar amb la relació d'aspecte esperada per a una carta de Pokémon, permetent un cert marge de tolerància. Només aquells blocs que complien aquest criteri dimensional es van considerar com a cartes vàlidament segmentades. Addicionalment, es va aplicar un filtre per descartar blocs amb una alçada insignificant.

8.2.1 Avantatges Observats Respecte a la Primera Aproximació

Robustesa a Cartes Clares: La morfologia horitzontal, en buscar connectar qualsevol element de la carta (per contrastat que sigui amb el seu entorn immediat), tendeix a formar blocs sòlids fins i tot si grans àrees de la carta són clares. No depèn tant d'un llindar global per definir tota la carta.

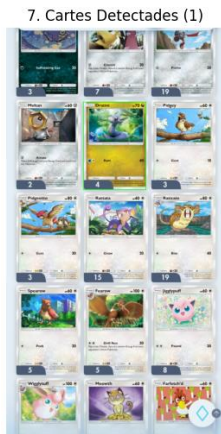


Fig. 1: Només detectem la carta que no es de color blanc

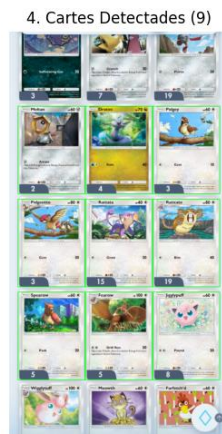


Fig. 2: Es detecten totes les cartes, incloent les clares

Tolerància a Petites Oclusions de la UI: Com que el mètode busca blocs sòlids dins de columnes definides i després filtra per ràtio, petites oclusions per elements de la UI a les vores o cantonades tenen menys probabilitat d'impedir completament la detecció de la carta, sempre que la part principal de la carta formi un bloc recognoscible.



Fig. 3: No detectem la carta d'abaix a la dreta per elements de la interfície



Fig. 4: Es detecten totes les cartes, incloent les obstuïdes per elements de la UI

Menys Sensibilitat al Soroll Intern: La morfologia ajuda a "netejar" i consolidar la representació de la carta.

Degut a la seva major robustesa davant dels problemes identificats en la primera aproximació, especialment en el tractament de cartes amb diferents tonalitats i la gestió d'interferències menors de la interfície d'usuari, aquesta segona metodologia de segmentació per anàlisi columnar i morfologia horitzontal es va seleccionar com el mètode preferit per generar els retalls individuals de cartes.

9 PREPARACIÓ DELS CONJUNTS DE DADES PER A LA CLASSIFICACIÓ DE CARTES

Un cop establert un mètode robust per a la segmentació de

cartes individuals a partir de les captures de pantalla, el següent pas crític va ser preparar els conjunts de dades necessaris per entrenar i avaluar els models de classificació. L'objectiu de la classificació es va definir com la identificació individual de cada carta (el seu "ID de Carta", que conceptualment representa la combinació del set i el número de la carta dins del set).

Es van utilitzar dues fonts principals de dades d'imatges:

Biblioteca d'Imatges de Referència: Es disposava d'una col·lecció d'imatges "netes" i d'alta qualitat [9], representant una instància de referència per a cada ID de carta conegut. Es va generar un índex que assignava un identificador numèric únic a cada ID de carta, essencial per a l'entrenament dels models. Aquest conjunt de dades de la biblioteca es va designar principalment com a font per a l'entrenament dels models de classificació.

Conjunt de Retalls de Captures de Pantalla Etiquetats Manualment: Aquest data set consistia en imatges de cartes individuals obtingudes mitjançant l'aplicació del mètode de segmentació a un conjunt divers de captures de pantalla reals. Aquests retalls van ser posteriorment etiquetats manualment amb el seu ID de carta corresponent utilitzant una eina d'etiquetatge (Label Studio [10]). Aquest conjunt, tot i ser de mida més limitada (300 imatges variades), es va considerar crucial i es va fer servir com a conjunt de test, reservat exclusivament per a l'avaluació final del rendiment de tots els models de classificació desenvolupats, proporcionant una mesura imparcial de la seva capacitat de generalització a dades reals no vistes prèviament durant cap fase d'entrenament.

10 ESTRATÈGIES DE CLASSIFICACIÓ DE CARTES

10.1 Metodologia de Classificació: Xarxes Neuronals Convolucionals (CNN) amb Aprenentatge per Transferència

La primera aproximació explorada per a la classificació dels IDs de carta es va basar en l'ús de Xarxes Neuronals Convolucionals (CNNs), donat el seu èxit demostrat en una àmplia gamma de tasques de visió per computador. Concretament, es va optar per l'Aprenentatge per Transferència. Aquesta tècnica permet aprofitar el coneixement après per models de CNN complexos que han estat pre-entrenats en conjunts de dades massius per a tasques de classificació d'imatges generals.

10.1.1 Arquitectura del Model

Es va seleccionar l'arquitectura EfficientNetB0 [8] com a model base pre-entrenat, coneguda per un bon equilibri entre precisió i eficiència computacional. Les capes convolucionals d'aquest model base es van mantenir inicialment "congelades" (els seus pesos no es modificaven durant les primeres etapes de l'entrenament).

Sobre aquest extractor de característiques base, es va afegir un nou "capçal de classificació" personalitzat. Aquest consistia en una capa de Dropout (una tècnica de

regularització per prevenir el sobre ajustament, desactivant aleatòriament neurones durant l'entrenament) seguida d'una capa Dense (totalment connectada) amb un nombre de neurones igual al nombre total d'IDs de carta únics a classificar, i una funció d'activació softmax per obtenir una distribució de probabilitat sobre les possibles classes.

10.1.2 Augmentació de Dades

Per compensar la limitació de mostres úniques a la biblioteca d'imatges de referència i per millorar la robustesa del model a les variacions visuals esperades en els retalls de captures reals, es va aplicar una sèrie d'**augmentacions de dades** durant l'entrenament. Les imatges de la biblioteca (utilitzades per a l'entrenament) es transformaven aleatòriament en cada època, aplicant operacions com petites rotacions, escalats (zoom), translacions, borrositat i ajustos de brillantor i contrast. L'objectiu era que el model aprengués a identificar les cartes basant-se en característiques invariants a aquestes petites pertorbacions.

10.1.3 Procés d'Entrenament

Fase Inicial: El model es va entrenar inicialment amb les capes del EfficientNetB0 [8] base congelades, de manera que només s'ajustaven els pesos del capçal de classificació afegit. L'entrenament es va realitzar utilitzant les imatges de la biblioteca (amb augmentació) i es va validar amb el conjunt d'imatges de la biblioteca sense augmentar.

Fase de "Fine-Tuning": Posteriorment, es va procedir a "descongelar" totes les capes del EfficientNetB0 [8] base, permetent que els seus pesos també s'actualitzessin. L'entrenament va continuar amb una taxa d'aprenentatge considerablement més baixa. Aquesta fase de fine-tuning té com a objectiu que les característiques pre-apreses pel model base s'adaptin de manera més específica a les particularitats visuals del conjunt de dades de cartes.

10.2 Metodologia de Classificació: Reconeixement per Comparació de Característiques Locals (Feature Matching)

Com es detallarà a la secció de Resultats, tot i que l'aproximació basada en Xarxes Neuronals Convolucionals va demostrar una capacitat de generalització considerable amb els retalls de captures reals, es va considerar valuós explorar metodologies de classificació alternatives que poguessin oferir beneficis addicionals en termes de simplicitat d'implementació o rendiment superior per a les característiques específiques de les dades.

10.2.1 Algorismes i Tècniques Clau Utilitzades

Detecció i Descripció de Característiques Locals amb ORB: Es va seleccionar l'algorisme per extreure punts clau distintius i els seus descriptors binaris associats de cada imatge.

Emparellament de Descriptors amb BFMatcher: Es va utilitzar un comparador de força bruta (BFMatcher) amb la

mètrica de distància de Hamming per trobar correspondències entre els descriptors d'una imatge de consulta i els de les imatges de la biblioteca.

Filtrat de Coincidències amb el Test de Ràtio de Lowe: Es va aplicar el test de ràtio de Lowe per seleccionar només les coincidències més fiables i distintives, descartant les ambigües.

10.2.2 Procés de Classificació

Fase de Pre-càlcul (Biblioteca): Per a cada imatge de la biblioteca de referència, es van extreure i emmagatzemar els seus descriptors ORB, associats al seu ID de carta.

Fase de Consulta (Retall de Captura): Per a un retall de captura a classificar, se li van extreure els seus descriptors ORB.

Comparació i Puntuació: Els descriptors del retall es van comparar amb els de cada imatge de la biblioteca, comptant el nombre de "bones coincidències" obtingudes després d'aplicar el test de ràtio.

Predicció: L'ID de carta de la imatge de la biblioteca amb el major nombre de bones coincidències es va assignar com a predicció.

11 DETECCIó DEL NOMBRE DE CòPIES DE CARTES

Una vegada les cartes han estat segmentades i identificades pel seu ID, el següent objectiu és determinar el nombre d'exemplars de cada carta que l'usuari posseeix. Aquesta informació es presenta a l'aplicació del joc mitjançant un número petit, de color blanc i envoltat d'una forma gris fosc, situat a la cantonada inferior esquerra de cada carta. Per a l'extracció d'aquest valor, es va dissenyar una pipeline de processament d'imatges i OCR enfocada a aquesta regió específica.

El procés per a la detecció del nombre de còpies d'una carta individual segueix els passos següents (Apèndix A4):

Definició de la Regió d'Interès (ROI): S'inicia amb el retall de la carta segmentada. Es defineix una Regió d'Interès (ROI) específica a la cantonada inferior esquerra d'aquesta imatge. Aquesta ROI es calcula mitjançant percentatges de l'amplada i l'alçada de la carta. Això assegura que el procés se centra exclusivament en l'àrea on s'espera trobar el número de còpies, reduint el soroll i la complexitat per a l'OCR.

Aïllament de la Forma Grisa de Fons: La imatge de la ROI es converteix a escala de grisos. Es realitza una llindarització de nivell de grisos utilitzant un rang de valors específic. Aquesta operació crea una màscara binària, on els píxels corresponents a la forma grisa són blancs i la resta són negres. Posteriorment, es detecten els contorns en aquesta màscara. Es busca el contorn de l'àrea més gran que s'ajusti als paràmetres d'àrea esperats per a aquesta forma grisa, i es crea una màscara refinada a partir

d'aquest contorn. Aquesta màscara actuarà com un filtre precís per a la següent etapa.

Preparació del Número per a l'OCR (Binarització i Mascarament): La ROI original (en escala de grisos) es binaritza de nou, aquesta vegada amb un llindar fix, amb l'objectiu d'obtenir el número blanc de les còpies sobre un fons fosc. Aquesta imatge binaritzada del número es combina amb la màscara refinada obtinguda en el pas anterior mitjançant una operació lògica. El resultat és una imatge on només el número de còpies, situat dins de la forma grisa aïllada, és visible. Això elimina eficaçment qualsevol altre element que pugui haver quedat a la ROI inicial.

Ajust Final per a l'OCR: La imatge resultant es inverteix. Això és un preprocesament comú per a molts motors OCR, que sovint funcionen de manera òptima amb text negre sobre fons blanc.

Reconeixement Òptic de Caràcters (OCR): La imatge finalment processada es passa a un motor de Reconeixement Òptic de Caràcters (OCR), concretament Tesseract OCR.

Es va utilitzar una configuració específica per a Tesseract (--psm 6, que assumeix un sol bloc de text uniforme, i un filtre de caràcters per acceptar només dígit 0-9) per optimitzar la lectura de números aïllats.

12 RESULTATS

12.1 Resultats de la segmentació de cartes

La fase de segmentació és fonamental per aïllar les cartes individuals de les captures de pantalla abans de procedir a la seva identificació. Es van avaluar dos enfocaments principals per a aquesta tasca sobre un conjunt de 115 captures de pantalla, les quals contenien un total de 1344 cartes.

12.1.1 Resultats de la Primera Aproximació: Segmentació Basada en Detecció de Vores (Canny) i Anàlisi de Contorns

Per avaluar quantitativament aquesta metodologia, es va realitzar una optimització sistemàtica dels seus paràmetres clau (mida del kernel gaussià, llindars de binarització i Canny, i tipus d'ompliment de forats) sobre el conjunt de 115 captures de pantalla. Les proves van revelar que el millor rendiment obtingut per aquesta aproximació va ser la detecció d'un màxim de 554 cartes del total de 1344 cartes presents en el conjunt de proves. Això representa una taxa de detecció d'aproximadament el 41.2%.

Es va observar que els dos mètodes d'ompliment de forats van oferir un rendiment pràcticament idèntic, amb 554 i 552 cartes detectades respectivament en les seves configuracions òptimes. L'optimització va indicar que, amb una configuració de `gaussian_kernel_size=(1,1)` i `threshold_value=190`, la variació dels llindars de Canny dins dels rangs provats tenia un impacte limitat en el nombre total de cartes detectades, suggerint que les

limitacions del mètode es trobaven en etapes prèvies o en la seva concepció fonamental.

Aquest resultat quantitatiu corrobora les limitacions qualitatives prèviament identificades, demostrant que aquesta aproximació no era prou robusta per a les necessitats del projecte.

12.1.2 Resultats de la Segona Aproximació: Segmentació per Anàlisi Columnar i Morfologia Horitzontal

La metodologia de segmentació per anàlisi columnar i morfologia horitzontal es va avaluar sobre el mateix conjunt de 115 captures de pantalla que contenien 1344 cartes.

Aquest enfocament va aconseguir detectar 1343 de les 1344 cartes presents, la qual cosa representa una taxa de detecció del 99.93%.

12.1.2 Comparativa de Resultats de la Segmentació

La Taula 3 (Apèndix A5) resumeix la comparativa de rendiment de les dues estratègies de segmentació:

Aquesta comparativa empírica justifica de manera concloent l'elecció del mètode de segmentació per anàlisi columnar i morfologia horitzontal com la metodologia preferida per al projecte. Els retalls de cartes generats per aquesta segona aproximació són els que s'han utilitzat com a entrada per a les etapes posteriors de classificació d'ID de carta i detecció del nombre de còpies.

12.2 Resultats de la Classificació d'IDs de Carta

La fase de classificació té com a objectiu principal assignar un ID de carta correcte a cada retall de carta segmentat. Es van explorar dues metodologies principals: les Xarxes Neuronals Convolucionals (CNN) i el Reconeixement per Comparació de Característiques Locals (Feature Matching). L'avaluació es va realitzar sobre un conjunt de test de 300 retalls de captures de pantalla reals etiquetats manualment, que representa fidelment les dades que el sistema haurà de processar en un entorn d'ús real.

12.2.1 Resultats de la Classificació amb Xarxes Neuronals Convolucionals (CNN)

L'enfocament basat en CNN va utilitzar l'aprenentatge per Transferència amb l'arquitectura EfficientNetB0 [8]. El model es va entrenar amb el conjunt complet d'imatges de la biblioteca de referència, i la seva generalització es va avaluar amb els retalls de captures reals.

La Taula 4 (Apèndix A6) resumeix l'impacte de diferents configuracions d'augmentació de dades en la precisió del model. La precisió sobre el "Test Net 845" es el rendiment amb imatges netes de la biblioteca vistes prèviament, mentre que la precisió sobre els "Retalls Etiquetats (300)" és el rendiment crucial sobre les dades reals del projecte.

L'anàlisi d'aquests resultats revela diversos punts clau:

Necessitat de l'Augmentació: Sense cap tipus

d'augmentació, el model aconseguia una precisió molt alta en el conjunt de dades netes de la biblioteca (98.70%), però la seva precisió sobre els retalls reals queia significativament (61.33%). Això confirmava el desajust de distribució i la tendència al sobreajustament.

Impacte Crític del Blur: L'addició del blur gaussià va ser el factor més determinant per millorar la generalització als retalls de captures. Les configuracions amb desenfoç van augmentar la precisió en els retalls etiquetats fins a superar el 70% i, en el millor dels casos, el 90%. Això es deu a que el desenfoç simula la pèrdua de nitidesa i les imperfeccions esperades en les captures reals.

Millor Rendiment de la CNN: La configuració òptima trobada per a la CNN va ser aquella que incloïa augmentació de dades amb desenfoç i translació, assolint una precisió del 90.67% en el conjunt de test de retalls de captures etiquetats. La translació és important perquè el mètode de segmentació per columnes a vegades es menja el primer 5% de la imatge.

12.2 Resultats de la Classificació amb Reconeixement per Comparació de Característiques Locals (Feature Matching - ORB)

La metodologia de Feature Matching, basada en la detecció de característiques ORB i l'emparellament, es va avaluar sobre el mateix conjunt de test de 300 retalls de captures de pantalla reals etiquetats. Es van provar diverses configuracions amb el BFMatcher i amb FLANN per optimitzar tant la precisió com el temps de processament.

La Taula 5 (Apèndix A7) resumeix els resultats de les proves de Feature Matching. Per a les configuracions FLANN, els paràmetres d'índex per a LSH es van establir generalment com a `table_number=6`, `key_size=12`, `multi_probe_level=1` i `checks=10`.

Precisió Excepcional del BFMatcher: El BFMatcher va aconseguir un 100% de precisió fins i tot amb 100 característiques, demostrant una capacitat de discriminació perfecta amb aquesta configuració.

Optimització de Velocitat amb N_FEATURES: La reducció del nombre de característiques té un impacte dràstic en el temps de processament, sense sacrificar la precisió fins a certs llindars.

Rendiment Sòlid de FLANN: El FLANN va mantenir una precisió del 99.67% en la majoria de configuracions, sent significativament més ràpid que el BFMatcher per a `N_FEATURES` similars. La configuració FLANN amb `N_FEATURES=125` és l'opció més equilibrada amb un temps d'inferència de només 0.012 segons per carta i una precisió del 99.67%.

Anàlisi de l'Error Residual (el 99.67%): S'ha investigat el motiu de la pèrdua del 0.33% de precisió en les configuracions FLANN (299/300). L'única carta mal classificada en tots els casos de 99.67% amb FLANN va ser la carta amb ID `A1a_063`. En aquests casos, mentre que el BFMatcher amb `N_FEATURES` iguals o superiors era capaç de classificar-la correctament, FLANN, tot i la seva eficiència, va tenir dificultats amb aquesta instància

particular. Això suggereix que, tot i la robustesa general de FLANN, pot haver-hi casos límit on la cerca aproximada pot no trobar la millor coincidència veritable o que la distintivitat de les característiques d'aquesta carta pot ser més subtil per a l'algorisme LSH.

12.2 Comparativa i Discussió Final de les Metodologies de Classificació

La Taula 6 (Apèndix A8) resumeix els resultats comparatius de les configuracions òptimes per a cada metodologia de classificació, avaluades totes elles sobre el mateix conjunt de 300 retalls de captures de pantalla reals.

La comparativa revela que:

Precisió: El Feature Matching supera la CNN en precisió de manera contundent, amb gairebé un 100% de detecció correcta. Aquesta diferència de més del 9% és significativa i crítica per a la fiabilitat d'un sistema automatitzat.

Velocitat d'Inferència: Ambdues metodologies ofereixen temps d'inferència per carta molt baixos. El Feature Matching optimitzat amb FLANN s'ha mostrat extremadament ràpid, situant-lo plenament en l'àmbit del processament gairebé en temps real.

Complexitat i Robustesa: La CNN va requerir una sintonització complexa de la pipeline d'entrenament i augmentació per superar el desajust de distribució. El Feature Matching, en canvi, va demostrar una precisió excepcional amb una configuració relativament més directa, aprofitant la consistència visual de les imatges digitals.

12.3 Resultats de la Detecció del Nombre de Còpies de Cartes

Per avaluar el rendiment d'aquest mòdul, es va realitzar un test amb un subconjunt de 50 retalls de cartes, obtinguts de captures de pantalla i processats pel mòdul de segmentació.

Resultats de l'OCR en Cartes Vàlides: Descomptant els 9 casos invàlids, la prova es va realitzar sobre 41 cartes vàlides on el número era potencialment llegible. De les 41 cartes vàlides, el sistema va fallar en la detecció del nombre de còpies en 3 ocasions. Això resulta en una precisió del mòdul de detecció de còpies del $(41 - 3) / 41 = 38 / 41 \approx 92.68\%$.

Anàlisi de les Limitacions i Errors:

Cartes Tallades: La principal limitació observada és l'existència de retalls on el número de còpies no és completament visible a causa d'un tall en la captura original o una segmentació imperfecta. Aquesta és una limitació inherent a la qualitat de la imatge d'entrada i a la robustesa de la segmentació inicial.

Mida Reduïda dels Dígits: Els números de còpies en els retalls de cartes són, elements visuals molt petits, compostos per un nombre reduït de píxels. Aquesta característica pot dificultar la lectura precisa per part de l'OCR, especialment si hi ha un lleuger desenfoç, compressió o artefactes en la imatge.

13 TREBALL A FUTUR

Tal com s'ha esmentat a la introducció, aquest Treball de Final de Grau constitueix el primer pas d'un projecte més ampli: el desenvolupament d'una aplicació mòbil destinada a facilitar la gestió d'intercanvis entre jugadors de Pokémon TCG Pocket.

En aquest context, i per tal de consolidar i expandir les capacitats del mòdul de visió per computador desenvolupat, que ja ha demostrat una alta precisió en la identificació de cartes, es proposen les següents línies de millora i investigació:

Investigació d'Object Detection: Actualment, la segmentació i la classificació són etapes seqüencials. Una alternativa a explorar seria l'ús de models de Detecció d'Objectes, com ara els basats en la família YOLO (You Only Look Once), que permeten realitzar ambdues tasques en un únic pas.

Detecció de Cartes en l'Obertura de Sobres: Implementar aquesta funcionalitat permetria als usuaris registrar ràpidament les noves adquisicions a la seva col·lecció sense necessitat de capturar la pantalla de la col·lecció completa.

14 CONCLUSIONS

Aquest Treball de Final de Grau ha abordat amb èxit el repte del reconeixement automatitzat de cartes de Pokémon TCG Pocket a partir de captures de pantalla digitals, establint les bases per a futures aplicacions de gestió d'intercanvis.

El projecte s'ha estructurat en tres grans mòduls: segmentació de cartes, identificació d'ID de carta i detecció del nombre de còpies. Per a la segmentació de cartes, es van avaluar dues aproximacions. La primera, basada en tècniques clàssiques de detecció de vores (Canny) i anàlisi de contorns, va demostrar limitacions significatives, detectant només el 41.2% de les cartes totals a causa de la seva sensibilitat a les variacions de color i la interferència d'elements de la interfície d'usuari. En contrast, la segona aproximació, que aprofita l'estructura columnar de l'aplicació i operacions morfològiques horitzontals, va assolir una excepcional taxa de detecció del 99.93%, confirmant-se com la metodologia òptima i robusta per a aquesta etapa.

Per a la identificació de l'ID de carta, es van explorar dues metodologies. L'enfocament basat en CNN amb aprenentatge per transferència i augmentació de dades va aconseguir una precisió del 90.67% sobre els retalls de captures de pantalla reals, demostrant la capacitat de les CNNs per generalitzar amb una configuració d'augmentació acurada.

No obstant això, la metodologia de Reconeixement per Feature Matching, va superar aquest rendiment de manera contundent, assolint una precisió del 99.67% sobre el mateix conjunt de dades. Aquesta superioritat en precisió, combinada amb una alta eficiència d'inferència (0.012 segons per carta), posiciona el Feature Matching com la solució més efectiva i fiable per a aquesta tasca en el context del projecte.

Finalment, el mòdul de detecció del nombre de còpies mitjançant OCR va demostrar una precisió del 92.68% en la lectura dels dígit. Aquest rendiment és totalment acceptable i contribueix a la funcionalitat completa del sistema.

En conjunt, el projecte ha establert una pipeline de visió per computador amb una alta precisió global en la detecció i identificació de cartes a partir de captures de pantalla.

S'han superat reptes significatius com la configuració d'entorns d'aprenentatge automàtic, la gestió de dades limitades per classe, el desajust de distribució entre les dades d'entrenament i les reals, i la robustesa davant les particularitats visuals de les captures del joc.

AGRAÏMENTS

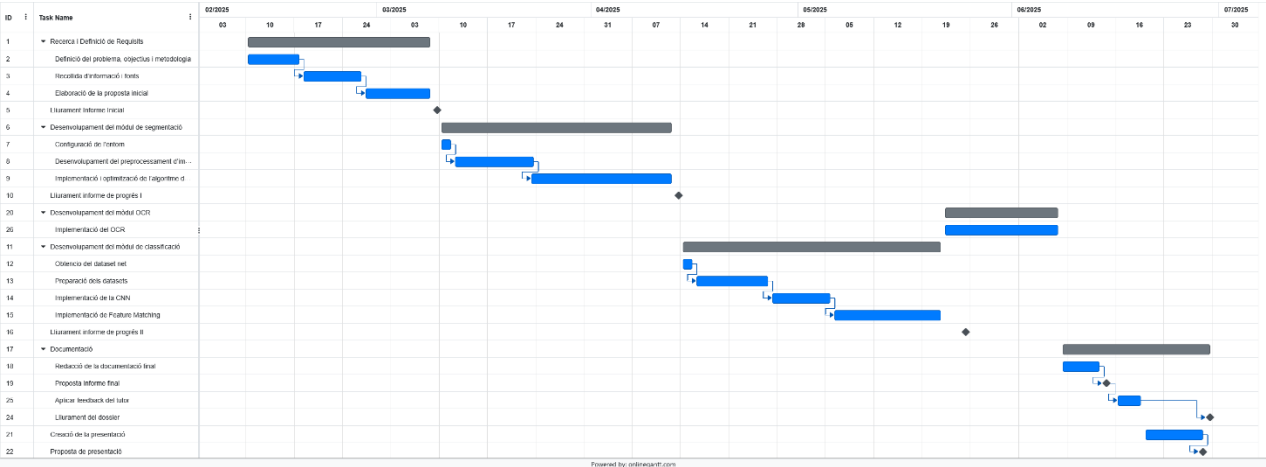
Vull agrair al meu tutor, Jorge Bernal, per la gran comprensió, flexibilitat i ajuda otorgada durant tot el treball. També dono les gràcies a la meua mare, per estar sempre al meu costat. I per últim als meus amics per confiar en mi.

BIBLIOGRAFIA

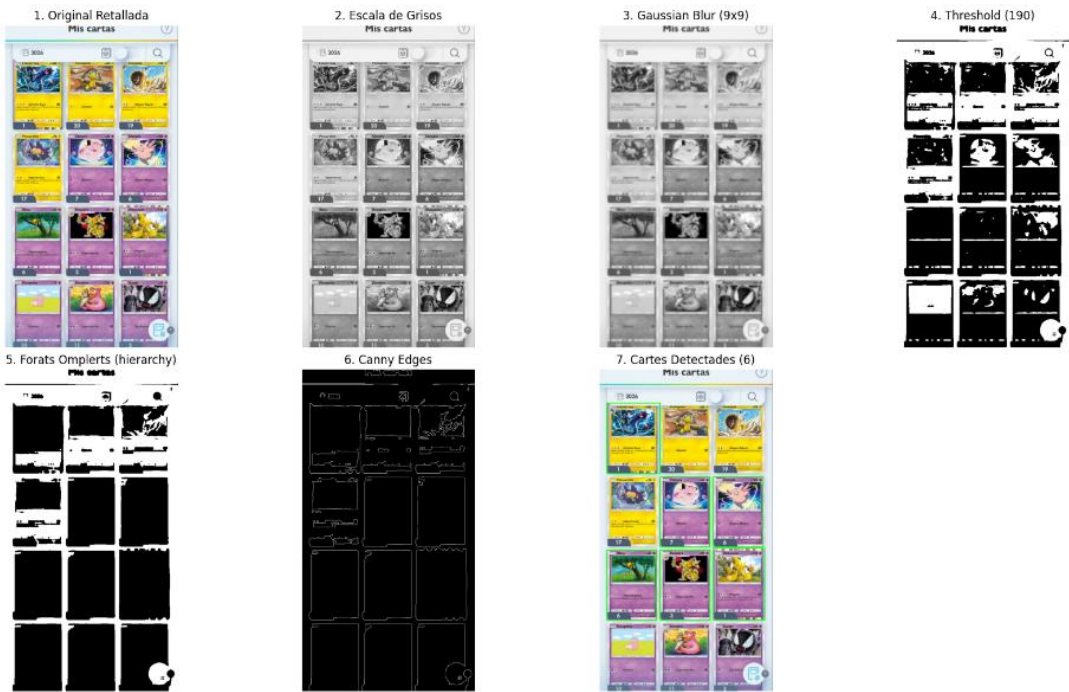
- [1] Pokémon TCG Pocket (@PokemonTCGP), [Tweet]. X (anteriorment Twitter), 14 de març de 2025. [En línia]. Disponible a: <https://x.com/PokemonTCGP/status/1900351172700959147> [Accedit el: 11 d'Abril de 2025].
- [2] tmikonen, "magic_card_detector: Python implementation of a Magic: the Gathering card segmentation and recognition." GitHub repository. [En línia]. Disponible a: https://github.com/tmikonen/magic_card_detector. [Accedit el: 5 de Maig de 2025].
- [3] F. W. Ostad, "OpenCV: Extract horizontal and vertical lines by using morphological operations." LearnOpenCV. [En línia]. Disponible a: <https://learnopencv.com/extract-horizontal-and-vertical-lines-by-using-morphological-operations/>. [Accedit el: 12 de Maig de 2025].
- [4] hj3yoo, "mtg_card_detector." GitHub repository. [En línia]. Disponible a: https://github.com/hj3yoo/mtg_card_detector. [Accedit el: 5 de Maig de 2025].
- [5] P. Perlin, "CS229 Project Report: Magic Card Detector (2017)." Stanford University. [En línia]. Disponible a: <https://cs229.stanford.edu/proj2017/final-reports/5233806.pdf>. [Accedit el: 15 de Maig de 2025].
- [6] A. Lowhur, "An interview with Anthony Lowhur - Recognizing 10,000 Yugioh Cards with Computer Vision and Deep Learning." PyImageSearch. [En línia]. Disponible a: <https://www.pyimagesearch.com/2020/09/21/an-interview-with-anthony-lowhur-recognizing-10000-yugioh-cards-with-computer-vision-and-deep-learning/>. [Accedit el: 28 de Maig de 2025].
- [7] u/redditor_123, "Identifying a Magic The Gathering card." Reddit r/computervision, [Post]. [En línia]. Disponible a: https://www.reddit.com/r/computervision/comments/9k0x0f/identifying_a_magic_the_gathering_card/. [Accedit el: 1 de Juny de 2025].
- [8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105-6114.
- [9] Limitless TCG, "Sets - Limitless TCG Pocket Database" [En línia]. Disponible a: <https://pocket.limitlesstcg.com/-/cards>. [Accedit el: 11 d'Abril de 2025].
- [10] Heartex, "Label Studio." [En línia]. Disponible a: <https://labelstud.io/>. [Accedit el: 10 de Juny de 2025].

APèNDIX

A1. DIAGRAMA DE GANNT AMB TOTA LA PLANIFICACIó



A2. PASOS DEL PROCESAMENT DE L'IMATGE AMB SEGMENTACIó BASADA EN CONTORNS



A3. PRINCIPALS PASOS DEL PROCESAMENT DE L'IMATGE AMB SEGMENTACIÓ BASADA PER COLUMNES I MORFOLOGIA HORITZONTAL



A4. Principals pasos del procesament de l'imatge amb detecció OC



A5. TAULA 3: COMPARATIVA DE RESULTATS DE LES METODOLOGIES DE SEGMENTACIó

Metodologia de Segmentació	Total de Cartes Detectades	Percentatge de Detecció
Detecció de Vores (Canny) i Anàlisi de Contorns	554 / 1344	41.2%
Anàlisi Columnar i Morfologia Horitzontal	1343 / 1344	99.93%

A6. TAULA 4: RENDIMENT DE LA CNN SEGONS LA CONFIGURACIó D'AUGMENTACIó DE DADES

Configuració d'Augmentació	Max Sigma (Blur)	Precisió (Test Net 845)	Precisió (300 Retalls Etiquetats)	Fine-tuning Necessari?
Sense Augmentació	0	98.70%	61.33%	No
Amb Aug. (variada + no blur)	0	60.24%	57.67%	Sí
Amb Aug. (blur 50% de les vegades)	2	80.47%	72.33%	Sí
Amb Aug. (blur 100%)	2	72.78%	73.33%	Sí
Amb Aug. (blur 100%)	1	94.79%	88.33%	No
Amb Aug. (blur 100%) + translació (0.1, 0.1)	0.5	94.67%	90.67%	Sí
Amb Aug. (blur 100%)	0.5	99.88%	83.33%	No
Amb Aug. (variada + blur)	0.5	57.28%	59.00%	Sí

A7. TAULA 5: RENDIMENT DEL RECONeixEMENT PER FEATURE MATCHING (ORB) SEGONS LA CON-FIGURACIÓ

Mètode d'Em- parellament	N_FEA- TURES	Correctes / Total	Precisió (%)	Temps (s)	Temps per Carta (s)	Notes Clau
BFMatcher	1000	300 / 300	100.00%	480.00	1.60	Referència base, alta precisió
BFMatcher	500	300 / 300	100.00%	165.53	0.55	Menys features, mateixa precisió, més ràpid
BFMatcher	250	300 / 300	100.00%	55.98	0.19	Molt ràpid mantenint perfecte
BFMatcher	125	300 / 300	100.00%	26.27	0.088	Altament eficient i precisa
BFMatcher	100	300 / 300	100.00%	19.42	0.065	El més ràpid amb 100% de precisió
BFMatcher	75	296 / 300	98.67%	15.19	0.051	Baixa lleugerament la precisió
FLANN (LSH)	1000	299 / 300	99.67%	71.82	0.239	
FLANN (LSH)	500	299 / 300	99.67%	24.21	0.081	
FLANN (LSH)	250	299 / 300	99.67%	8.48	0.028	Excel·lent equilibri precisió-temps
FLANN (LSH)	125	299 / 300	99.67%	3.60	0.012	Configuració òptima (velocitat i precisió)
FLANN (LSH)	100	298 / 300	99.33%	2.98	0.010	Més ràpid, però perd precisió lleugerament
FLANN (LSH - Paràmetres Alts)	500	299 / 300	99.67%	3251.00	10.84	Paràmetres: checks=100, table_number=20, multi_probe_level=2. Notablement més lent.

A8. TAULA 6: COMPARATIVA DE RENDIMENT DE METODOLOGIES DE CLASSIFICACIÓ (CONFIGURA-CIONS ÒPTIMES)

Mètrica / Metodolo- gia	Xarxes Neuronals Convolu- cionals (CNN)	Reconeixement per Fea- ture Matching (ORB + BFMatcher)	Reconeixement per Feature Matching (ORB + FLANN)
Precisió (Accuracy) en el Conjunt de Test	90.67%	100.00%	99.67%
Temps d'Inferència per Carta (s)	Generalment ràpid (estimació: mil·lisegons)	0.065	0.012
Complexitat d'Imple- mentació/En- trenament	Alta (requereix entrenament iteratiu, GPU, dades)	Moderada (pre-càlcul, sense entrenament iteratiu)	
Robustesa al Des- ajust de Distribució	Bona (amb augmentació es- pecífica)	Excel·lent (per a dades consistents)	
Requisits de Dades d'Entrenament	Biblioteca extensa + augmenta- ció + val/test reals	Biblioteca extensa (sense entrenar directament)	