



TRABALLO FIN DE GRAO
GRAO EN ENXEÑARÍA INFORMÁTICA
MENCIÓN EN COMPUTACIÓN



Axuste adaptativo de rexións de interese na detección de tomates en secuencias de vídeo

Estudiante: Nicolás Martínez González

Dirección: Outro Nome Completo

A Coruña, abril de 2024.

Dedicatoria

Agradecementos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Resumo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Palabras chave:

- First itemtext
- Second itemtext
- Last itemtext
- First itemtext
- Second itemtext
- Last itemtext
- First itemtext

Keywords:

- First itemtext
- Second itemtext
- Last itemtext
- First itemtext
- Second itemtext
- Last itemtext
- First itemtext

Índice Xeral

1	Introdución	1
1.1	Contexto general	1
1.2	Objetivos del proyecto	1
1.2.1	Propuesta	2
1.3	Trabajo relacionado/Estado del arte	2
1.4	Estructura de la memoria	2
2	Fundamentos teóricos	4
2.1	Visión artificial	4
2.2	Aprendizaje automático	4
2.3	Detección de objetos	4
2.4	Deep Learning	4
2.5	CNNs	4
2.6	Transferencia de aprendizaje (Transfer Learning)	4
2.7	EfficientDet	4
2.7.1	Arquitectura	4
2.7.2	Anchors y Compound Scaling	5
2.8	YOLOv5	5
2.8.1	Arquitectura	5
2.9	SSD	5
2.9.1	Arquitectura	5
2.10	Faster R-CNN/FCOS	5
3	Fundamentos tecnológicos	6
3.1	Librerías y lenguajes de programación	6
3.1.1	Python	6
3.1.2	NumPy	7
3.1.3	Tensorflow	7

3.1.4	PyTorch	7
3.1.5	Torchvision	7
3.2	Dataset	7
3.2.1	Forma del Dataset (Pascal VOC)	7
3.2.2	Weighted Boxes Fusion (WBF).	7
3.2.3	Integración de imágenes de otros datasets.	8
3.3	Herramientas auxiliares	8
3.3.1	TensorBoard	8
3.3.2	GitHub	8
3.3.3	LaTeX (Overleaf)	8
4	Metodología	9
4.1	Planificación	9
4.2	Requisitos, actores, casos de uso	9
4.3	Seguimiento + incidencias	9
4.4	Costes	9
4.4.1	Costes de software	9
4.4.2	Costes materiales	9
4.4.3	Costes de recursos humanos	9
5	Arquitectura del sistema?	10
5.1	Análisis teórico-práctica preliminar	10
5.1.1	Herramientas disponibles: tensorflow vs pytorch	10
5.1.2	Modelos disponibles	10
5.1.3	Datasets disponibles	10
5.1.4	Comparativa y discusión !?	10
5.2	Conjuntos de entrenamiento	10
5.3	Adaptaciones del dataset	11
5.4	Ajuste adaptativo de regiones de interés ??	11
6	Pruebas y análisis	12
6.1	Pruebas iniciales	12
7	Conclusiones	14
7.1	Situación final del proyecto	14
7.2	Trabajo futuro	14
A	Material adicional	16

Bibliografía

19

Índice de Figuras

6.1 Inferencias con el modelo YOLOv56s	13
--------------------------------------------------	----

Índice de Táboas

Capítulo 1

Introducción

En este capítulo presentaremos el contexto en el que se plantea el proyecto, los objetivos que persigue además de las propuestas halladas para llevarlos a cabo, junto al estado del arte y la estructura de la propia memoria.

1.1 Contexto general

A lo largo de los últimos años, en específico en la última década, el uso de la inteligencia artificial como herramienta para todo tipo de aplicaciones ha experimentado un aumento muy notable. Podemos encontrarla incluida en todo tipo de contextos: vehículos autónomos, asistentes por voz, sistemas de reconocimiento facial/biométrico, sistemas de recomendación. Un ámbito de aplicación de la inteligencia artificial que ha recibido especial atención desde la mitad del siglo pasado ha sido la robótica aplicada a la agricultura. En este aspecto, los sistemas buscan una automatización del control de la producción agrícola en general (plantado, supervisión del crecimiento, recogida, ...) mediante diversos métodos como la lógica difusa, las redes de neuronas artificiales y los algoritmos genéticos, entre otros. Debido a un auge en la investigación en el campo de la inteligencia artificial a principios de la década de los 2010s, a partir de 2014, un nuevo modelo

1.2 Objetivos del proyecto

Para este proyecto vamos a poner nuestro enfoque en el uso de modelos de redes de neuronas convolucionales(CNNs) para la detección de tomates en secuencias de vídeo. Partiendo de la base de emplear una base de datos ya etiquetada conformada por fotos tomadas en un invernadero los dos principales objetivos de este proyecto son:

- Comparar el rendimiento y el coste de entrenamiento de arquitecturas CNN recientes empleadas para detectar objetos en dispositivos de bajas prestaciones.

- Analizar modificaciones en la arquitectura de alguna CNN para conseguir un ajuste adaptativo o más fino de los umbrales usados para determinar las regiones de interés.

1.2.1 Propuesta

La propuesta a la que pretendemos adherirnos se centra en el uso de modelos de redes neuronales convolucionales ya disponibles, junto con un conjunto de imágenes etiquetadas para llevar a cabo la detección de tomates. En este aspecto cabe describir los principales requisitos del proyecto: - La elección de los modelos debe estar supeditada a alcanzar la máxima eficiencia con los mínimos recursos, tanto en cuestión de espacio para alojar dicho modelo, como en cuestión del coste computacional del entrenamiento del mismo. Lo que se busca es construir un sistema capaz de funcionar en un equipo de bajas prestaciones. - El conjunto de datos debe ajustarse al contexto que pretendemos modelar. Esto es, imágenes fieles a un entorno real con condiciones de visibilidad variables y no siempre variables, y un número de fotos no especialmente amplio.

1.3 Trabajo relacionado/Estado del arte

Comentar ejemplos citados en los informes de las reuniones de las semanas 2-4. - Paper de Irene - Paper de Magalhães

1.4 Estructura de la memoria

En esta sección se explicará la organización seguida para presentar esta memoria.

- Capítulo 1: Introducción. Es el capítulo actual, sirve para presentar las motivaciones detrás de este proyecto, sus objetivos y su descripción general.
- Capítulo 2: Fundamentos teóricos. En este capítulo se comentan los conceptos necesarios para comprender la librería empleada en el proyecto y las familias de modelos de CNN empleadas.
- Capítulo 3: Fundamentos tecnológicos. En este capítulo se explican las tecnologías y herramientas empleadas para llevar a cabo el proyecto.
- Capítulo 4: Metodología. En este capítulo se comentan tanto la planificación del proyecto como los requisitos, desarrollo y costes del mismo.
- Capítulo 5: Diseño e implementación. En este capítulo se razona el diseño final del sistema además de los principales detalles de implementación.

- Capítulo 6: Pruebas. En este capítulo se muestran los resultados de las pruebas ejecutadas sobre el sistema y se analizan dichos resultados.
- Capítulo 7: Conclusiones. En este capítulo final se examina el estado final del proyecto y se comentan posibles líneas de trabajo futuro.

Fundamentos teóricos

2.1 Visión artificial

2.2 Aprendizaje automático

2.3 Detección de objetos

2.4 Deep Learning

2.5 CNNs

Explicar aspectos generales de las cnns. No entrar en detalles muy específicos a nivel de funciones de activación, optimización, hiperparámetros... Marcar líneas generales para desarrollar cálculo de loss más adelante -> optimización con parámetros específicos de pytorch

2.6 Transferencia de aprendizaje (Transfer Learning)

2.7 EfficientDet

2.7.1 Arquitectura

Backbone (EfficientNet), BiFPN, cabezas de predicción de bbox y de clase.

2.7.2 Anchors y Compound Scaling

2.8 YOLOv5

2.8.1 Arquitectura

Backbone (CSPDarknet53), PANet

2.9 SSD

2.9.1 Arquitectura

Backbone (VGG16), extractor de características (VGG16)

2.10 Faster R-CNN/FCOS

Fundamentos tecnológicos

En esta sección comentaremos los contenidos necesarios para comprender la base tecnológica sobre la que se cimienta el proyecto, es decir, tanto los lenguajes de programación, aplicaciones usadas para escribir el código y visualizarlo, además de llevar un historial del mismo, ahondando tanto en el hardware y el software tenidos en cuenta a la hora de desarrollar el proyecto.

3.1 Librerías y lenguajes de programación

Para este trabajo, el principal lenguaje de programación ha sido Python, además de diversas librerías adaptadas a la computación científica y al aprendizaje automático, tales como NumPy, PyTorch, TensorFlow, además de algunas derivadas de las ya mencionadas, TorchMetrics, TorchVision, Pytorch Lightning, entre otras.

3.1.1 Python

Python es un lenguaje de programación de alto nivel interpretado, con un tipado dinámico, caracterizado por su versatilidad y legibilidad. Es un lenguaje que se centra en tener un sintaxis clara y una estructura basada en la indentación, lo cual facilita su comprensión de forma intuitiva. Uno de sus principales puntos fuertes es su amplia biblioteca estándar, además de las específicas elaboradas por su comunidad, debido a la propia naturaleza de código abierto de Python. Este aspecto es el que ha fomentado la inclusión de Python en la creación de bibliotecas y marcos de trabajo. Dichas bibliotecas creadas por la comunidad de Python abarcan muchos ámbitos, encontrándose aplicadas a un gran número de ámbitos tales como el desarrollo web, el análisis de datos o la inteligencia artificial.

3.1.2 NumPy

Numpy es una librería de Python que proporciona un soporte de matrices y arrays multi-dimensionales, así como de varias funciones matemáticas para realizar operaciones de forma eficiente sobre estos tipos de datos. El principal uso derivado de esta librería radica en el uso de arrays para los datos que maneja el modelo (bounding boxes, scores, clases, ...).

3.1.3 Tensorflow

TensorFlow es una de las dos principales librerías de código abierto para aprendizaje automático dentro del ecosistema de Python (junto a PyTorch). Diseñada por Google, permite la creación y entrenamiento de redes neuronales de un gran número de aplicaciones: reconocimiento de imágenes, procesamiento de lenguaje natural, aprendizaje profundo, modelos de clasificación/regresión,... La principal aplicación de esta librería yace en el hecho de que algunos modelos usados como esqueleto para otros son traducciones a PyTorch de modelos nativos de TensorFlow. También cabe destacar que la librería TensorBoard, detallada en adelante, aunque adaptada para su uso de forma conjunta con PyTorch, está diseñada originalmente por TensorFlow.

3.1.4 PyTorch

PyTorch es la otra de las dos librerías más populares de aprendizaje profundo en Python. Tiene un enfoque dinámico, en aspectos de visualización de datos y depuración, además de

3.1.5 Torchvision

3.2 Dataset

Descripción

Hablar de cómo se conformó, condiciones en que se tomaron las fotos, sesgo en las anotaciones (esto quizás comentarlo más adelante).

3.2.1 Forma del Dataset (Pascal VOC)

Describir el esquema del dataset y las modificaciones que tuvieron que aplicarse para adaptarlo al modelo.

3.2.2 Weighted Boxes Fusion (WBF).

Función de postprocesado de resultado en forma de bounding boxes opuesto a Supresión de No Máximos (NMS), con el fin de hallar el área exacta de solape de varias bounding boxes.

3.2.3 Integración de imágenes de otros datasets.

Comentar propuesta de incluir imágenes de otros conjuntos para que sea más robusto. -> Explicar cómo esto hace que se aleje de la idea inicial de que sea fiel a un contexto real.

3.3 Herramientas auxiliares

3.3.1 TensorBoard

Herramienta para cargar información registrada por los modelos en ciertos puntos de los procesos de entrenamiento, validación y test, a fin de poder visualizar parámetros como la precisión, `recall` o la pérdida del modelo.

3.3.2 GitHub

3.3.3 LaTeX (Overleaf)

Capítulo 4

Metodología

4.1 Planificación

Explicar la metodología ágil que hemos seguido. Hablar de la distribución inicial de trabajo y cómo ha cambiado con el tiempo.

4.2 Requisitos, actores, casos de uso

4.3 Seguimiento + incidencias

Reuniones y problemas encontrados -> cambios de planes sobre la marcha

4.4 Costes

4.4.1 Costes de software

Gratis

4.4.2 Costes materiales

Amortización + horas -> pc y rtx

4.4.3 Costes de recursos humanos

Nico -> director, analista, programador Carlos -> consultor

Arquitectura del sistema?

En este capítulo se explican los detalles de la arquitectura del sistema y cómo ciertos aspectos han tenido que modificarse para adaptarse a los modelos usados.

5.1 Análisis teórico-práctica preliminar

5.1.1 Herramientas disponibles: tensorflow vs pytorch

evaluar herramientas para nuestro problema

5.1.2 Modelos disponibles

analizar los diferentes modelos disponibles en los repositorios Explicar diferencias con otras familias de modelos y por qué sólo hemos empleado EfficientDet (u otros).

5.1.3 Datasets disponibles

base de datos de Portugal y otras encontradas

explicar lo de la "secuencia de vídeo": la de Oporto "salen" de un vídeo (se puede observar fácilmente), pero no tenemos el vídeo original para "extraer" otros frames, excepto los ya etiquetados

5.1.4 Comparativa y discusión !?

justificar las decisiones tomadas requisitos

5.2 Conjuntos de entrenamiento

Comentar los mecanismos empleados para dividir los conjuntos de entrenamiento, validación y test. Métricas para compararlos

5.3 Adaptaciones del dataset

Explicar la dicotomía entre usar recortes de las imágenes del tamaño de entrada requerido por el sistema o redimensionarlas.

5.4 Ajuste adaptativo de regiones de interés ??

algoritmo propuesto: - es independiente del modelo (sólo utiliza la salida de cada imagen: el histograma de los scores de los bounding boxes que se han detectado y calcula el umbral mínimo de score válido)

1. estudio de los histogramas de los scores de los bounding boxes detectados por diferentes imágenes (y modelos !?)
2. propuesta de cálculo automático del umbral mínimo de score (y alternativas)

Pruebas y análisis

6.1 Pruebas iniciales

Mostrar resultados de pruebas con modelos sin reentrenamiento

Formato en pruebas, la idea es agrupar las imágenes por familias de modelos probados pero no de esta forma

Imágenes de pruebas:

1. Inferencias con otros modelos sin reentrenamiento (**baseline**)
 - > Inferencias sobre un mismo conjunto de imágenes del dataset de Oporto de varios modelos disponibles en la librería torchvision (FasterRCNN, SSD300_VGG16, RetinaNet, FCOS)
 - tiempos de inferencia (recursos hw), errores, etc etc
2. Inferencias con EfficientDet entrenando todo el modelo con umbrales distintos (prediction_confidence_threshold, wbf_iou_threshold, skip_box_thr)
3. Diferencias entre inferencias de EfficientDet con redimensionado de las imágenes y recortando
4. Reentrenamiento completo de la red con diferentes proporciones de conjuntos de datos de entrenamiento/validación/testeo (40-20-40, 50-20-30, 60-20-20, 70-15-15, 80-10-10).
 - (a) Inferencias de todos los modelos en una imagen
 - (b) Histogramas de tamaños de bounding box para mostrar el posible sesgo estadístico, además de histogramas de confianza en las predicciones.
 - (c) Selección de la mejor opción.
5. Gráficas de TensorBoard (Loss, precisión, recall, ...) => para todos los entrenamientos (variación temporal) e inferencias (por batch !?)

Conclusións

DERRADEIRO capítulo da memoria, onde se presentará a situación final do traballo, as conclusións aprendidas, a relación coas competencias da titulación en xeral e a mención en particular, posibles liñas futuras,...

7.1 Situación final del proyecto

7.2 Trabajo futuro

Apéndices

Material adicional

EXEMPLO de capítulo con formato de apéndice, onde se pode incluír material adicional que non teña cabida no corpo principal do documento, suxeito á limitación de 80 páxinas establecida no regulamento de TFGs.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique

neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Bibliografía
