

Práctica final: clasificación de monedas

Como el propio título de la práctica indica, el objetivo de la práctica final de la asignatura visión artificial (2021-2022) va a ser la elaboración, entrenamiento y validación de un sistema de detección y clasificación de monedas en imágenes.

Para el desarrollo de la práctica, dispondréis de un conjunto de imágenes de train y test, más un conjunto extra de imágenes, con las que podréis entrenar un clasificador. El objetivo de la práctica es que apliquéis todas las técnicas de preprocesamiento, extracción de característica, o cualquier otra técnica vista en clase (o no), con el objetivo de obtener el mejor clasificador.

A continuación, vamos a ir desgranando los aspectos más importantes de la práctica.

Fase 1. Entrenamiento de un clasificador de monedas

Este es la parte más importante de la práctica y a la que más tiempo debéis dedicar. Disponéis de un conjunto de imágenes de train/test dividido en 8 clases: 1 céntimo, 2 céntimos, 5 céntimos, 10 céntimos, 20 céntimos, 50 céntimos, 1 euro y 2 euros.

Fase 1.1. Análisis del conjunto de datos

Una vez descargado del dataset, lo primero que deberías realizar es un análisis inicial de los datos (comúnmente llamado análisis exploratorio de datos) que os sirva para conocer en profundidad el dataset: distribución train/test, número de clases, número de ejemplos por clase, dimensión de las imágenes, etc. En esta fase también podéis realizar un análisis en busca de valores atípicos (outliers), valores anormales, ruido, etc. y aplicar las técnicas que consideréis oportunas para su corrección. El objetivo de esta fase es que tengáis un conocimiento en profundidad del problema que debéis solucionar antes de empezar a aplicar cualquier técnicas de aprendizaje.

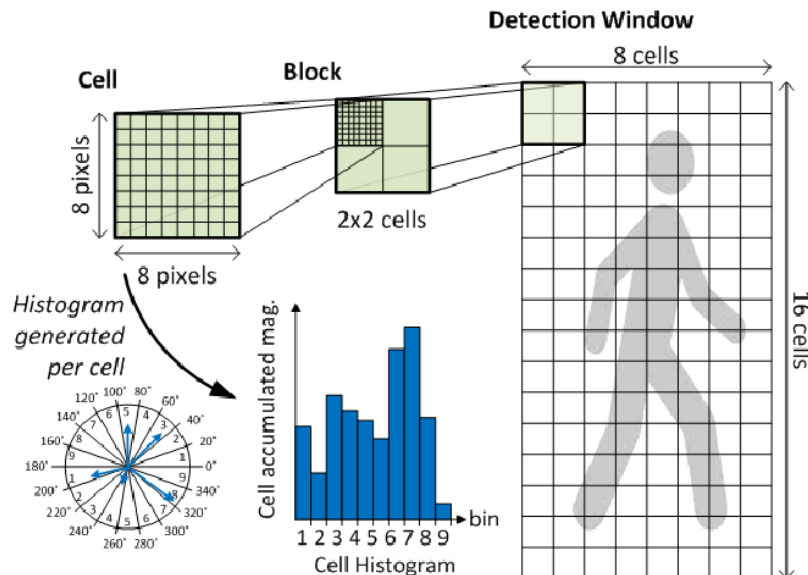
Fase 1.2. Extracción de los datos de entrenamiento

Una vez analizado el dataset y corregidos sus posibles fallos, es hora de generar el conjunto de entrenamiento que va a alimentar al clasificador. Cuando los datos de entrada de un sistema de clasificación supervisado están formados por imágenes, tenemos dos alternativas, principalmente. La primera, consiste en transformar cada imagen en un vector de píxeles y utilizar directamente la información de las intensidades como características de entrada. Esto a veces puede ser suficiente, especialmente en imágenes pequeñas y simples. Sin embargo, cuando la complejidad de las imágenes es alta o cuando hay mucha variación en intensidad, iluminación, rotación o posición de los objetos, esto puede hacer que el clasificador no consiga aprender nada.

Para solucionar este problema existen las técnicas de extracción de características, que tratan de transformar la imagen en un vector de características que sea más adecuado para el aprendizaje del clasificador. Generalmente, estas características que se extraen de las imágenes son de más alto nivel que las intensidades y, por tanto, son más estables ante iluminaciones diferentes, posición de los objetos, etc. y sirven mejor para discriminar las distintas clases del problema. Una de las técnicas de extracción de características más habitual en la clasificación de imágenes es el histograma de gradientes orientados, o HOG por sus siglas en inglés. Este algoritmo de extracción de características divide la imagen en pequeños bloques de píxeles (8x8,

16x16, etc.) y calcula, para cada píxel del bloque, la magnitud de gradiente y su orientación. A partir de dicha información, se calcula un histograma de orientación de gradientes. Es decir, de alguna manera se identifica, para cada bloque, cuántos píxeles tienen gradiente en dirección horizontal, cuántos en vertical, etc. Así, dicho histograma nos ayuda a identificar cuál es el patrón de formas más habitual dentro del bloque. Si repetimos esta información para cada bloque de la imagen y concatenamos los histogramas (y algún proceso más que puede buscar en internet), construimos el vector de características de HOG asociado a la imagen

achieve the human detection chain.



Evidentemente, este no es el único método de extracción de características que existen en la literatura. Podéis buscar alternativas y evaluar si alguna puede ser útil para la clasificación de monedas.

Fase 1.3. Entrenamiento del clasificador

Una vez que tenemos los datos de entrenamiento preparados, es hora de elegir y entrenar modelos de clasificación. Disponéis de una colección amplia de métodos que habéis estudiado en la asignatura Aprendizaje Formal y conocéis técnicas para buscar los mejores hiperparámetros para dichos modelos. Ahora debéis ponerlos en práctica para obtener el mejor modelo.

Como hemos comentado anteriormente, disponéis de un conjunto de entrenamiento y un conjunto de test para que podáis comprobar la eficacia real de los modelos entrenados. Además, nosotros guardaremos un conjunto extra de validación, que no será público, en el que evaluaremos vuestros modelos.

Todos los datos de entrenamiento y test los tenéis disponibles en una competición privada de Kaggle que podéis encontrar en la siguiente dirección:

Rellenar url

En esta competición podréis subir también vuestros resultados en el conjunto de test, de modo que la competición registrará el resultado de cada equipo y elaborará un ranking público con los resultados.

Fase 2. Detección y clasificación de monedas

La segunda fase de este proyecto es una fase optativa que no es necesario realizar para aprobar esta práctica final. Básicamente, consiste en elaborar un sistema inteligente capaz de tomar una imagen que contiene un conjunto de monedas, y devolver el valor en euros de dicha imagen (es decir, cuánto dinero hay representado en la imagen). En las imágenes, aunque complejas, asumimos que las monedas están siempre visibles (no hay solapamiento).

Para realizar esta tarea, necesitas un método de detección de posibles monedas, previo a la aplicación del clasificador de la fase anterior. Este método de detección debe ser capaz de localizar en la imagen aquellas regiones susceptibles de contener una moneda. La forma más fácil de hacer esto sería tratando de aplicar algún método de visión artificial capaz de localizar formas circulares en la imagen. Una vez obtenida la región, mandaríamos dicha región al clasificador de monedas para que nos diga a qué moneda corresponde. Sumando así todos los valores, obtendríamos la cuantía total.

Instrucciones del trabajo

Para la entrega de la práctica final deberéis enviar en un archivo comprimido todos los códigos utilizados y estandarizados según las instrucciones que os proporcionaremos. Además, deberéis de subir a Kaggle el resultado de vuestra propuesta sobre los datos de test públicos. Por nuestra parte, utilizaremos el código para evaluar sobre un conjunto de test privado.

Además de lo anterior, deberéis realizar una presentación grabada (5-10 minutos) en la que contéis los pasos principales que habéis dado para alcanzar la mejor solución. En un documento extra, podéis también incorporar un resumen de todas las pruebas que habéis realizado para alcanzar dicho resultado.

La fecha de entrega será el 9 de enero de 2021. Se habilitará una tarea en MiAulario para su entrega.

Los grupos para la realización del trabajo estarán preferiblemente formados por dos personas. Excepcionalmente, podrán estar formados por una o tres personas, previa aprobación por nuestra parte.