Sistemas Inteligentes para la Gestión en la Empresa

Práctica 2: Deep Learning para clasificación

Objetivos y evaluación

En esta segunda práctica de la asignatura Sistemas Inteligentes para la Gestión en la Empresa estudiaremos cómo desarrollar un modelo de clasificación basado en redes neuronales profundas capaz de trabajar con datos multimodales.

La práctica consistirá en la resolución de un problema de aprendizaje utilizando técnicas de Deep Learning.

La práctica se desarrollará en grupos de dos personas (preferiblemente) o individualmente (previa consulta al profesor).

La calificación constituirá el 25% de la nota final de la asignatura (2.5 puntos).

Se evaluará la calidad de la memoria presentada y la elaboración de las soluciones propuestas. El tamaño de la **memoria** será **de máx. 5 páginas.** Se valorará especialmente la claridad en la redacción y el apoyo con materiales gráficos: diagramas, gráficas, etc. en definitiva, resultados cuantitativos.

La entrega se realizará a través de Prado, en el enlace que se habilitará al efecto en la misma página de la asignatura.

Descripción del problema

Se trabajará con el conjunto CUB-200-2011[1], que contiene más de 10.000 imágenes sobre pájaros pertenecientes a 200 categorías.

El problema de clasificación con CUB consiste en predecir la categoría de un pájaro a partir de los datos disponibles.

Respecto a los datos disponibles, tenemos dos fuentes: imágenes de los pájaros (varias imágenes de pájaros para cada categoría) y metadatos (valores numéricos de atributos para cada pájaro). En esta práctica, se deberán utilizar al menos las imágenes para la clasificación.

[1] https://www.vision.caltech.edu/datasets/cub_200_2011/

Datos

Los datos originales se encuentran disponibles en la web de los autores de CUB, aunque se recomienda utilizar la versión incluida en el fichero *pr2-starting-package.zip en la pagina de la asignatura* en Prado.

Se han creado dos datasets de imágenes {x20, x200}, que contienen las imágenes de 20 y 200 categorías, respectivamente. No hay particiones en este dataset.

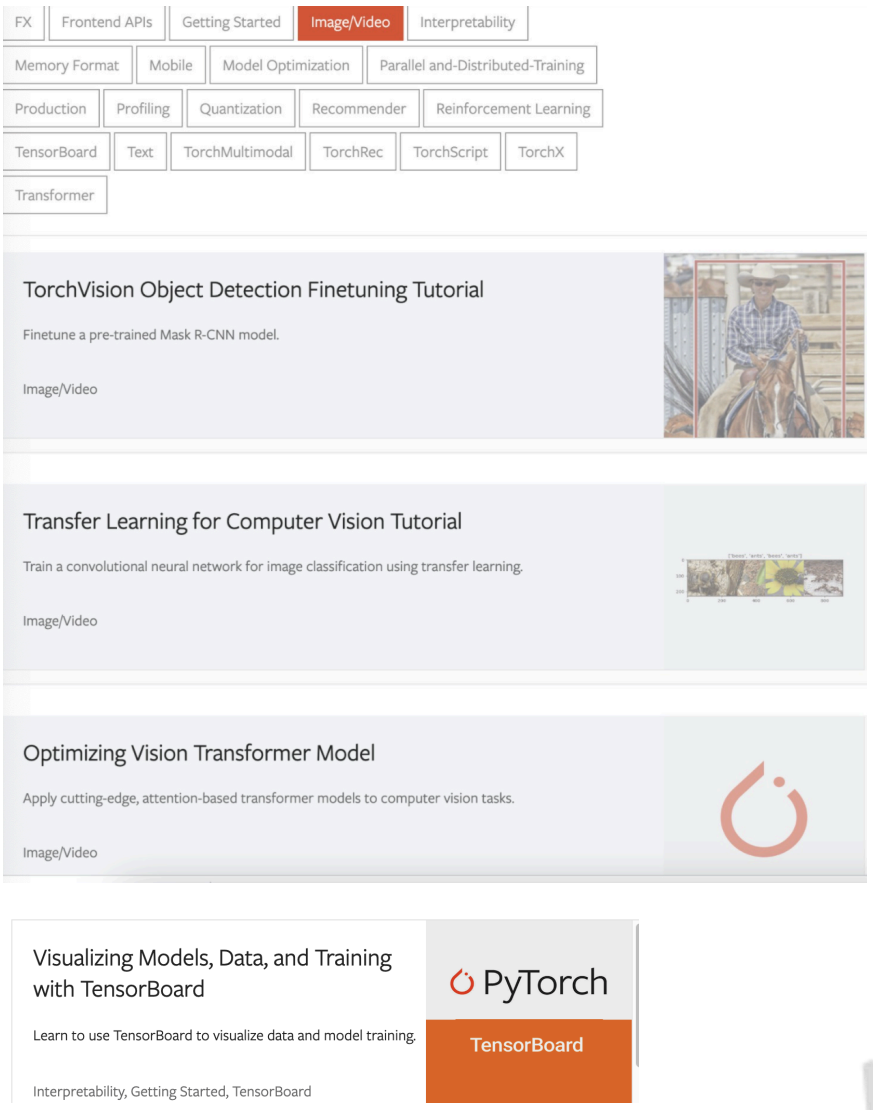
En la carpeta *data_additional* se incluyen varios ficheros de interés adicionales:

- classes.txt: lista de categorías de pájaros (200 <id_categoría, nombre_categoría>)
- images.txt: lista de imágenes (11788 <id_imagen, nombre_imagen>)
- image_class_labels.txt: listado con la categoría a la que pertenece cada imagen (11788 <id_imagen, id_categoría>)
- attributes.txt: lista de atributos descriptivas de las categorías (312 <id_atributo, nombre_atributo>)
- image_attribute_labels.txt: para cada imagen, atributo y valor binario (<id_imagen, id_atributo, valor sí/no, certeza del valor, tiempo de anotación>

Materiales

En Python: Si no se tiene experiencia previa en librerías de Deep learning (Keras, Tensorflow, etc.), se recomienda usar PyTorch. Para familiarizarse con la herramienta, se recomienda hacer los tutoriales nativos de PyTorch, por ejemplo, en PyTorch, ver:

1. DEEP LEARNING WITH PYTORCH: A 60 MINUTE BLITZ
2. Especialmente útiles son aquellos tutoriales sugeridos si seleccionamos el filtro ***Image***:



Tareas y evaluación

Se entregará la solución del problema (código Python) y una memoria descriptiva de las tareas realizadas.

La memoria explicará qué tareas para construir el modelo de clasificación y mejorar su calidad se han llevado a cabo y con qué objetivo, así como los resultados obtenidos – en términos de la calidad del clasificador y del tiempo requerido para su entrenamiento.

Las tareas que se pueden llevar a cabo y su puntuación asociada son las siguientes:

7 puntos: Resolución del problema de clasificación con imágenes

- Utilización de subconjunto de 20 categorías
- Análisis exploratorio
- Particionamiento de datos
- Clasificación multiclase
- Ajuste de hiperparámetros, topología de la red, función de coste y optimizador

- Aplicación y estudio de 2 técnicas para mejora del aprendizaje

+1 punto: Ampliación de la solución básica para las 200 clases

+1 punto: Uso de Weights And Biases (u otra herramienta de logging de modelos y parámetros como WeightsAndBiases (WANDB) o para gestión del proceso de aprendizaje y busqueda de mejores parámetros - parameter finetuning)

+1 punto: Por cada una de las siguientes ampliaciones:

- Aprendizaje multimodal (imágenes + datos tabulares + descripciones en lenguaje natural u otros, e.g. ver inspiración de https://imagebind.metademolab.com/)
- Aprendizaje con coste variable (por ejemplo, modificación de la función de pérdida[2])
- Ajuste automático de hiperparámetros (usando, por ej., optuna, hyperopt, otros)
- Métodos de tipo *ensemble* e híbridos
- Transferencia de aprendizaje o *fine tuning*
- Explicabilidad del modelo (por ejemplo, se puede utilizar un modelo explicable o composicional con la metodología X-NeSyL[3], que además use como entrada los atributos del dataset -ala roja, pico amarillo-)

Se podrá obtener más de 10 puntos en esta práctica, hasta un máximo de 13 puntos. El exceso sobre 10 se añadirá a la nota global de la asignatura, pudiendo sumar hasta 3.25 en esta práctica.

Entrega

Límite: Ver fecha límite de entrega en Prado.

Contenidos: Un fichero .zip, incluyendo:

- Código (.py, .ipynb, otros. Adicionalmente y obligatoriamente, una url a un Google Colab ejecutable y ejecutado y una copia del mismo impresa en pdf)
- Memoria
  - Portada: nombre de los autores, título
  - Índice
  - Contenidos
    - Fundamentos teóricos
    - Descripción de las técnicas empleadas
    - Discusión de resultados
    - Conclusiones
  - Bibliografía

[2] Por ejemplo, regularización para hacer "cost-sensitive learning / class weighting", como se ha visto en clasificación no balanceada y en multi-clase. Ver ejemplo a modo de tutorial: https://tensorflow.rstudio.com/guides/keras/training_with_built_in_methods#using-sample-weighting-and-class-weighting

[3] Díaz-Rodríguez, N., Lamas, A., Sanchez, J., Franchi, G., Donadello, I., Tabik, S., ... & Herrera, F. (2022). EXplainable Neural-Symbolic Learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. *Information Fusion*, *79*, 58-83. https://www.sciencedirect.com/science/article/pii/S1566253521001986 o https://arxiv.org/abs/2104.11914