Deep Fake Detector

Por Pablo Guilló y Javier Franco

Introducción

A lo largo de este proyecto hemos desarrollado un modelo de detección para imágenes de rostros falsas. Desarrollado en python bajo el framework de deep learning pytorch, este modelo debería ser capaz de diferenciar rostros falsos de reales. Se han desarrollado dos versiones distintas del modelo, una completamente independiente y desde cero y otra a través de transfer learning y el modelo resnet18. Todo esto desarrollado y ejecutado valiéndose de jupyter notebook y las máquinas y GPUs de Google Collab.

Dataset

El dataset a emplear se encuentra disponible en Kaggle y se trata de un conjunto de imágenes de rostros, tanto verdaderos como falsos, que se emplean para entrenar el modelo. Las imágenes están divididas por dificultad existiendo tres tipos, fáciles, medias y difíciles. Además, se decidió ampliar el dataset mediante data augmentation, es decir, generar nuevas imágenes a través de las originales para aumentar el tamaño de los datos de entrenamiento y en consiguiente la información a emplear.

Método optimizador y función de pérdida

En lo que a los métodos de optimización se refiere, se empleó la variante ADAM en ambos modelos. Además, al modelo con fine tunning se le añadió la variante ADAMW como una forma de regularización. Respecto a la función de coste, se usó la entropía cruzada al tratarse de un problema de clasificación binaria estándar.

Modelo FaceAuthenticityCNN

El modelo FaceAuthenticyCNN fue desarrollado directamente desde cero y no posee capas provenientes de otras estructuras. Consiste en dos capas convolucionales con 32 y 64 salidas, kernels 3x3 y un pixel de padding, 2 capas de maxpooling (2,2) y dos capas fully connected que desembocan en 2 salidas. Posee un rendimiento bastante pobre, probablemente debido al estructura básica y la falta de datos de calidad, incluso después del augmentation.

Modelo FineTunning

El modelo con fine tunning emplea las capas preentrendas del modelo resnet 18. Una particularidad de este modelo es que hubo que cambiar las dimensiones de las imágenes a (256,256). Se reentrenaron las capas fc y se obtuvo un rendimiento superior al modelo básico 70% vs 60%

Conclusión

En conclusión, el proyecto de detección de rostros falsos ha sido un proceso de exploración y experimentación utilizando dos enfoques distintos: un modelo desarrollado desde cero (FaceAuthenticityCNN) y otro aprovechando la técnica de fine-tuning con un modelo preentrenado (FineTuning). A pesar de que el modelo inicial mostró limitaciones en términos de rendimiento, el enfoque de transferencia de aprendizaje demostró ser más efectivo, logrando mejoras significativas en la precisión.