

Universidad de Antioquia

Maestría en ingeniería



Fundamentos de Deep Learning

Entrega 1 Proyecto del Curso

Estudiantes:

Ricardo Tangarife González 1045025315

Fabian Stiven Duque Duque 1026149866

Docentes:

Raul Ramos Pollan

Julian David Arias londoño

Semestre 2022/1

Contexto de Aplicación.

Nuestros trabajos de investigación de maestría están enmarcados dentro de un proyecto que busca dar apoyo al ámbito de la seguridad ciudadana, específicamente a través de señales acústicas y visuales.

Por tanto, a partir de modelos se desea detectar eventos relacionados a la seguridad ciudadana, mediante el sensado de sonido y video.

Las zonas urbanas de las ciudades emiten diferentes tipos de información que comunican eventos de alerta, riesgo, o sucesos anómalos. Se pretende poder recopilar algunos de estos eventos y que sean reconocidos de forma autónoma, que permita generar alertas a partir de dicha situación.

Con el contexto anterior, planteamos como proyecto para el curso generar dos modelos, uno para la clasificación de dos eventos de sonido y un modelo de detección de peleas a través de videos, que apoyen el ámbito de la seguridad ciudadana.

Nota: Dependiendo de la calidad de los modelos y tiempos de desarrollo de los mismos, se plantea también poder probar la construcción de un modelo fusionado a partir de los dos modelos generados previamente, probando algunas de las técnicas para fusión de modelos; tal que permita el reconocimiento de los eventos de video y sonidos con un modelo unificado.

Objetivo de Machine Learning.

Objetivo 1: Clasificar los eventos sonoros de disparos y sirenas correspondientes a eventos que apoyan el ámbito de la seguridad ciudadana.

Objetivo 2: Identificar eventos de peleas a través de videos que generen alertas como apoyo a la seguridad ciudadana.

Datasets.

Sonido:

Para el entrenamiento del modelo de predicción acústico se seleccionarán dos sonidos del dataset UrbanSound8k.

El dataset UrbanSound8k es un dataset público con 8730 sonidos etiquetados, utilizado en diferentes tareas abiertas en kaggle, su tamaño en disco es aproximadamente 7GB y los archivos de sonido están en formato .wav; la distribución de clases es la siguiente:

- 0 = air_conditioner = 873
- 1 = car_horn = 873
- 2 = children_playing = 873
- 3 = dog_bark = 873

4 = drilling = 873
5 = engine_idling = 873
6 = gun_shot = 873
7 = jackhammer = 873
8 = siren = 873
9 = street_music = 873

Video:

UBI-Fights es un conjunto de datos a gran escala de 80 horas de video completamente anotado a nivel de cuadro. Consta de 1000 videos, donde 216 videos contienen un evento de lucha y 784 son situaciones normales de la vida diaria, la dimensión 640 x 360 píxeles y la velocidad de fotogramas es de 30 fps y el tamaño en disco es cercano a 7.6 GB.

El conjunto de datos cuenta con las siguientes anotaciones:

- Fight: F_id_environment_camera_color.mp4
- Normal: N_id_environment_camera_color.mp4
- Environment: Indoor (0) / Outdoor (1);
- Camera: Fixed (0) / Rotated (1) / Movable (2);
- Color: RGB (0) / Grayscale (1);

Métricas de desempeño.

Para evaluar los resultados de los modelos, se utilizará métricas como:

- Matriz de confusión
- La exactitud (Accuracy) de predicción
- La puntuación F1 (F1-Score)

Referencias y Resultados Previos.

UrbanSound8k:

En Kaggle, con algunos resultados previos:

<https://www.kaggle.com/datasets/chrisfilo/urbansound8k>

Acceso Dataset:

<https://urbansounddataset.weebly.com/urbansound8k.html>

UBI-Fights:

Acceso Dataset:

<http://socia-lab.di.ubi.pt/EventDetection>

Resultados previos:

https://www.di.ubi.pt/~hugomcp/doc/Events_PRL.pdf

Métricas de desempeño:

https://fayrix.com/machine-learning-metrics_es