

Detección de somnolencia utilizando técnicas de visión artificial en entornos móviles.

Autores: Macarena Quiroga, Emilio Melo

Director: Martín Bilbao

Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco,
Comodoro Rivadavia, Chubut. {meloemilio.prog,msinf.quiroga}@gmail.com,
martinbilbao@ing.unp.edu.ar

Resumen Este trabajo tiene como objetivo presentar el desarrollo e implementación de un dispositivo inteligente de detección de somnolencia en conductores empleando principalmente técnicas de visión artificial y una Raspberry Pi 4, con el fin de alertar al conductor y prevenir la ocurrencia de potenciales accidentes de tránsito. En primera instancia, se realizará una introducción a la temática abordada, junto los conceptos claves necesarios para el desarrollo del trabajo. Posteriormente, se hará mención sobre el desarrollo de los datasets, las comparativas y estudios de escalabilidad usando los modelos preentrenados MobileNetV2 e InceptionV3, el proceso de detección de somnolencia empleando como técnicas de visión artificial los modelos de CNN mencionados y los clasificadores basados en cascadas de Haar, finalizando con las pruebas en la Raspberry Pi en un entorno real de conducción.

Keywords: Detección de somnolencia · Raspberry Pi · Aprendizaje por transferencia · Vision artificial.

Contexto

El trabajo forma parte de la tesina de grado realizada para la obtención del título "Licenciatura en Informática", presentada en Agosto de 2022 en la Universidad Nacional de la Patagonia San Juan Bosco, sede Comodoro Rivadavia. El objetivo principal ha sido desarrollar e implementar un dispositivo inteligente que permita detectar, a través de una cámara y en tiempo real, si un conductor presenta indicios constantes de somnolencia. La motivación para su desarrollo fueron los altos índices de mortalidad en Argentina y en el mundo ocasionados por los accidentes de tránsito, siendo la somnolencia una de sus causantes.

1. Introducción

La somnolencia puede ser definida como la necesidad de conciliar el sueño o una tendencia a quedarse dormido. Es un proceso, resultado del ritmo biológico humano normal, el cual consiste en ciclos que involucran contraer sueño y estar despierto[1].

Las Redes Neuronales Convolucionales (CNN) y el empleo de técnicas de visión artificial, permiten retratar el comportamiento del ojo humano, posibilitando así el reconocimiento de objetos a través de la captura de imágenes, con el objetivo de realizar algún tipo de inferencia y actuar en consecuencia.

Las técnicas de visión artificial están logrando avances significativos en diversos ámbitos siendo ampliamente utilizadas en diferentes tipos de aplicaciones, ofreciendo grandes ventajas que conllevan a generar un impacto favorable en la sociedad, contrarrestando las problemáticas existentes en la actualidad. Una de ellas, es la alta tasa de accidentes viales, causada en gran medida por conductores somnolientos.

El 30% de los accidentes graves están relacionados a la somnolencia y/o fatiga. Al conducir cansado o con sueño, por consumo de medicamentos contra-productivos, ingesta de bebidas alcohólicas, contar con trastornos del sueño, la conducción nocturna, entre otros factores, hacen que el riesgo de un accidente aumente[2]. Es por ello, que el desarrollo de un dispositivo inteligente portable para detectar la somnolencia, permitiría disminuir la probabilidad de ocurrencia de accidentes viales, reduciendo por ende, las pérdidas humanas y/o lesiones.

2. Desarrollo

La implementación del dispositivo inteligente involucró una serie de desarrollos, los cuales han sido la elaboración de datasets de imágenes de ojos abiertos y cerrados de rostros de personas, desarrollo de modelos basados en la arquitectura MobilenetV2 e InceptionV3, empleando la técnica transfer learning (o aprendizaje por transferencia); el ensamble del prototipo utilizando como componente principal la Raspberry Pi 4 modelo B y un script para realizar el proceso de detección de somnolencia.

2.1. DATASETS

Los datasets empleados para el entrenamiento de los modelos preentrenados seleccionados, han sido elaborados utilizando capturas de ambos ojos, pertenecientes a rostros de 5 (cinco) personas. Para realizar dichas capturas, se desarrolló un script en Python el cual utiliza la librería OpenCV para la interacción con la cámara OV5647 de la Raspberry Pi, junto con los clasificadores basados en cascadas de Haar *haarcascade_righteye_2splits.xml* y *haarcascade_lefteye_2splits.xml*, obtenidos del repositorio oficial de OpenCV en GitHub, para la detección de ambos ojos. En el cuadro 1 se refleja la composición de los dos datasets elaborados.

Dataset	Cantidad de imágenes						Total
	Entrenamiento		Validación		Prueba		
	Abiertos	Cerrados	Abiertos	Cerrados	Abiertos	Cerrados	
1	500	500	200	200	14	14	1428
2	200	200	50	50	14	14	528

Cuadro 1: Composición de los dataset elaborados.

2.2. TRANSFER LEARNING

Aprendizaje por transferencia es una técnica que permite transferir conocimiento adquirido utilizando modelos preentrenados que poseen un gran conjunto de datos, para que puedan ser personalizados en una tarea determinada. Esto permite hacer uso de arquitecturas que están altamente probadas, permitiendo desarrollar modelos de manera rápida y eficiente, brindando soluciones a problemas complejos como son los de visión artificial. Las arquitecturas basadas en CNN para la detección de objetos que han sido utilizadas en el trabajo fueron MobilenetV2 e InceptionV3.

MobileNet es una arquitectura propuesta por Google. Fue diseñada para obtener una máxima eficiencia en la precisión, teniendo en cuenta los recursos limitados, para una aplicación de visión artificial, en los dispositivos móviles e integrados. Pueden utilizarse para detección, clasificación, incrustación y segmentación. Por otro lado, InceptionV3 es un modelo muy utilizado para el reconocimiento de imágenes. El mismo fue entrenado con el conjunto de datos de ImageNet, y ha demostrado lograr una exactitud mayor al 78.1% [4].

Para cada modelo entrenado utilizando aprendizaje por transferencia, se realizó su correspondiente estudio de escalabilidad. La escalabilidad, en términos de un algoritmo de deep learning, hace referencia a la capacidad del mismo de adaptarse, manteniendo o mejorando su eficiencia cuando se modifica el tamaño o el volumen de entrada de los datos.

La metodología empleada para realizar el estudio de escalabilidad consistió en evaluar los modelos propuestos sobre los dos datasets presentados en la sección 5.1, aplicando tamaños de imágenes de entrada de 165x165 píxeles y 96x96 píxeles. Para cada tamaño, se realizó el entrenamiento partiendo de 4 épocas, aumentando este número hasta alcanzar una precisión igual o superior al 90%, o hasta ver una decadencia de ésta u otras métricas que indiquen que el modelo no está mejorando. Finalmente, los dos modelos seleccionados, basados tanto en MobileNetV2 como en InceptionV3, fueron aquellos que obtuvieron un 100% en exactitud, precisión, recall y F1 en las pruebas previas.

2.3. DISPOSITIVO INTELIGENTE

La construcción del prototipo partió de la utilización de una Raspberry Pi 4 con 4GB de memoria RAM, a la cual le fueron conectados una serie de dispositivos.

Uno de ellos fue una cámara modelo OV5647 para realizar las correspondientes capturas del rostro del conductor, permitiendo además su utilización durante la noche gracias a los módulos infrarrojos incorporados. Adicionalmente, también se hizo uso de un buzzer activo de 5V y 90 decibeles para la alerta sonora, y un cooler de 5V para la refrigeración de la Raspberry. Por otra parte, se utilizó una memoria de clase 10 de 32 GB para el sistema operativo Raspberry Pi OS, el sistema operativo oficial.

Los modelos desarrollados fueron convertidos a un formato apto para poder ser ejecutados en la Raspberry Pi. El proceso de detección de somnolencia se llevó a cabo mediante un script desarrollado con el lenguaje de programación Python, utilizando los clasificadores basados en cascadas de Haar para detectar las regiones de interés (ojos), de las cuales se realizan capturas que luego son enviadas al modelo a fin de que el mismo prediga el estado de los ojos (abiertos o cerrados). La alarma de detección de somnolencia, entonces, es emitida si los ojos del conductor se detectan como cerrados durante 3 o más cuadros, emitiendo así la alerta sonora e incrementando un contador de somnolencia, el cual indica la cantidad de veces que la alarma fue disparada. En la figura 1 se puede observar el proceso de detección de somnolencia.

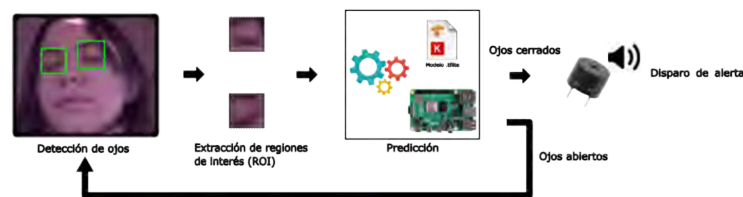


Figura 1: Esquema simplificado del proceso de detección de somnolencia.

3. Resultados

Se realizaron pruebas con ambos modelos durante el día y la noche, en el cual el modelo MobileNetV2 presentó un mejor desempeño en comparación a los tiempos de retardo en la detección de ojos y la emisión de alerta sonora de InceptionV3. Los tiempos correspondientes al retardo de captura y emisión de alerta utilizando MobileNetV2 fueron de 2.45 y 3.50 respectivamente, mientras que en caso de InceptionV3 fueron de 3.28 y 4.49 segundos. En vista de los tiempos antes mencionados fue que se decidió utilizar como modelo final en el dispositivo aquel basado en MobileNetV2, el cual ofrece una diferencia de 830 milisegundos con respecto al retardo de captura de la detección y 1 segundo en la emisión de la alerta.

En la figura 2 se puede observar la detección de somnolencia durante la noche, y en la figura 3 se visualiza la detección de ojos abiertos durante el día con la presencia de un falso positivo, detectando la boca como un ojo.



Figura 2: Detección de somnolencia durante la noche.



Figura 3: Detección de ojos abiertos durante el día.

Los scripts, datasets y modelos entrenados se encuentran disponibles para su acceso, descarga e instalación en el repositorio de GitHub [5].

4. Conclusiones y trabajos futuros

Se ha logrado desarrollar un dispositivo inteligente capaz de detectar indicios de somnolencia y alertar mediante una alarma sonora al conductor, haciendo uso de clasificadores Haar y un modelo entrenado basado en MobileNetV2 con una precisión del 100 %, demostrando así el beneficio de utilizar las técnicas de aprendizaje por transferencia y visión artificial en aplicaciones de tiempo real.

El dispositivo desarrollado, además de brindar una efectividad satisfactoria, sienta un precedente para mejoras y funcionalidades futuras, como incorporar nuevas técnicas y librerías que permitan ampliar los tipos de indicios de somnolencia detectados por el dispositivo, tales como el bostezo y cabeceo, permitir la detección de los ojos abiertos y cerrados con cascadas de Haar cuando el conductor utiliza anteojos, entrenando nuevos detectores que permitan alcanzar este objetivo y añadir un sistema de autenticación, ya sea por reconocimiento facial o tarjeta RFID, de tal forma que el conductor pueda registrarse al ingresar al vehículo y obtener así datos estadísticos por usuario.

Referencias

1. Aleksandar Colic, Design and implementation of driver drowsiness detection system, 2014.
2. Organización civil Luchemos por la vida, <http://luchemos.org.ar>. Ultimo acceso:10-06-22.
3. Mark Sandler and Andrew Howard and Menglong Zhu and Andrey Zhmoginov and Liang-Chieh Chen, MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2019.
4. Google Cloud, Advanced Guide to Inception v3, <https://cloud.google.com/tpu/docs/inception-v3-advanced>. Ultimo acceso: 10-03-22.
5. Repositorio Drowsiness Pi, <https://github.com/DevTeamCR/DrowsinessPi>.