



Instituto Tecnológico y de Estudios Superiores de Monterrey

**Momento de Retroalimentación: Reto Privacidad y Seguridad de los Datos (Portafolio
Análisis).**

Integrantes:

Diego Alberto Baños Lopez	A01275100
Carlos David Lozano Sanguino	A01275316
Carol Arrieta Moreno	A01275465

Monterrey, Nuevo León, México 27 de Octubre del 2023

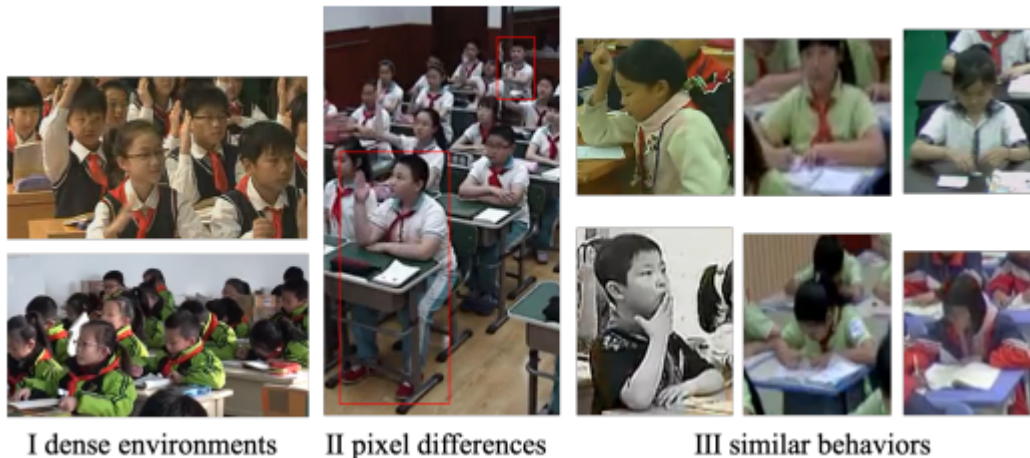
Inteligencia artificial avanzada para la ciencia de datos II (Gpo 501)

En este proyecto es importante además de los datos que recolectemos, garantizar la anonimidad de dichos datos o más bien garantizar que videos, imágenes u otro recurso a utilizar no sea comprometedor para las personas que participen o se usen en el producto, en algún entrenamiento que tengamos planeado utilizar u algún otro uso que le demos a dichos datos.

Para el reto se partió la idea de hacer dos distintos modelos con un objetivo diferente cada uno, el primero sería el reconocimiento facial que detectara el rostro de los estudiantes mientras que el segundo modelo detectara cuando algún estudiante levante el brazo para participar.

Para el modelo de reconocimiento facial se planea usar la librería de python `face_recognition`, el cual es un modelo que en general ayuda a identificar las caras de los hipotéticos estudiantes, este modelo presume de usar de información de dominio público y usar componentes open source, que siempre y cuando uno haga referencia a que se usó este modelo, no debería de haber ningún problema con la normativa que sigue el socio en el trato de la privacidad, el cual se basa en el respeto de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares y las ley ARCO que protegen a las personas en México en que se exponga datos vulnerables privados y se comprometa el derecho individual a la privacidad, de igual manera en el caso de los rostros de prueba se procurará cumplir con dicho trato a través de un acceso limitado y usando rostros públicos o que sean personas que estén conscientes de lo que se trata y de cómo serán utilizados, el almacenamiento se procura que sea controlado únicamente por los miembros del equipo responsables de desarrollar el proyecto.

Por otro lado el modelo para el reconocimiento de brazos o levantamiento de brazos para la participación está planeado que sea basado en el modelo Yolov8 por Ultralytics para la detección de objetos, a diferencia del anterior modelo, este se entrenará bajo un dataset de estudiantes levantando la mano, tras haber investigado diferentes fuentes, papers e investigaciones, encontramos un dataset de acceso público denominado como Student Class Behaviour Datasets (SCB Dataset), este dataset es un compilacion de videos/fotos de estudiantes que levantan la mano, escriben o leen que recopilaron ellos mismos para su trabajo en detección del comportamiento de los estudiante en clase con Yolov5 y Yolov7 el cual compartieron mediante su paper y github para acceso libre a cualquier futura investigación teniendo en cuenta la falta de datasets públicos para los proyectos con Inteligencia Artificial que radican actualmente, este datasets los actualizan constantemente para sus posteriores investigaciones y proyectos para la detección de participación y comportamiento en clase.



Los autores de estos papers y creación del datasets son YANG FAN y WANG TAO de la universidad Chengdu Neusoft en China donde recopilaron estas imágenes de acceso libre con la seguridad de no afectar la confidencialidad de los usuarios que residen en las imágenes ya que se han usado para sus posteriores trabajos por la institución mencionada anteriormente.

SCB-dataset [↗](#)

Student Classroom Behavior dataset

Paper Name	Paper Link	Dataset Link	Trained weights Link
SCB-dataset 1 SCB-dataset: A Dataset for Detecting Student Classroom Behavior	4k hand-raising paper	Baidu Netdisk dataset extraction code: ZDBG	Baidu Netdisk models extraction code: aq7v

Para la obtención del dataset se descargó de la nube Baidu Netdisk que es una aplicación similar a Google Drive para el almacenamiento en la nube donde se encuentra toda la info necesaria para usar el dataset, para mayor seguridad del uso de los datos nuestro equipo envió un correo de autorización para permitir el uso del dataset en nuestro proyecto y así tener acceso libre a estos datos con mayor seguridad.



Yang.Fan <Yang.Fan@nsu.edu.cn>

Para: Carlos David Lozano Sanguino



Mié 25/10/2023 18:58

Iniciar respuesta con:

Of course, you are free to use it, I have made SCB-Dataset public, but please add the source or paper references when using it.

从 Windows 版邮件发送

发件人: Carlos David Lozano Sanguino

发送时间: 2023年10月26日 4:44

收件人: Yang.Fan@nsu.edu.cn

抄送: Diego Alberto Baños Lopez

主题: HELP WITH Github: SCB-Dataset

Hello! my name is Carlos, the reason for my mail is due I saw your GitHub with the work with yolo and the SCB-Dataset, is awesome by the way, I wanted to ask you if there is a data license or policy, permission to use that dataset, I wanted to know or consult it because I would like to use that dataset for a participation recognition project by raising my hand but they ask me for a license or terms of use for data protection Could you help me with this question please?

Here is my github account if you want it or need it 😊 : <https://github.com/CDLS12>

Debido a que este dataset es de acceso libre para cualquier usuario y es creado por Yan y Wang Tao no nos corresponde la forma de trabajo de los datos, mas que el contacto directo con los autores por correo o github como se vio anteriormente para tener autorizacion del uso de los datos teniendo como condicion que su trabajo sera citado en cualquier investigacion externa como la nuestra para evitar o violar normas del derecho de autor, estos datos sera almacenados en nuestros dispositivos o drive para el entrenamiento del modelo y se mencionara el trabajo de Yan y Wang Tao que sirvio de referencia para nuestro actual reto, no es necesario la firma de algun documento dado que es de acceso abierto a diferencia de la mayoría de datasets que actualmente son privados o necesitan permiso o firma para el uso de estos.

Referencias:

-Fan, Y. (s/f). SCB-Dataset3: A benchmark for detecting student classroom behavior. Arxiv.org. Recuperado el 27 de octubre de 2023, de <http://arxiv.org/abs/2310.02522>

-Fan, Y. (s/f-a). SCB-dataset: A Dataset for Detecting Student Classroom Behavior. Arxiv.org. Recuperado el 27 de octubre de 2023, de <http://arxiv.org/abs/2304.02488>

-Yang, F. (s/f). SCB-dataset: Student Classroom Behavior dataset. de <https://github.com/Whiffe/SCB-dataset>

Ageitgey, A. (2021). face_recognition: The world's simplest facial recognition api for Python and the command line [Código fuente]. https://github.com/ageitgey/face_recognition/tree/master

