Universidad Tecnológica Nacional
AJEA – Actas de Jornadas y Eventos Académicos de UTN
Trabajo Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1c38
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Colección JIT en AJEA https://rtyc.utn.edu.ar/index.php/ajea/JITs

Entrenamiento de Modelo Clasificador por Imágenes para Museografía Interactiva.

Image Classifier Model Training for Interactive Museography.

Presentación: 11/09/2023

Francisco Levrino
Universidad Tecnológica Nacional (UTN) - Facultad Regional San Francisco.
flevrino@facultad.sanfrancisco.utn.edu.ar

Lucía V. Castellano
Universidad Tecnológica Nacional (UTN) - Facultad Regional San Francisco.
lcastellano@facultad.sanfrancisco.utn.edu.ar

Rocío Boriglio
Universidad Tecnológica Nacional (UTN) - Facultad Regional San Francisco.
rboriglio@facultad.sanfrancisco.utn.edu.ar

Resumen

El presente trabajo expone las primeras etapas en el entrenamiento de un clasificador con el uso de la red neuronal YOLO. Se propone emplear esta herramienta en el Museo Interactivo de Ciencias (MuIC) para fomentar la museografía interactiva, implementando el reconocimiento por imágenes con el uso de dispositivos móviles. El objetivo es la inserción de este mecanismo de identificación para brindar acceso a guías audiovisuales a los visitantes de las muestras del MuIC. Con esta propuesta, los visitantes podrán gozar de una experiencia independiente, guiada por su curiosidad ante los juegos propuestos y otorgará limpieza visual en los mismos al eliminar el uso de un código *QR*.

Palabras clave: Reconocimiento por imágenes, Red neural, Museografía interactiva, Experiencia de usuario, guías interactivas.

Abstract

This work presents the first steps on the classifier training using a neuronal network “YOLO”. The proposal is to employ this tool at the Museo Interactivo de Ciencias (MuIC) to promote interactive museography, introducing image recognition with mobile devices. The aim is to integrate this identification mechanism to provide audiovisual guides to MuiC visitors. In this way, the users will be able to enjoy an independent experience, guided by their curiosity on the proposal games, and give a visual cleaning removing the QR codes.

Keywords: Image recognition, Neuronal network, Interactive museography, User experience, interactive guides.

AJEA – Actas de Jornadas y Eventos Académicos de UTN
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Texto Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1638

Introducción

El Museo Interactivo de Ciencias (MuIC) es un grupo de investigación que pertenece al departamento de materias básicas de la Universidad Tecnológica Nacional (UTN) Facultad Regional San Francisco, que cuenta con una muestra interactiva denominada “ConCiencia”. El mismo está constituido por docentes, estudiantes y graduados de las distintas carreras de ingenierías con las que cuenta la Facultad. Este equipo se encarga de diseñar cada una de las experiencias interactivas y de brindar acompañamiento a los visitantes en durante las muestras.

El MuIC está constantemente explorando nuevos métodos de aplicación de conocimientos por lo que, es necesario conocer los distintos medios en que la tecnología puede hacerse presente para colaborar en la enseñanza de las diferentes temáticas de la ingeniería y contribuir al aprendizaje de manera efectiva.

El grupo lleva a cabo sus actividades mediante una muestra permanente que recibe distintas instituciones tanto locales como regionales. Los visitantes abarcan desde estudiantes de nivel inicial, secundario y terciario hasta el público en general de la localidad de San Francisco y sus alrededores.

Dado que los objetivos del grupo requieren adaptarse a la tecnología y a las nuevas aplicaciones de la misma, es fundamental destacar que, a través del proyecto PID SIPPSF0009830, se está desarrollando una herramienta que permite la clasificación de tres de los juegos exhibidos en la muestra y brinda la posibilidad de la detección e identificación de los mismos a través de dispositivos móviles.

Actualmente el MuIC cuenta con guías en formato papel que permiten a los visitantes llevar a cabo su experiencia con cierto grado de libertad, sin requerir ayuda de los miembros del museo para poder interactuar con las actividades. Además de estas guías impresas, un grupo dentro del museo está desarrollando guías en formato video que constituyen una alternativa más amigable a la escrita en formato papel, las cuales debido a su adaptabilidad y fácil accesibilidad permiten alcanzar una óptima comprensión de los temas a tratar independientemente de la clase de audiencia a la que el mensaje es dirigido (Gilli et al., 2023).

Los visitantes pueden acceder a estas guías a través de sus propios dispositivos móviles o utilizando dispositivos proporcionados por el museo, y las primeras pruebas se realizan mediante códigos QR (Levrino et al., 2023). A través de estas nuevas guías, se podrá brindar una mejor y más independiente experiencia a lo largo de la muestra, en especial en visitas de grandes grupos donde se dificulta el acceso simultáneo de todos a las guías en formato papel.

Para lograr una mayor limpieza visual de la muestra y así mejorar la experiencia del usuario, se están evaluando las opciones de reconocimiento de imágenes con la utilización de técnicas de visión artificial, convenientes para sentar bases en el desarrollo de una herramienta que permita el acceso a las guías sin la necesidad del uso de códigos QR. En este trabajo, se detalla la metodología utilizada para la selección de las experiencias, la preparación del dataset y el proceso de entrenamiento. Posteriormente, se presenta un análisis detallado de los resultados obtenidos y, finalmente, se resumen las conclusiones alcanzadas.

Metodología

Este trabajo presenta el empleo de la red neuronal YOLO (Bochkovskiy et at., 2020) en su versión V4 utilizando como *backbone* Darknet53 (Chien-Yao et al., 2020). Esta red permite de manera sencilla y rápida el entrenamiento de un modelo basado en un *dataset* personalizado y clasificado mediante la aplicación “labelImg” (Tzutalin, 2015). Estas herramientas fueron seleccionadas debido a su simplicidad, facilidad de uso y familiarización de miembros participantes del museo con las mismas.

Selección de juegos

AJEA – Actas de Jornadas y Eventos Académicos de UTN
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Texto Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1638

En base a la muestra con la que cuenta el grupo de investigación MuIC, especialmente las experiencias que se hallan en exposición permanente y las cuales son las más utilizadas en las presentaciones del grupo, se seleccionaron tres para poder comenzar las pruebas de entrenamiento del modelo clasificador. Las mismas cuentan con guías audiovisuales por lo que, en el caso de éxito en las pruebas, son idóneas para continuar con el prototipo de una herramienta que ayude al usuario a acceder a las guías. En la Figura 1 se encuentran los juegos seleccionados para el desarrollo del entrenamiento del modelo clasificador.

Figura 1. Experiencias interactivas “Cosa de Parejas”, “Jugando con el submarino” y “Tangram”.

Preparación del *dataset*

El *dataset* en este caso es un conjunto de imágenes que contienen los elementos a identificar. Además, cada imagen está acompañada por un archivo de texto que identifica el objeto y su posición en la imagen.

Para la confección de este *dataset*, se inicia con clips de videos cortos, cada uno con una duración aproximada de 10 segundos para cada una de las tres actividades seleccionadas en esta instancia de pruebas. Estos videos fueron grabados en distintos entornos, con los objetos ubicados en diversas superficies y, en algunos casos, con varios objetos compartiendo el mismo encuadre. La diversidad de escenarios y las disposiciones contribuyen a brindar variedad en las imágenes y contribuye a mejorar los resultados del proceso de entrenamiento.

Una vez capturados los videos, se separaron en imágenes mediante una herramienta realizada en Python por miembros del museo. Esta herramienta extrae un fotograma del video cada cinco fotogramas y lo guarda como una imagen. De este modo, se obtuvieron un total de 758 imágenes.

El siguiente paso consistió en el etiquetado de las imágenes. Para ello, el *dataset* fue dividido en cuatro grupos de imágenes, distribuidas entre los autores y colaboradores del presente artículo, así se pudo terminar de manera más rápida esta tarea que, si bien no es compleja, es bastante tediosa y repetitiva. Luego de esto, el *dataset* fue rearmado obteniendo así un total de 408 etiquetas para “Tangram”, 320 etiquetas para “Cosas de pareja” y 336 para “Submarino”.

Entrenamiento

Luego de la preparación del *dataset*, el mismo fue enviado a uno de los servidores proporcionados por una empresa especializada en soluciones de visión artificial, que permitió utilizar tiempo de servidor para llevar a cabo los entrenamientos. Este servidor cuenta con una placa de video Nvidia A-100 la cual constituye el tope de gama de los productos Nvidia orientados al entrenamiento de modelos de visión artificial.

Durante el proceso de entrenamiento, se necesitan dos grupos de datos: el conjunto de datos de entrenamiento (*train)* y el conjunto de datos de evaluación (*test)*, los cuales cumplen funciones íntegramente diferentes y de vital importancia. El entrenamiento se desarrolla en iteraciones, donde se suministran los datos del grupo *train* al modelo y se evalúa su respuesta para que pueda aprender de las etiquetas adjuntas a cada imagen. Una vez terminada esta etapa de aprendizaje, el modelo es

AJEA – Actas de Jornadas y Eventos Académicos de UTN
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Texto Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1638

alimentado con datos del grupo *test* y se compararan los resultados que se hayan generado, con las etiquetas realizadas por las personas. Esto permite evaluar la precisión del modelo en la detección.

El entrenamiento fue realizado con un máximo de 6000 iteraciones. Aproximadamente en la iteración 900, se comienza a observar un mAP% *(mean Average Precision)* superior al 50% y una disminución en la función de pérdidas, que alcanzó valores por debajo de 3. Además, en la Figura 2 se aprecia que a partir de la iteración 2000, el mAP% comienza a estabilizarse en un rango de valores entre 95% y 99%. Por otro lado, es notable que la función de pérdida, a partir de la iteración 4000 permanece casi constante en un intervalo de valores de 0,5 a 0,7, por lo que se estima que aumentar la cantidad de iteración puede no mejorar el modelo o incluso empeorar los resultados obtenidos. Resulta interesante llevar a cabo un entrenamiento más prolongado para evaluar si los resultados se mantienen, mejoran o empeoran.

Resultados

Luego de completar la etapa de entrenamiento, se obtuvo un modelo que demostró una gran capacidad para detectar las tres experiencias contando con una buena precisión, teniendo en consideración la cantidad de imágenes y las configuraciones que se utilizaron. A continuación, en la Figura 2 se observa la progresión de mejora en la precisión del modelo a medida que aumentan las iteraciones o pasadas (línea roja), al mismo tiempo que se aprecia una disminución en la función de pérdidas (línea azul).

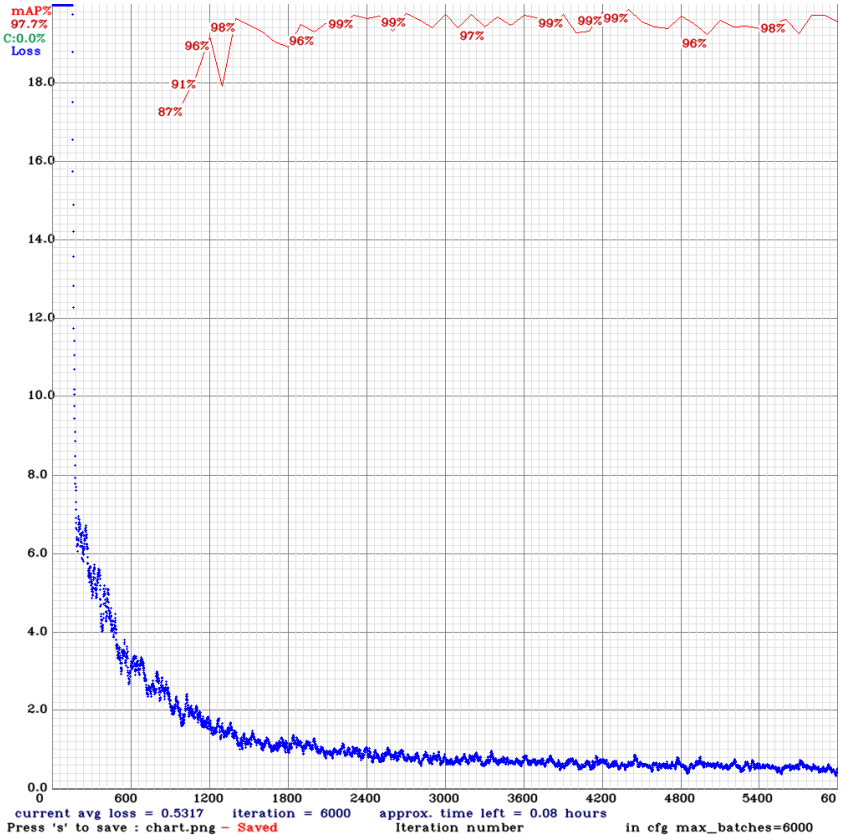

Figura 2. Progresión del entrenamiento.

De esta manera, se destaca la obtención de un *mAP%* del 97,7%, valor que indica cuál fue la medida de precisión alcanzada por el modelo. Este valor se calcula tomando la media de los promedios de precisión (AP) obtenidos para cada una de las experiencias analizas a largo de todas las iteraciones. Para el modelo obtenido, los valores de AP de cada una de las actividades fueron del 99,69% para Tangram, del 95,18% para Cosa de parejas y del 98,11% para Submarino.

AJEA – Actas de Jornadas y Eventos Académicos de UTN
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Texto Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1638

Asimismo, se considera el valor de *average loss*, que refleja el rendimiento general del modelo durante el entrenamiento en una variedad de tareas de *machine learning*, donde el objetivo es minimizar la función de pérdida. En este caso, el valor conseguido fue de 0,5317, el cual se considera un resultado aceptable para la aplicación propuesta.

A continuación, en la Figura 3 se muestra que el modelo logró detectar los tres juegos en una misma toma. Esta imagen se generó levantando el modelo con un script de Python utilizando la librería de *OpenCV*. Es importante destacar que la imagen utilizada no forma parte del *dataset* de entrenamiento, sino que representa una situación real dentro del ámbito del museo. Esta elección se llevó a cabo para poder separar los objetos a reconocer del entorno en los que fueron detectados para el entrenamiento. Los resultados obtenidos mostraron una confianza del 94% para cada una de las experiencias, lo cual se considera un resultado adecuado para la finalidad del modelo.

Figura 3. Juegos detectados por el modelo.

Figura 4. Detección de las experiencias con guías de fondos.

AJEA – Actas de Jornadas y Eventos Académicos de UTN
Libro de actas de resúmenes DOI: https://doi.org/10.33414/ajea.1640.2024
Texto Completo Publicado en AJEA – Full Text Published in AJEA
https://rtyc.utn.edu.ar/index.php/ajea/article/view/1638

Por otro lado, en la Figura 4 se puede observar que al colocar las guías en formato papel, que están presentes en el museo, detrás de las experiencias “Cosa de Parejas” y “Jugando con el Submarino”, la certeza en la detección de cada una disminuye de 94% (como se ve en la Figura 3) a 60% y de 94% (Figura 3) a 65%, respectivamente. De esta forma, se deduce que el *dataset* necesita aún más imágenes para obtener mejores resultados en situaciones donde la actividad es difícil de distinguir de su entorno. Por lo tanto, se plantea como una actividad a realizarse en un futuro trabajo para continuar mejorando el modelo obtenido.

Conclusiones

Con el fundamento de las pruebas realizadas se concluye que YOLO es una alternativa factible y adecuada a la hora de inmiscuirnos en la museografía interactiva a través del reconocimiento por imágenes, lo que permite una rápida identificación de las experiencias presentes en la muestra del MuIC. Con esta implementación, se abre la posibilidad de desarrollar una herramienta que permita a los usuarios acceder a las guías interactivas sin la necesidad de leer un código QR, que puede irrumpir la imagen visual de la muestra.

Basados en los conocimientos adquiridos durante la realización de esta investigación, se concluye que es factible continuar refinando la fórmula e intentando la obtención de mejores resultados mediante el uso de *datasets* más grandes y mejores configuraciones para el entrenamiento. Además, una tarea interesante a considerar es la evaluación de la velocidad con la que diferentes dispositivos pueden capturar una imagen y procesarla para obtener así la lista de juegos detectados, de esta manera se pueden comparar las técnicas empleadas en términos de usabilidad, tal como se propone en (Levrino et al., 2023).

Referencias

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *Yolov4: Optimal speed and accuracy of object detection.* arXiv preprint arXiv:2004.10934.

Gilli L., Cordoba R., & Pipino H. A. (2023). *Desarrollo de Museografía Interactiva Animada*. Jornadas de Ciencias y Tecnología 2023. San Francisco (Córdoba).

Levrino F., Castellano L., Fantin M., Mulassano M., & Pipino H. (2023). *Museografía Interactiva con Acceso mediante Código QR*. Jornadas de Ciencias y Tecnología 2023. San Francisco (Córdoba).

Tzutalin (2015). *LabelImg*. Git code. Disponible en: https://github.com/tzutalin/labelImg

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). *CSPNet: A new backbone that can enhance learning capability of CNN*. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 390-391).

Agradecimientos

Queremos hacer una mención especial a Matías Fantín y al Dr. Ing. Hugo A. Pipino, miembros de MuIC, por su ayuda en el etiquetado de imágenes. Además, deseamos expresar las gracias a la Ing. Micaela S. Mulassano, por su ayuda en el entrenamiento del modelo. Por último, no podemos dejar de mencionar nuestro sincero agradecimiento a ARGIA TECH S.R.L por haber prestado los equipos en los cuales se desarrollaron los entrenamientos.