https://repository.uaeh.edu.mx/revistas/index.php/icbi/issue/archive
Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI

Pädi

ISSN: 2007-6363

Publicación Semestral Pädi Vol. 9 No. Especial (2021) 106-109

Detección de estudiantes que copian en el aula usando Redes Neuronales Convolucionales

Identification of students who copy in the classroom using Convolutional Neural Network

R. Cruz Guerrero [a,*], K. Gutiérrez Fragoso [a]

[a] *Sistemas Computacionales, Instituto Tecnológico Superior del Oriente de Hidalgo, Apan Hidalgo, México. C.P. 43900.*

Resumen

La modernización del proceso educativo implica la automatización de las actividades académicas y administrativas que fomenten un entorno inteligente. Incorporar tecnologías emergentes en las instituciones educativas de nivel superior permitirá transitar hacia la conversión de procesos rutinarios para mejorar la calidad del servicio educativo. El presente trabajo consiste en automatizar la detección de alumnos que copian durante la aplicación de sus exámenes en las aulas utilizando técnicas de Deep Learning con redes neuronales convolucionales. Se obtuvo una precisión de un 95.75% en el modelo de clasificación después de experimentar con diferentes parámetros y arquitecturas de una red neuronal convolucional.

Palabras Clave:
Aprendizaje profundo, redes neuronales convolucionales, clasificación, reconocimiento de imágenes.

Abstract

The modernization of the educational process implies the automation of academic and administrative activities that promote an intelligent environment. Incorporating emerging technologies in higher-level educational institutions will make it possible to move towards the conversion of routine processes to improve the quality of the educational service. The present work consists of automating the detection of students who copy during the application of their exams in their classrooms using Deep Learning techniques with convolutional neural networks. An accuracy of 95.75% was obtained in the classification model after experimenting with different parameters and characteristics of a convolutional neural network.

Keywords:
Deep learning, convolutional neural networks, classification, image recognition.

1. Introducción

La modernización del proceso educativo implica el uso de tecnologías emergentes para automatizar en la mayoría de lo posible las diversas actividades que se efectúan dentro de las escuelas. En los últimos años, se han desarrollado diversos trabajos de investigación que han servido de apoyo para automatizar procesos en las instituciones educativas. Ejemplo de ello es el uso de sistemas biométricos para la identificación de personas que consiste en integrar diversos dispositivos de hardware que pueden trabajar en red para generar un sistema de alertas tempranas al identificar amenazas de seguridad al interior de los planteles educativos (Delgado, et. al., 2019). El desarrollo del proyecto describe el uso de placas raspberry, cámara, software y los elementos se integran en una arquitectura del internet de las cosas, considerando un modelo de capas: captura de datos, almacenamiento, procesamiento y administración.

En (Vázquez, 2014) se realiza reconocimiento facial mediante técnicas de visión tridimensional, se expone una técnica novedosa para la identificación de personas con imágenes 3D, reemplazando al método típico del uso de contraseñas o por identificadores de radio frecuencias (RFID). Aplicando técnicas basadas en biometría, obtuvieron una solución más confiable ya que toma en cuenta los rasgos físicos. Aplicaron la reconstrucción tridimensional, la cual tiene un amplio campo de aplicación, particularmente con la visión estereoscópica que es más cercana a la forma como lo hace el ojo humano.

En el trabajo de Sudhakar, et. al., los autores abordan el problema la detección de rostros con múltiples vistas. Se enfocan principalmente en la anotación de puntos de referencia faciales,

*Autor para la correspondencia: rcruz@itesa.edu.mx
Correo electrónico: rcruz@itesa.edu.mx (René Cruz-Guerrero), kgutierrez_@itesa.edu.mx (Karina Gutiérrez-Fragoso)

Fecha de recepción: 15/05/2021 Fecha de aceptación: 31/05/2021 Fecha de publicación: 05/08/2021
https://doi.org/10.29057/icbi.v9iEspecial.7492

proponen un método que no requiere anotación de pose / punto de referencia y es capaz de detectar rostros en una amplia gama de orientaciones utilizando un sólo modelo basado en redes neuronales convolucionales profundas (Sudhakar, et. al., 2015).

Las evaluaciones en conjuntos de datos de referencia de detección de rostros populares muestran que el algoritmo tiene un rendimiento similar o mejor en comparación con los métodos típicos, que son más complejos y requieren anotaciones de diferentes poses o marcas faciales. La mayoría de los trabajos desarrollados para identificar rostros en una escuela es con la finalidad de permitir o negar el acceso a cierta área de las instalaciones, sin embargo, hay diversos campos donde estas técnicas pueden aplicarse.

En el presente trabajo se propone la aplicación de Redes Neuronales Convolucionales para detectar si los estudiantes copian en un examen. La aplicación de evaluaciones escritas es una actividad que se realiza con mucha frecuencia en las aulas, ya sea durante el curso de un periodo escolar en los procesos de admisión al inicio de cada semestre. Debido a que esta actividad demanda de una constante supervisión por parte de los docentes, se propone aportar una herramienta que les sirva de apoyo, para ello se utiliza la tecnología de Deep Learning en específico el uso de redes neuronales convolucionales.

2. Visión computacional

2.1 Deep Learning

En los últimos años ha emergido un área específica del Machine Learning llamada Deep Learning o Aprendizaje Profundo, que se basa en algoritmos de aprendizaje en múltiples niveles de representación y de abstracción con el fin de modelar relaciones más complejas entre los datos. Las fases corresponden con distintos niveles de conceptos, donde los más altos vienen definidos por los más bajos de forma jerárquica. Esta idea de representaciones sucesivas por capas es lo que da el nombre de "profundo" a este tipo de aprendizaje, siendo la profundidad del modelo el número de capas que contiene (Deng & Yu, 2013).

En Deep Learning, estas representaciones por capas forman una Red Neuronal (Neural Network) que, al fin y al cabo, es un conjunto estructurado de neuronas. Estos términos provienen de la neurobiología ya que estas redes se inspiran en el funcionamiento del cerebro humano. Las capas están formadas por un número determinado de neuronas, donde cada neurona recibe información de la capa anterior por medio de estímulos externos a través de sus conexiones de entrada. Las neuronas realizan cálculos internamente y devuelven un valor de salida que se transmite a las neuronas de la capa siguiente. En la última capa, la salida es la predicción buscada (Chollet, 2017).

La ventaja del aprendizaje en Deep Learning es que permite que todas las capas del modelo aprendan al mismo tiempo, en lugar de continuamente. Cuando el modelo ajusta un parámetro, el resto de los parámetros que dependen de él se ajustan automáticamente. De esta forma, todo está supervisado bajo una única señal de realimentación.

Otras propiedades importantes que hacen del Deep Learning una revolución en el mundo de la inteligencia artificial son la simplicidad, la escalabilidad y la versatilidad. La simplicidad se debe a la automatización del feature engineering, esto es, el proceso de creación de características que tanto tiempo y esfuerzo supone en las etapas previas al desarrollo de un modelo. La escalabilidad permite trabajar con conjuntos de datos de diferentes tamaños, mientras que la versatilidad hace que algunos modelos ya entrenados se puedan emplear para otros propósitos.

2.2 Redes neuronales convolucionales

Por sus siglas en inglés Convolutional Neural Networks (CNN). Se trata de una categoría de redes neuronales respetables en el campo del reconocimiento y la clasificación de imágenes. Las CNN utilizan perceptrones multicapa que requieren un preprocesamiento mínimo para entrenar la arquitectura para que realice la tarea de reconocimiento / clasificación de manera muy eficaz. Tienden a funcionar mejor que otros algoritmos de reconocimiento de objetos y video en los campos de clasificación de imágenes y procesamiento del lenguaje natural.

Este modelo ha logrado un gran avance en el campo de la clasificación de imágenes y la detección de objetos. Las CNN profundas introducen una gran cantidad de capas ocultas, lo que reduce la dimensionalidad de la imagen y permite que el modelo extraiga características de imagen escasas en un espacio de baja dimensión.

2.3 Arquitecturas de las CNN

Diversas organizaciones como IBM y Google han contribuido con arquitecturas CNN como ALexNet, LetNet y GoogleNet. A continuación, se exponen algunas de las más utilizadas en la literatura.

LeNet.- Es una de las primeras aplicaciones exitosas de Redes Convolucionales que fueron desarrolladas por Yann LeCun en la década de 1990. Es una de las primeras arquitecturas que se aportaron, utilizadas inicialmente para procesos como: leer códigos postales, caracteres y dígitos. (LeCun, Bengio y Hinton, Deep learning. Nature 2015). En la Figura 1 se muestran las capas que la constituyen.

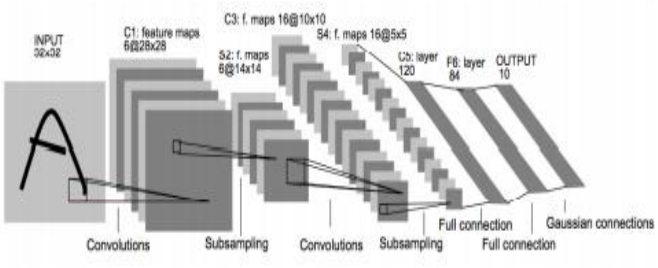

Figura 1: Arquitectura LeNet.

AlexNet.- El primer trabajo que popularizó las redes convolucionales en visión artificial fue AlexNet, desarrollado por investigadores encabezados por Alex Krizhevsky. Tiene una arquitectura muy similar a LeNet, pero con mayor profundidad, más grande y presenta capas convolucionales apiladas una encima de la otra (Krizhevsky, Sutskever y Hinton 2012). En la Figura 2 se muestra su estructura.

ZF Net.- Es una mejora en AlexNet al ajustar los hiperparámetros de la arquitectura, en particular al expandir el tamaño de las capas convolucionales medias y reducir el tamaño del paso y el filtro en la primera capa. Su autor fue el ganador del ILSVRC 2013 y comúnmente se aplica para el tratamiento de imágenes y audio. (Zeiler & Fergus, 2013). En la Figura 3, se muestra la distribución de sus capas.

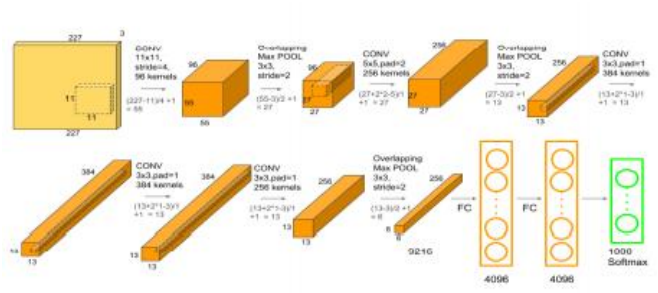

Figura 2: Arquitectura AlexNet.

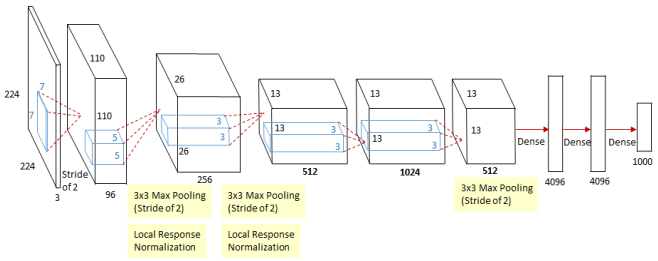

Figura 3: Representación de arquitectura ZF Net.

3. Desarrollo

La metodología de desarrollo incluye los siguientes procesos: preparación de datos, creación del modelo de clasificación, desarrollo de la interface de usuario y pruebas. En la Figura 4, se muestran los procesos principales que intervienen para desarrollar el sistema de identificación de imágenes.

3.1 Preparación de Datos

En esta etapa, primero se recopilaron las imágenes que conforman el conjunto de datos original utilizado para crear el modelo de clasificación. Dichas imágenes se obtuvieron de la web mediante el uso de la técnica Webscraping con librerías de Python tomando en cuenta las dos clases a considerar en el modelo, es decir imágenes que representan comportamientos que tienen los estudiantes que copian y que no copian, ejemplo de ello se muestra en las Figuras 5 y 6, respectivamente.

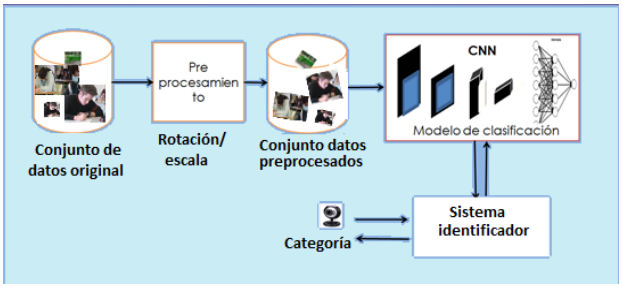

Figura 4: Procesos de la metodología propuesta.

Para la homologación del tamaño, cada imagen recopilada se transformó en un formato de 240 x 240 pixeles. El conjunto inicial está conformado por 117 imágenes de la clase copian y 123 de la clase no copian. A este conjunto se le aplicó el pre-procesamiento estandarizando su tamaño. Posteriormente, con la finalidad de tener un conjunto de datos más variado y enriquecido, a las imágenes iniciales se les aplicaron transformación de rotación y escalado.

Referente al proceso de rotación, se aplicó seleccionando algunas imágenes del conjunto original de manera aleatoria y aplicando diversos ángulos de inclinación utilizando la ecuación 1.

$$x' = x_c + (x - x_c)\,cos\theta - (y - y_c)\sin\theta$$
$$y' = y_c + (x - x_c)\,sin\theta + (y - y_c)\cos\theta \qquad (1)$$

Después de aplicar las operaciones de transformación, el conjunto de datos a utilizar para crear el modelo de clasificación, quedó conformado por 298 imágenes de la clase copia y 274 de la clase no copia.

Figura 5: Imágenes donde el estudiante está copiando.

3.2 Creación del modelo

Con el objetivo de encontrar el mejor desempeño en el modelo clasificador, se realizó la experimentación en dos etapas. La primera consistió en evaluar el desempeño de una red neuronal convolucional típica (LetNet) probando diversos valores con sus diferentes parámetros como número de filtros, épocas, pasos y tamaños de filtro, entre otros. Por su parte, en la segunda etapa se realizaron pruebas con arquitecturas como ZFNet, ResNet y AlexNet.

Figura 6: Imágenes donde el estudiante no está copiando.

El modelo de clasificación fue creado en Python utilizando las librerías Tensorflow y Keras. Para dicha creación, del conjunto de datos, se asignó el 80% de las imágenes para el entrenamiento (X_Train) y el 20% para prueba (X_Test).

En la Tabla 1, se muestran los resultados obtenidos después de realizar pruebas con diversos parámetros. Como se puede observar el mejor desempeño del clasificador se obtuvo después de aplicar 30 épocas, 15 pasos y 32 filtros de tamaño 5x5.

Tabla 1: Resultados con diferentes valores de los parámetros

No prueba	Num. Epocas	Num. Pasos	Num. Filtros	Tamaño de filtro	Accuracy
1	10	10	32	3x3	0.86
2	10	10	64	3x3	0.7143
3	16	10	32	3x3	0.6813
4	16	15	64	5x5	0.860
5	30	15	32	5x5	**0.9575**
6	30	15	64	5x5	0.9475

En la Figura 7, se muestra un ejemplo de los resultados obtenidos al generar el modelo, en este caso corresponde a la prueba número 4, donde se aplicaron 16 épocas, 15 pasos y 64 filtros de 5x5.

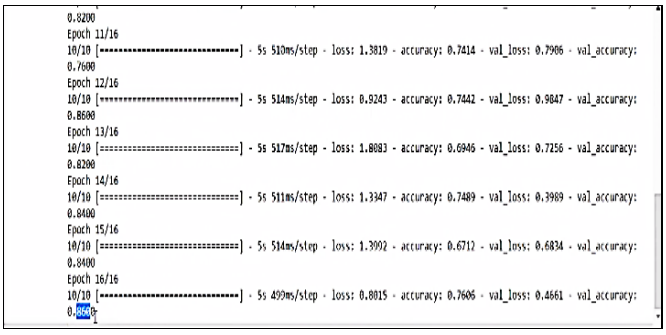

Figura 7: Resultado obtenido aplicando 16 épocas y 64 filtros.

En la Tabla 2, se muestran los resultados obtenidos con las diferentes arquitecturas aplicadas, como se puede observar los mejores desempeños se obtuvieron con las Arquitecturas LetNet y ZFNet con un 0,9575 y 0,9570 respectivamente.

El desempeño obtenido en el entrenamiento del modelo fue de 0.9575 y el de las pruebas realizadas con el conjunto de prueba fue de 94.92% lo cual ratifica que es un modelo de clasificación adecuado.

Después de crear el modelo clasificador, se programó la interface de usuario en Python utilizando la librería CV2.

Tabla 2: Resultados con diversos parámetros de una CNN típica

Arquitectura	Num. Epocas	Num. Filtros	Tamaño de filtro	Accuracy
LetNet	30	32	5x5	**0,9575**
ResNet	30	32	5x5	0,9487
AlexNet	30	32	5x5	0,9474
ZF Net	30	32	5x5	**0,9570**

4. Conclusiones

En las instituciones educativas se están implementando nuevas tecnologías para automatizar más sus procesos. El sistema propuesto tiene como finalidad proveer a los docentes de una herramienta que permita automatizar la supervisión de los estudiantes durante la aplicación de sus exámenes. Los resultados muestran que el uso de las técnicas de aprendizaje profundo brinda buen desempeño en el tratamiento de las imágenes. Las arquitecturas LetNet y ZFNet mostraron mejor desempeño que AlexNet y ResNet en varias pruebas realizadas.

Referencias

Delgado, A., Timana, D., Golondrino, E. and Villalba K., (2019). Arquitectura IoT para la identificación de personas en entornos educativos, Revista Ibérica de Sistemas y Tecnologías de Información, 841-853.

Krizhevsky, A, I Sutskever, y G Hinton (2012), Imagenet classification with deep convolutional neural networks, In Advances in neural information processing systems, 1097-1105.
DOI:10.1145/3065386

LeCun, Y. (2012). Learning Invariant Feature Hierarchies. European Conference on Computer Vision, European Conference on Computer Vision.
DOI: 10.1007/978-3-642-33863-2_51

LeCun, Y, Y Bengio, y G Hinton. (2015). Deep learning. Nature. 521(7553), 436-444.

Sudhakar, S., Saberian, M. y Li, J. (2015). Multi-view face detection using Deep Convolutional Neural Networks. Sitio oficial de la Association for Computing Machinery.
DOI: 10.1145/2671188.2749408

Szegedy, C. (2015), Going deeper with convolutions, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
DOI:10.1109/CVPR.2015.7298594

Triantafyllidou, D. y Tefas, A. (2016) Face detection based on deep convolutional neural networks exploiting incremental facial part learning, 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico.,3560-3565.
DOI:10.1109/ICPR.2016.7900186

Vázquez, M. (2014) Sistema de reconocimiento facial mediante técnicas de visión tridimensional [Tesis de maestría. Centro de investigaciones en Óptica]. Repositorio Institucional.

Zeiler, M D, y R Fergus. Visualizing and understanding convolutional networks. Lecture Notes in Bioinformatics, 2014.
DOI:10.1007/978-3-319-10590-1_53