Reconocimiento de emociones con redes neuronales

ITESM Campus Querétaro
Visión Robótica, Grupo 1
Rodrigo Gutiérrez Alvarez A01207532
Roberto Carlos Guzmán Cortés A01702388
Semestre FJ-2022
13 Junio, 2022

Índice

Introducción

Durante los últimos años hemos sido testigos de cómo la inteligencia artificial ha sido cada vez más parte de nuestras vidas: desde asistentes de voz hasta aplicaciones que desbloquean un celular con sólo mirar nuestro rostro. Todo lo que realiza la tecnología hoy en día parece magia y más aún cuando se le dota de comportamientos inteligentes para aprender de los datos que le son suministrados por infinidad de usuarios cotidianamente.

Recientemente se ha hablado de cómo las redes neuronales han revolucionado el campo de la inteligencia artificial, dando paso a una nueva área de estudio conocida como aprendizaje profundo o aprendizaje a profundidad (del inglés *deep learning*). Sin embargo, esta idea no es del todo nueva ya que antes de este siglo, hubo numerosos autores que enriquecieron esta área y acuñaron este concepto.

A lo largo de este documento explicaremos un poco sobre lo que es deep learning, arquitecturas, herramientas comúnmente usadas en la creación de aplicaciones con comportamientos inteligentes y sobre todo la explicación de cómo implementar un proyecto de este tipo: reconocimiento de edad y emociones con redes neuronales.

Marco teórico

La inteligencia artificial ha sido definida de muchas maneras dependiendo del autor a abordar; desde una perspectiva humana se ha definido como una ciencia que intenta replicar aquellos comportamientos inteligentes propios del ser humano en máquinas[1].

Comúnmente gracias a los medios de comunicación se ha relacionado a la inteligencia artificial con robots que superan a seres humanos en alguna competencia, más en realidad esta es un área de aplicación (robótica), ya que como se mencionó en el párrafo anterior se trata de emular comportamientos inteligentes independientemente de su área de aplicación. Algunas de las principales áreas de investigación de las inteligencia artificial son[1]:

- Natural Language Processing (NLP)
- Knowledge representation and reasoning
- Automated planning
- **Machine learning (ML)**
- Machine perception
- Intelligent robots

En los siguientes apartados nos enfocaremos en una subárea de Machine Learning (ML) cuya principal aplicación son las redes neuronales, dicha subárea es mejor conocida con el nombre de Deep Learning (DL).

Deep Learning

Mientras que Machine Learning (ML) es el área en la que la finalidad es que una sistema aprenda a partir de los datos que le son suministrados, deep learning (DL) es una representación más abstracta de cómo se logra dicha meta a través de capas de abstracción conformadas a su vez por neuronas (funciones de activación, regresión, etc.) que transmiten información sintetizada por cada una de ellas a sus capas aledañas[2].

En los apartados de redes neuronales convolucionales y descripción del proyecto conoceremos más sobre la forma en que dichas capas y neuronas se organizan para lograr satisfacer con gran exactitud dichas tareas.

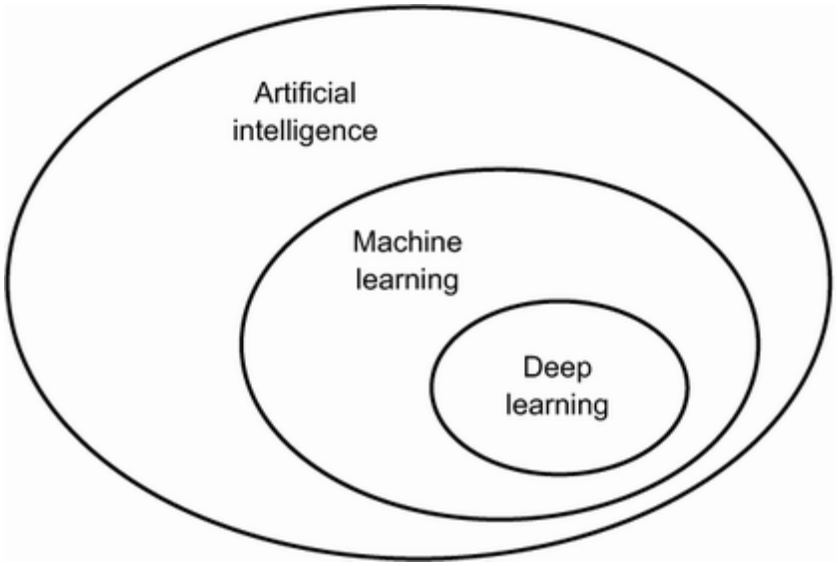

Figura 1. Relación entre IA, ML y DL[2]

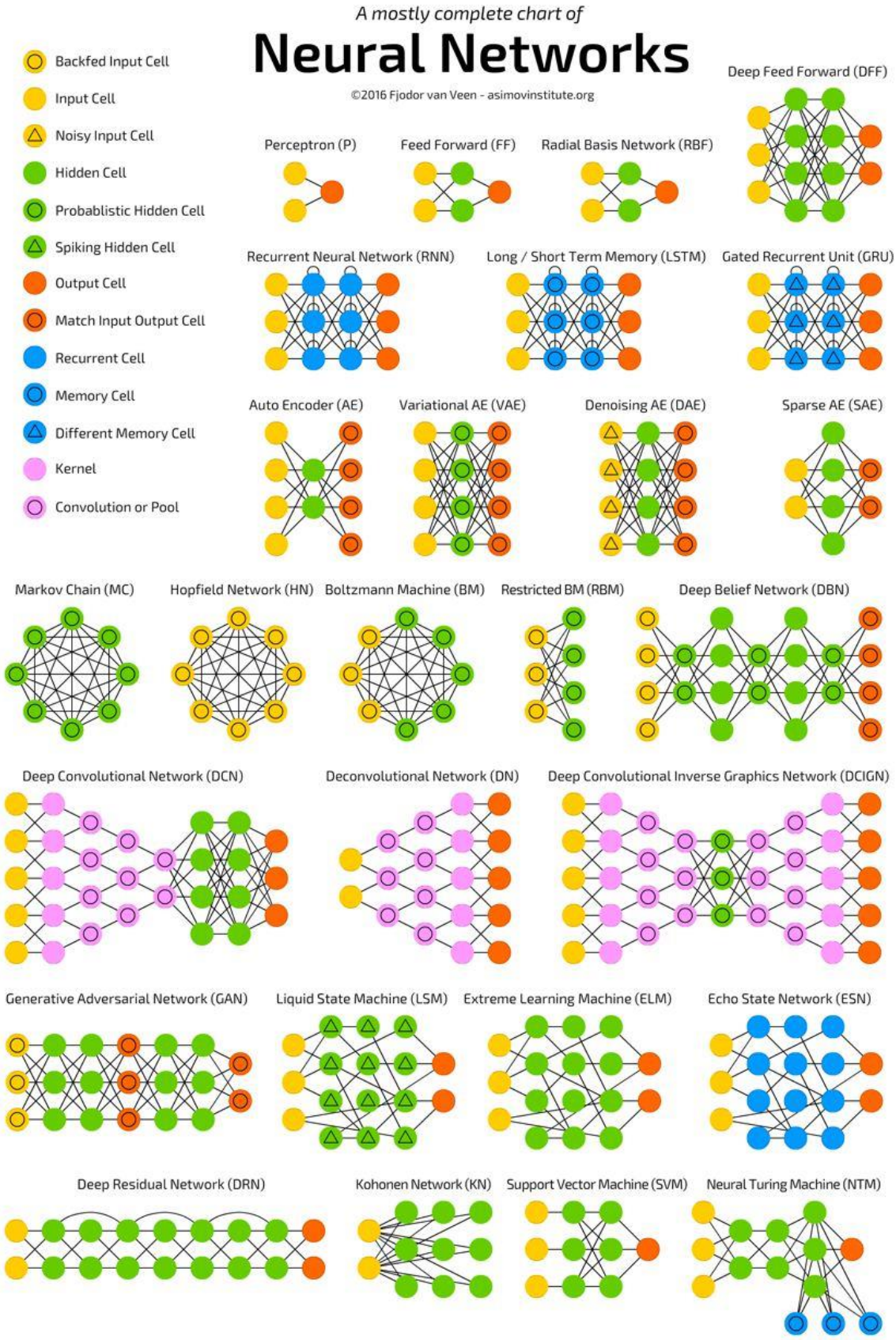

Figura 2. Neural Network architectures[3]

Librerías

A lo largo de las décadas pasadas entre más progresaba el campo de la inteligencia artificial se fueron creando frameworks y librerías con la finalidad de facilitar la vida a la comunidad de desarrolladores. A continuación se presentan las librerías más comunes en torno al desarrollo de aplicaciones que hacen uso de redes neuronales:

Keras

Keras es una librería open source para Python, enfocada en redes neuronales. Keras funciona como una interfaz para TensorFlow. [4]
Keras esta designada para realizar experimentos rápidos de redes neuronales, con un enfoque user-firendly. Su desarrollo fue resultado del proyecto ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System).[5]

Pytorch

PyTorch es un framework de código abierto enfocado al machine learning basado en la librería Torch, comúnmente usado en aplicaciones de visión computacional y procesamiento de lenguajes, desarrollado principalmente por MetaAI. [6]

OpenCV

OpenCV es una librería abierta, enfocada en visión computacional y machine learning. OpenCV fue desarrollada para proveer una infraestructura común para aplicaciones de visión y acelerar el uso de percepción de la máquina para usos comerciales.[7]

Scikit-learn

Scikit-learn es una librería gratuita para machine learning implementada en el lenguaje de programación python. Algunas de sus características son clasificación, regresión y agrupamiento de algoritmos, incluyendo support-vector machines, random forests, gradient boosting y DBSCAN. [8]

Caffe

Caffe es un framework de aprendizaje profundo hecho con expresión, velocidad y modularidad. Fue desarrollado por Berkeley AI Reaserch (BAIR), junto con la contribución de la comunidad. [9]

Frameworks

Algunas de las librerías anteriormente mencionadas forman parte de marcos de trabajo (frameworks) los cuales nos proveen de un ambiente con múltiples herramientas (no sólo librerías) para poder desarrollar más rápidamente aplicaciones que usen técnicas de inteligencia artificial, aunado a que definen estándares de calidad durante su proceso de creación.
Entre los frameworks para inteligencia artificial más ampliamente conocidos por la comunidad de desarrolladores tenemos:

Tensforflow

TensforFlow es una librería open source gratuita dedicada a machine learning e inteligencia artificial. Está librería suele ser usada en múltiples plataformas, con múltiples propósitos, principalmente el entrenamiento de redes neuronales[10].
TensforFlow puede ser implementado en múltiples lenguajes de programación tales como Python, Javascript, C++ y Java[11].

Microsoft Cognitive Toolkit / CNTK

El CNTK o Microsoft Cognitive Toolkit es una herramienta de tipo open source comercial enfocada al deep learning. Ayuda a describir redes neuronales dentro de una serie computacional mediante grafos directos. La herramienta permite a los usuarios combinar múltiples modelos. [12]

Redes Neuronales Convolucionales

Las redes neuronales convolucionales son un tipo de redes neuronales donde las neuronas artificiales, corresponden a campos receptivos de una manera muy similar a las neuronas de la corteza visual de un cerebro biológico. Este tipo de redes son variaciones de un perceptrón multicapa, sin embargo, debido a que su aplicación es realizada en matrices bidimensionales, son muy efectivas para tareas de visión artificial, como en la clasificación y segmentación de imágenes, entre otras aplicaciones. [13]

La principal operación que permite lo mencionado anteriormente recibe el nombre de convolución, la cual consiste en hacer deslizamientos de una matriz bidimensional (en el caso de las imágenes) llamada kernel, la cual dará como salida los filtros requeridos para la tarea de reconocimiento, en el siguiente gráfico se muestra un ejemplo de ello.

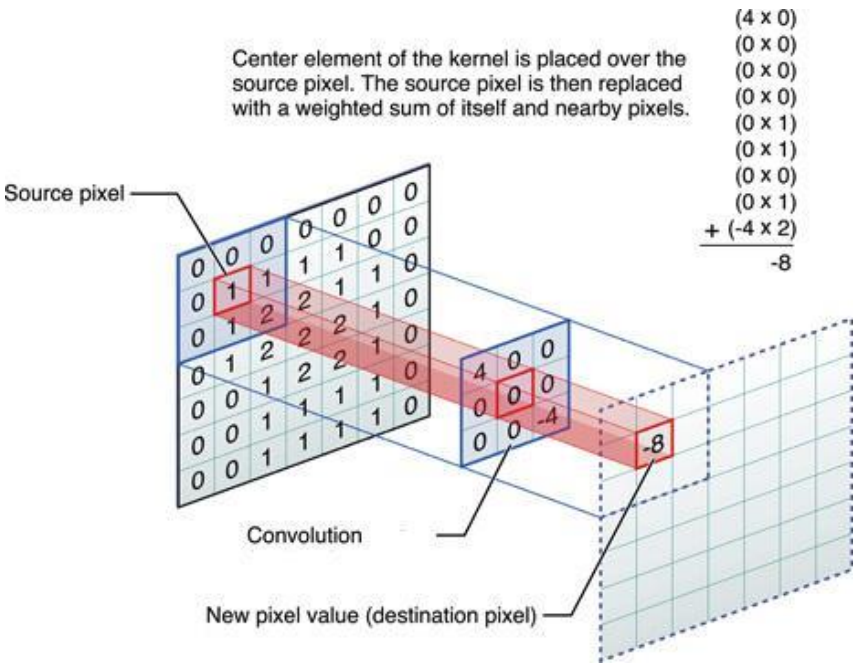

Figura 3. Convolución 2D sobre una imagen[14]

Con el paso de nuestra imagen tras las distintas capas convoluciones, *maxpool* y *dropout* se irá robusteciendo un tensor 3D conocido como “el mapa de características de la imagen” (*feature map*), el cual contiene en su dimensión de profundidad las características detectadas en la imagen de entrada. El siguiente gráfico representa claramente como ocurre dicho proceso:

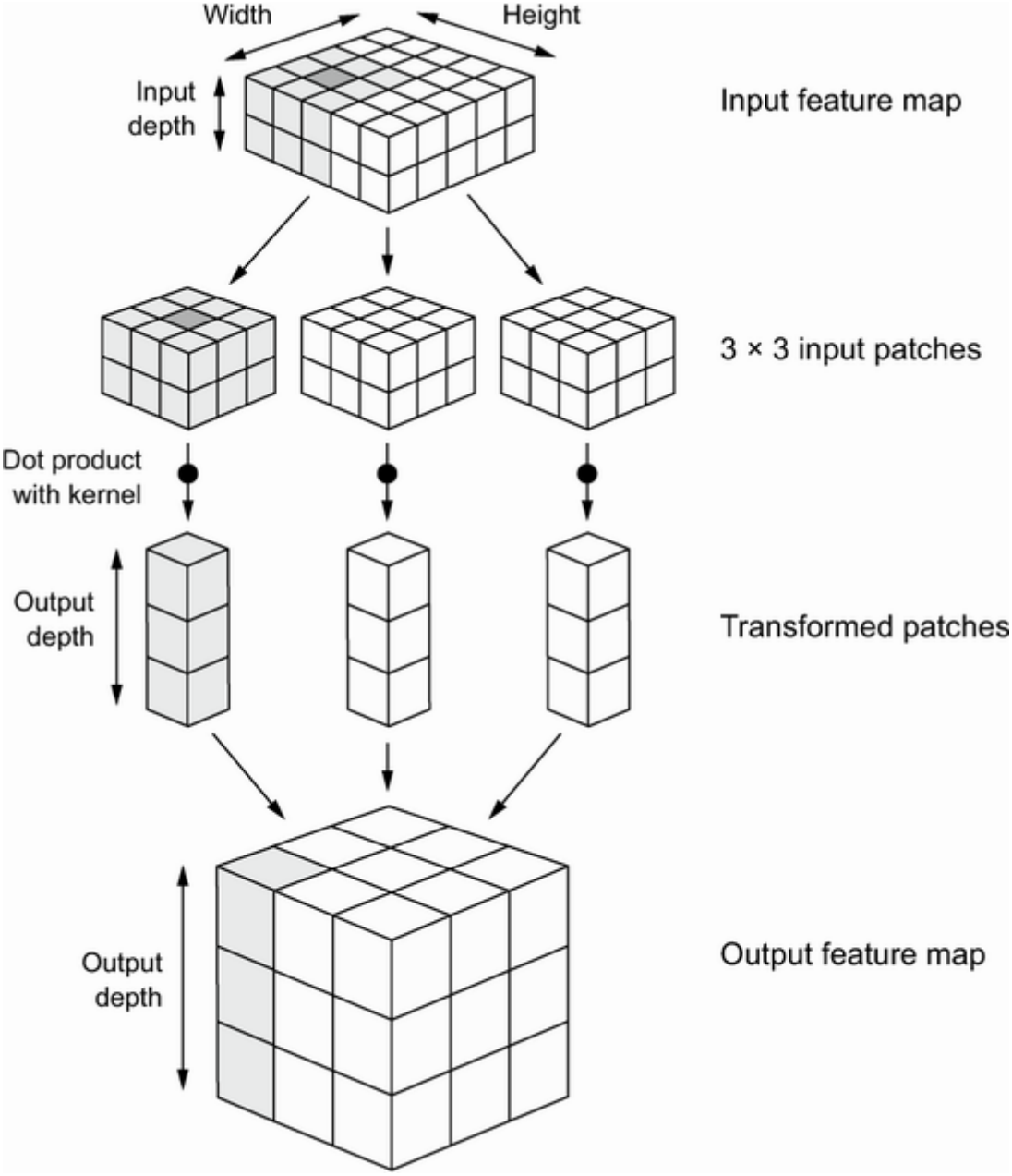

Figura 4. Proceso de conformación del *feature map*[2]

Arquitecturas

Debido a las diversas formas en que pueden estar organizadas las capas de una red, la cantidad de ellas, el número de neuronas y las relaciones que establecen entre sí, podemos disponer de diversas arquitecturas, tales como las que veremos en los próximos apartados.

VGG

Creada por el grupo “Visual Geometry Group” de la universidad de Oxford, esta arquitectura usa algunas ideas de su predecesor (AlexNet) las cuales ha conseguido mejorar significativamente, en 2014 destacó como un modelo “estado-del-arte” para la resolución de problemas que involucran reconocimiento y clasificación de imágenes.
Una de sus principales características es la combinación de capas convolutivas junto con capas del tipo “*Maxpool*” y “*Dropout*”[15]. Algunas de sus tantas variantes son:

- VGG11
- VGG13
- VGG16
- VGG19

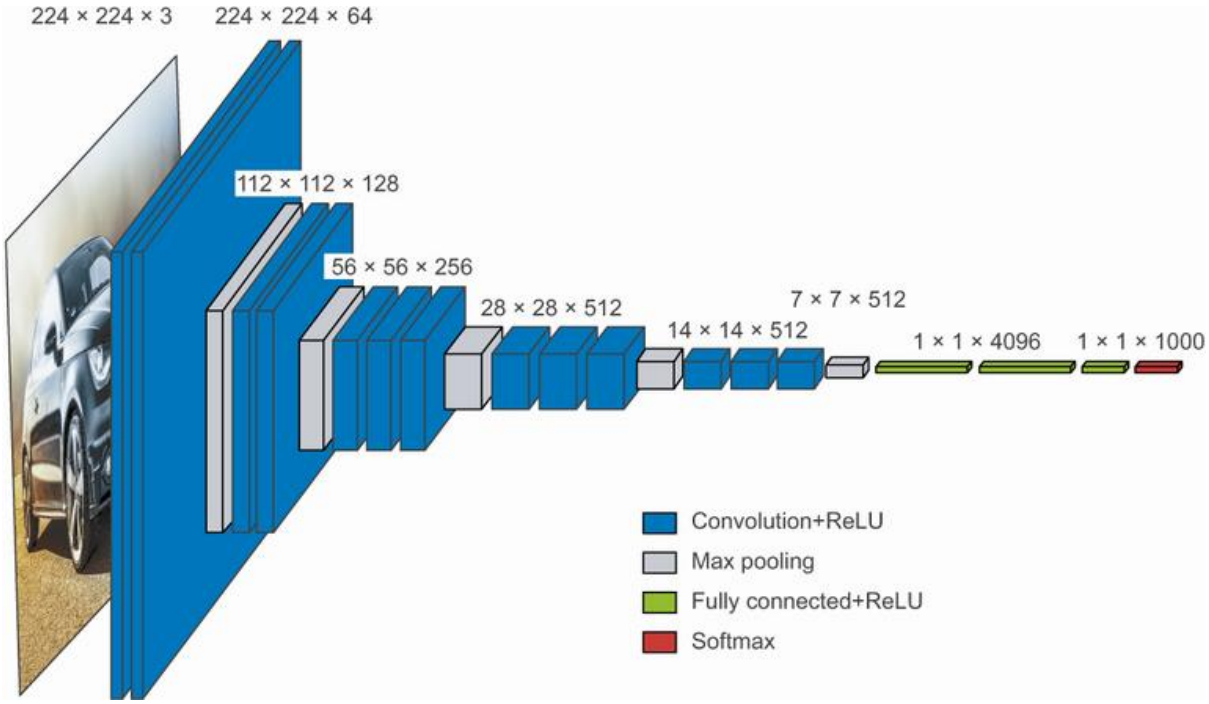

Figura 5. Ejemplo de una red neuronal convolucional, arquitectura VGG16[2]

GoogLeNet

Presentada por google en 2014 para resolver el *ImageNet Large-Scale Visual Recognition Challenge 2014* (ILSVRC14) es otro tipo de red neuronal convolucional que consiste originalmente de 22 capas la cual tiene diversos usos tales como reconocimiento de rostros, reconocimiento de texto en anuncios de publicidad, clasificación y reconocimiento de objetos, etc.

Su enfoque modular le permite agregar o quitar capas de distinta índole dándole con ello capacidad de resolver diversas tareas[16]. En el siguiente gráfico podremos apreciar mejor lo anteriormente explicado:

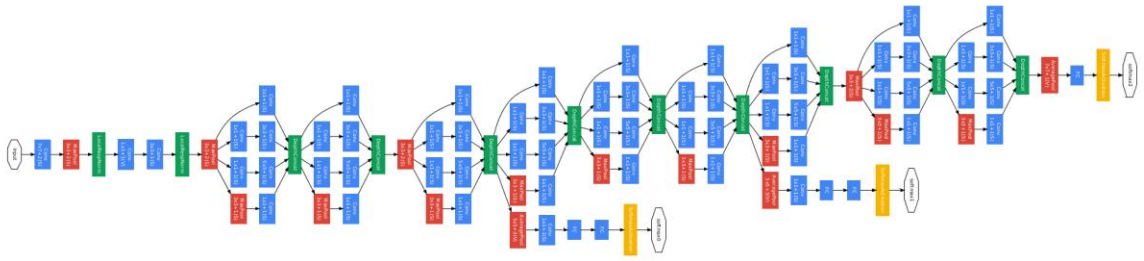

Figura 6. Estructura GoogLeNet[16]

ResNet

Investigadores en el área de las redes neuronales encontraron que a mayor número de capas se requería mayor poder de cómputo, sin embargo, también notaron que podrían obtenerse resultados cada vez más interesantes, ya sea por su exactitud (*accuracy*), comportamiento en entrenamiento, etc. Quedando pendiente un gran problema por resolver: el desvanecimiento del gradiente (*vanishing gradient and gradient exploding* en inglés), dicho problema consistía en que los pesos que eran transmitidos entre capas tendían a reducirse a cero o incrementar a cantidades infinita y con ello se limitaba el número de capas a usar en una red neuronal. La solución: residual networks[17].

A manera de síntesis podemos decir que se crean conexiones adicionales entre las capas para con ello manejar de mejor manera “el proceso de aprendizaje de la red”, controlando con ello mejor la actualización de los pesos de las capas de neuronas mediante comparaciones con capas previas. En la siguiente imagen se puede visualizar mejor dicho proceso:

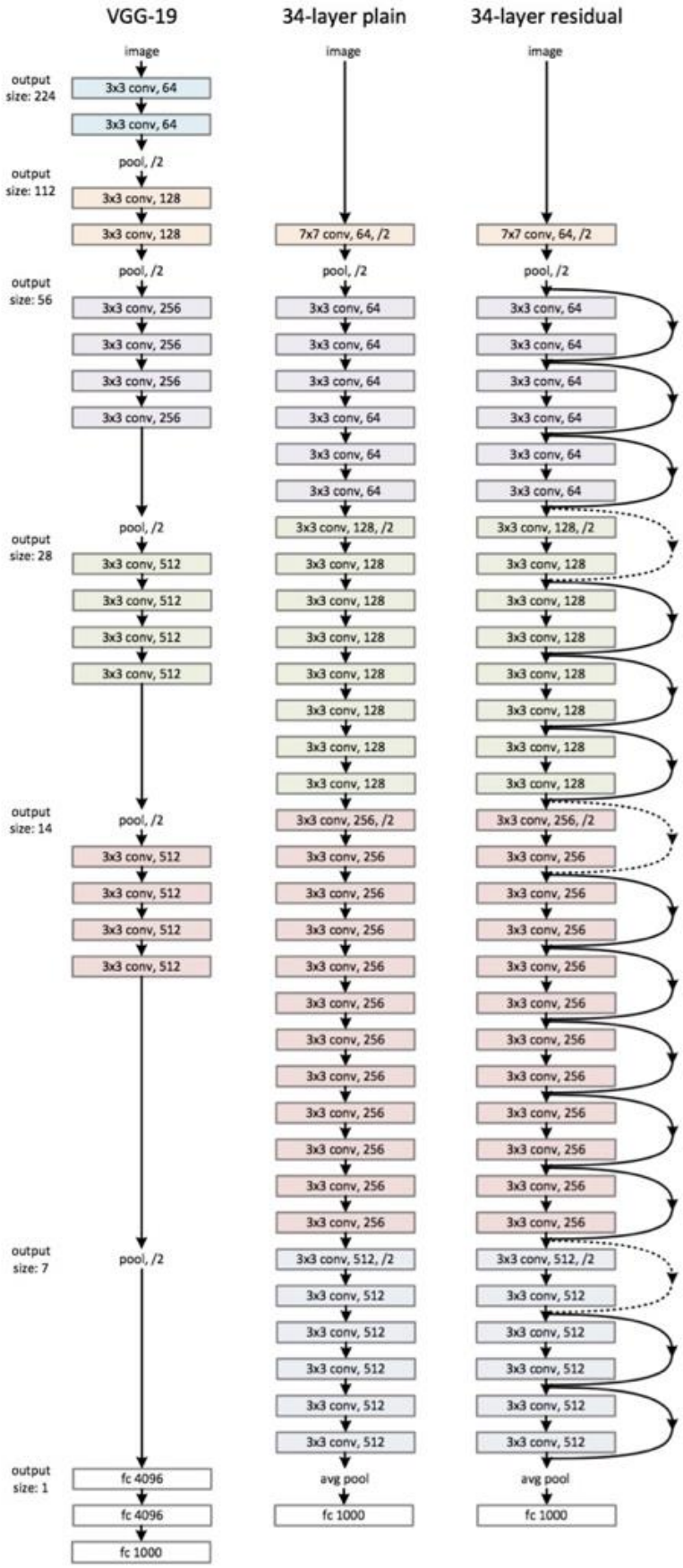

Figura 7. Comparación entre diversas redes neuronales[17]

Documentación

En los siguientes apartados se darán a conocer los detalles técnicos del presente proyecto, así como los pasos necesarios para poder ejecutarlo y hacer uso del mismo.

Stack tecnológico

El conjunto de herramientas que hicieron posible la implementación de este proyecto en sus distintas etapas son:

Aspecto	**Herramienta**
Hardware de optimización	GPU
Framework	Tensorflow
Librerías	Keras OpenCV
Tipo de red neuronal	Convolucional
Arquitectura	VGG11
Lenguaje de programación	Python Javascript
Editor de código fuente	Google colab
Navegador	Google Chrome

Datasets

Para que nuestra red neuronal convolucional cumpliera su objetivo descargamos el siguiente dataset en el sitio oficial de Kaggle[18]:
https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

El cual contiene originalmente dos folders (training and validation) con sus correspondientes 7 clases de emociones:

- Angry
- Disgust
- Fear
- Happy
- Neutral
- Sad
- Surprise

Sin embargo, debido al grado de exactitud que se lograba en entrenamiento con las 7 emociones (60% aproximadamente), se decidió eliminar 3 clases para con ello incrementar el grado de exactitud durante la fases de entrenamiento y validación, así como las correspondientes pérdidas de errores (training and validation loss) en los mismos.

Las clases que al final permanecieron fueron:

- Angry
- Happy
- Neutral
- Sad

Adicionalmente, cabe destacar que cada imagen tiene una resolución de 48x48 y están en escala de grises, lo anterior es comprensible debido a que una mayor resolución y la presencia de colores aumentarán la demanda de poder de cómputo para hacer un procesamiento a profundidad de cada imagen.

Setup

El presente proyecto se basó en la implementación de Abhishek Sharma[19] dónde también hace un sistema de detección de emociones en tiempo real. Sin embargo, esta implementación posee una limitante, la cual es el entorno de ejecución, se necesita copiar e instalar en el ordenador de cada usuario todo el conjunto de librerías y dependencias y fue esta área de oportunidad la cual decidimos abordar.

Google Colab es un entorno de desarrollo online que nos brinda demasiadas herramientas para el desarrollo de aplicaciones que incorporan inteligencia artificial, tales como librerías de deep learning como keras, frameworks de inteligencia artificial como tensorflow, aceleradores de ejecución como Graphic Processor Units Nvidia y Tensor Processor Units.

Para hacer uso del presente proyecto basta con crear una copia de proyecto en la unidad particular de google drive. En el siguiente link se encuentra el acceso:

https://colab.research.google.com/drive/1EZCHCZDIwWcrFne086EIoYZCtnHtIx8n?usp=sharing

Posteriormente, para su correcta ejecución se recomienda hacer click en el logotipo redondo que podremos ver en la siguiente imagen:

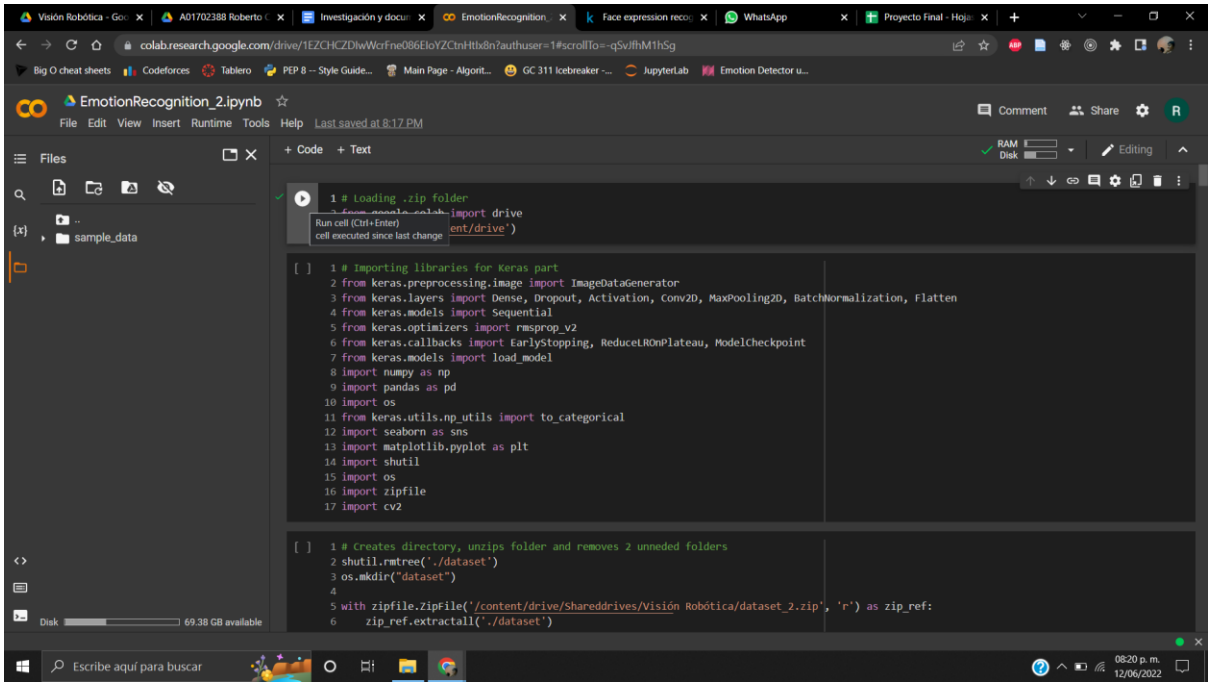

Figura 8. Emotion Recognition project en Google Colab

Es necesario ejecutar cada *code snippet* en orden descendente, es decir, primero los bloques de arriba y esperar a que completen su ejecución (mostrarán una palomita verde) para proseguir en el uso correcto de la aplicación.

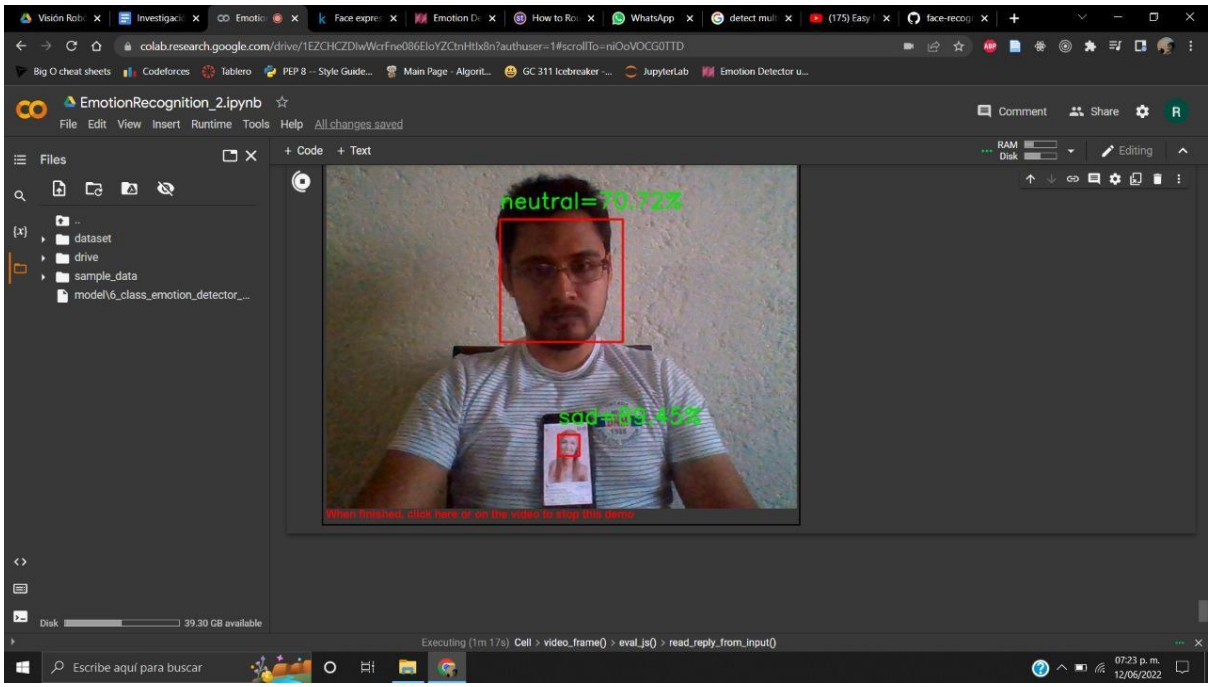

Figura 9. Ejecución y resultados obtenidos

Resultados

Tras remover 3 clases del sistema se obtuvieron las siguientes mejoras:

- El parámetro **training accuracy** pasó de 60% a un 70% en el caso promedio y 75% en el mejor de los casos
- El parámetro **validation accuracy** pasó de un 50% a un 60% en el caso promedio y hasta 62% en el mejor de los casos

En las siguiente gráficas se podrá apreciar mejor lo descrito anteriormente:

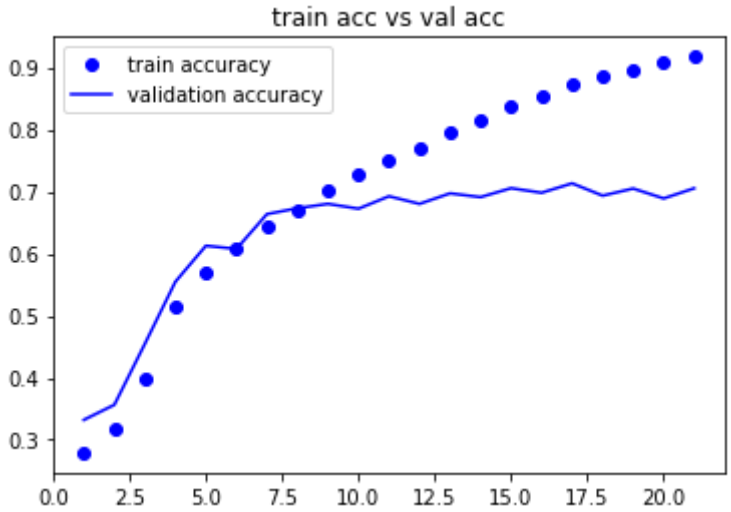

Figura 10. Training accuracy vs validation accuracy

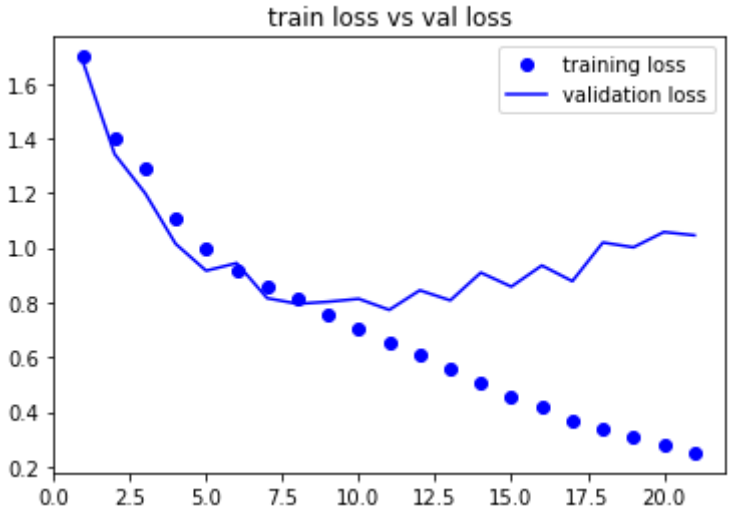

Figura 11. Training loss vs validation loss

Como se puede apreciar, los procesos relacionados con validación tienden a alcanzar un mínimo global para posteriormente tomar un comportamiento que perjudica el desempeño de la red neuronal: **validation accuracy** decrementando y **validation loss** incrementando. La solución a esto recibe el nombre de *early stopping* lo cual consiste en detener el entrenamiento del modelo cuando las métricas de su rendimiento comienzan a tornarse perjudiciales para el funcionamiento del mismo y al momento de aplicar esta medida correctiva se retoman los valores que existían durante el mínimo global para futuras predicciones, es decir, se guarda memoria del mejor momento en que el modelo se había entrenado.

Áreas de aplicación

Existen numerosas áreas en las que la inteligencia artificial puede ayudarnos a obtener provecho de los datos recopilados y no precisamente sólo aquellos que pueden tomarse de forma numérica (escalares) y textual sino también en formatos multimedia como lo son las imágenes.

En los siguientes apartados podremos apreciar como el reconocimiento de emociones apoyado por las tecnologías previamente explicadas pueden brindar apoyo a profesionistas en distintas áreas laborales.

Reconocimiento de emociones en audiencias

¿Cuántos conferencistas, docentes y expositores han querido generar una impacto cada vez mayor en sus audiencias al momento de hablar? Comúnmente se puede hablar de un ambiente se siente agradable o “pesado” y que tanto el rostro como la postura postura del público dicen más que 1000 palabras. Sin embargo, dejando a un lado las corazonadas ¿hay alguna forma de conocer lo que está sintiendo la audiencia en ese momento?¿Podrá la inteligencia artificial ayudarnos en esta misión?

En el artículo escrito por Janith Kodithuwakku, Dilki Dandeniya Arachchi y Jay Rajasekera titulado “*An Emotion and Attention Recognition System to Classify the Level of Engagement to a Video Conversation by Participants in Real Time Using Machine Learning Models and Utilizing a Neural Accelerator Chip*”, nos ejemplifican que es posible llegar a conclusiones cada vez más profundas cada que analizamos la emociones que manifiesta nuestra audiencia[20].

Las propuestas de aplicación por parte de los autores son de sumo interés debido a que pueden ayudar a mejorar el aprendizaje de cualquier individuo si este se suma a colaborar en el proceso, partiendo desde la simple toma de asistencia en videollamada, la realización de actividades que ayuden a ponerlo “en sintonía” con la emociones que le ayudarán a aprender y enfocarse mejor en los mensajes emitidos por el conferencista, etc.

Reconocimiento de emociones en home office

El reconocimiento facial para detección de emociones puede apoyar a las empresas para saber un estado de ánimo aproximado de los empleados y con ello poder apoyarlos a estar en un estado más integral, al mismo tiempo que se busque el aumentar la eficacia de los recursos humanos.

Si bien este proyecto no tiene un enfoque clínico, no descartamos su aplicación en el monitoreo de emociones en pacientes que están llevando alguna especie de tratamiento relacionado con problemas de ansiedad, falta de atención, depresión, etc. Siempre y cuando estos sistemas no resulten invasivos y aseguren la privacidad de la información brindada en pro de él o los pacientes que estén siendo parte del proceso de sanación.

Conclusión

Los sistemas de reconocimiento facial y sus aplicaciones como lo son la detección de emociones en este caso, pueden ser muy útiles para poder empatizar con una audiencia o mejorar el estado emocional de la gente. Las aplicaciones pueden ir desde un sondeo para saber el bienestar de la gente, hasta aplicaciones para análisis de las mismas y aumentar la recepción de la gente ante ciertos temas.

Las áreas de mejora que han sido consideradas van desde la selección de un dataset más rico en resolución y clases de emociones, hasta en el análisis de métricas en tiempo de ejecución de la aplicación, ya que entre los profesionistas de TI es bien sabido que python tiene un tiempo ejecución relativamente más lento que otras tecnologías como C++, Php, javascript, golang, etc.

Por otro lado, los autores de dicho proyecto somos conscientes de que para obtener mayor eficacia en el reconocimiento de emociones en nosotros los seres humanos se necesitan más que un conjunto de imágenes, debido al hecho de que se pueden incorporar métricas de sustancias secretadas por el cuerpo tales como adrenalina, cortisol, serotonina, oxitocina, noradrenalina, dopamina, entre otras. Las cuales nos darán información más valiosa sobre el estado interno de algún individuo y proseguir con un enfoque más clínico ya sea médico o psicológico para que profesionales de la salud acompañen a las personas en su proceso de curación, readaptación, transformación, etc.

Finalmente, vale la pena reflexionar sobre las clasificaciones que hacen los sistemas de inteligencia artificial, todo esto debido a que los sesgos que existen en nuestra sociedad salen a relucir en aplicaciones que predicen el porcentaje de peligrosidad de una persona con sólo analizar su rostro o aquellos sistemas a los que les pides imágenes de chefs y desafortunadamente el género femenio de raza clara es el que predomina en dichas consultas, demostrando así que el desarrollo de la tecnología debe que ir a la par del desarrollo de las leyes y el involucramiento de más áreas como lo son las ciencias sociales para así asegurar un manejo más armonioso de estas tecnologías de vanguardia en pro de la humanidad.

Referencias

[1]S. Russell and P. Norvig, Artificial intelligence, 4th ed. 2020.
[2]Chollet, F., 2022. Deep Learning With Python. 2nd ed. Greenwich, USA: Manning Publications.
[3]"Neural Networks: Chapter 6 - Neural Architectures", Chronicles of AI, 2022. [Online]. Available: https://chroniclesofai.com/neural-networks-chapter-6-neural-architectures/. [Accessed: 20- May- 2022].
[4]K. Team, “Keras: the Python deep learning API,” *Keras.io*, 2022. https://keras.io/ (accessed May 27, 2022).
[5]Wikipedia Contributors, “Keras,” *Wikipedia*, Apr. 05, 2022. https://en.wikipedia.org/wiki/Keras (accessed May 29, 2022).

[6]Wikipedia Contributors, “PyTorch,” *Wikipedia*, May 23, 2022. https://en.wikipedia.org/wiki/PyTorch (accessed May 29, 2022).
[7]“About - OpenCV,” *OpenCV*, Nov. 04, 2020. https://opencv.org/about/ (accessed May 29, 2022).
[8]Wikipedia Contributors, “scikit-learn,” *Wikipedia*, Jan. 14, 2022. https://en.wikipedia.org/wiki/Scikit-learn (accessed May 29, 2022).
[9]“Caffe | Deep Learning Framework,” *Berkeleyvision.org*, 2012. http://caffe.berkeleyvision.org/ (accessed May 29, 2022).
[10]“TensorFlow,” TensorFlow, 2022. https://www.tensorflow.org/ (accessed May 27, 2022).
[11]Wikipedia Contributors, “TensorFlow,” Wikipedia, May 01, 2022. https://en.wikipedia.org/wiki/TensorFlow (accessed May 27, 2022).
[12]chrisbasoglu, “The Microsoft Cognitive Toolkit - Cognitive Toolkit - CNTK,” *Microsoft.com*, Feb. 16, 2022. https://docs.microsoft.com/en-us/cognitive-toolkit/ (accessed May 27, 2022).
[13]C. de, “clase de las redes neuronales profundas, más comúnmente aplicada al análisis de imágenes visuales,” *Wikipedia.org*, Jun. 23, 2014. https://es.wikipedia.org/wiki/Red_neuronal_convolucional (accessed May 31, 2022).
[14]M. Basavarajaiah, "6 basic things to know about Convolution", Medium, 2022. [Online]. Available: https://medium.com/@bdhuma/6-basic-things-to-know-about-convolution-daef5e1bc411. [Accessed: 06- Jun- 2022].
[15]A. Kaushik, "Understanding the VGG19 Architecture", OpenGenus IQ: Computing Expertise & Legacy, 2022. [Online]. Available: https://iq.opengenus.org/vgg19-architecture/. [Accessed: 21- May- 2022].
[16]R. Alake, "Deep Learning: GoogLeNet Explained", Medium, 2022. [Online]. Available: https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765. [Accessed: 06- Jun- 2022].
[17]P. Ruiz, "Understanding and visualizing ResNets", Medium, 2022. [Online]. Available: https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8. [Accessed: 06- Jun- 2022].
[18]J. Oeix, “Face expression recognition dataset”. https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset
[19]A. Sharma, "Emotion Detector using Keras - with source code - easiest way - easy implementation - 2022 - Machine Learning Projects", Machine Learning Projects, 2022. [Online]. Available: https://machinelearningprojects.net/emotion-detector-using-keras/. [Accessed: 13- Jun- 2022].
[20]J. Kodithuwakku, D. Arachchi and J. Rajasekera, "An Emotion and Attention Recognition System to Classify the Level of Engagement to a Video Conversation by Participants in Real Time Using Machine Learning Models and Utilizing a Neural Accelerator Chip", Algorithms, vol. 15, no. 5, p. 150, 2022. Available: https://www.mdpi.com/1999-4893/15/5/150. [Accessed 13 June 2022].