

## **Investigue e indique en qué casos son útiles las siguientes arquitecturas, agregue imagenes si esto le ayuda a una mejor comprensión**

### **a. GoogleNet (Inception)**

GoogleNet, o Inception, revolucionó el campo de la visión por computadora al ganar el concurso ImageNet en 2014. Su diseño innovador introduce "módulos de Inception", que son combinaciones de diferentes operaciones de convolución y pooling ejecutadas en paralelo. Estos módulos permiten que la red capture características en múltiples escalas, haciendo de Inception una opción poderosa para tareas de reconocimiento visual en conjuntos de datos grandes.

### **b. DenseNet (Densely Connected Convolutional Networks)**

DenseNet es una arquitectura caracterizada por sus conexiones densas entre las capas, donde cada capa recibe entradas de todas las capas anteriores. Esta estructura de conexión densa mejora el flujo de información y gradientes a través de la red, y permite un uso más eficiente de los parámetros. DenseNet es particularmente útil en situaciones donde la eficiencia de la memoria es crítica y en conjuntos de datos con recursos limitados.

### **c. MobileNet**

Diseñado para ser ligero y eficiente, MobileNet es ideal para dispositivos con restricciones de cómputo como teléfonos móviles y dispositivos embebidos. Esta arquitectura logra su eficiencia mediante el uso de convoluciones separables en profundidad, que reducen significativamente la cantidad de cálculos necesarios. Aunque puede haber un ligero compromiso en la precisión, su rapidez y tamaño reducido lo hacen perfecto para aplicaciones en tiempo real en dispositivos móviles.

### **d. EfficientNet**

EfficientNet destaca por su método de escalado compuesto que determina cómo crecer la red, equilibrando ancho, profundidad y resolución. Esta estrategia permite que EfficientNet alcance niveles de precisión comparables a otros modelos más grandes, pero con un número significativamente menor de parámetros. Es una excelente opción cuando se busca alta precisión con un modelo eficiente en términos de tamaño y cómputo.

## **¿Cómo la arquitectura de transformers puede ser usada para image recognition?**

Vision Transformer (ViT) es un ejemplo de una arquitectura que usa esta idea y ha demostrado tener un rendimiento comparable o incluso mejor que las arquitecturas CNN tradicionales en conjuntos de datos grandes como ImageNet.

Es importante notar que los transformers requieren grandes cantidades de datos y cómputo para entrenarse desde cero, pero han demostrado ser altamente efectivos cuando se preentrenan en grandes conjuntos de datos y luego se afinan en conjuntos de datos más pequeños

## Referencias

VisibleBreadcrumbs. (2023). Mathworks.com.  
<https://es.mathworks.com/help/deeplearning/ref/googlenet.html>

Papers with Code - DenseNet Explained. (2020). Paperswithcode.com.  
<https://paperswithcode.com/method/densenet>

VisibleBreadcrumbs. (2023). Mathworks.com.  
<https://es.mathworks.com/help/deeplearning/ref/mobilenetv2.html>

Papers with Code - EfficientNet Explained. (2020). Paperswithcode.com.  
<https://paperswithcode.com/method/efficientnet>