**MobileNetV2** es una arquitectura de red neuronal convolucional (CNN) diseñada específicamente para dispositivos móviles y sistemas con recursos limitados. Fue introducida por Google en 2018 como una mejora de la arquitectura original MobileNet, optimizando aún más su eficiencia en términos de velocidad y consumo de recursos, sin sacrificar significativamente el rendimiento.

Principales características de MobileNetV2:

1. **Bloques de convolución inversa residual ("Inverted Residual Blocks")**
   * A diferencia de las redes tradicionales que expanden características y luego las contraen, MobileNetV2 utiliza bloques que **comprimen características primero y luego las expanden**. Esto mejora la eficiencia computacional.
   * Cada bloque incluye un **atajo residual**, lo que ayuda a preservar información importante durante el flujo de datos.
2. **Convoluciones separables en profundidad ("Depthwise Separable Convolutions")**
   * Estas convoluciones descomponen el proceso en dos pasos:
     + Una convolución que opera independientemente en cada canal (profundidad).
     + Una convolución estándar que mezcla las características entre canales.
   * Este enfoque reduce drásticamente el número de operaciones y parámetros necesarios.
3. **Bajo consumo de memoria y cálculo**
   * Está diseñada para trabajar en dispositivos móviles o integrados con recursos limitados (baja memoria y potencia de cálculo), lo que la hace ideal para aplicaciones en tiempo real, como clasificación de imágenes, detección de objetos y segmentación semántica.
4. **ReLU6 y activación lineal**
   * Utiliza **ReLU6**, una variante de ReLU que limita los valores máximos a 6, para mejorar la estabilidad en dispositivos con precisión de punto flotante limitada.
   * En capas de baja dimensión, utiliza activaciones lineales para minimizar la pérdida de información.

Ventajas de MobileNetV2:

* **Eficiencia:** Menor cantidad de operaciones computacionales y parámetros en comparación con modelos más grandes como ResNet o VGG.
* **Velocidad:** Adecuada para aplicaciones en tiempo real.
* **Escalabilidad:** Puede ajustarse mediante un parámetro llamado *width multiplier*, que permite controlar el número de canales en las convoluciones para equilibrar precisión y eficiencia.

Aplicaciones comunes:

* Clasificación de imágenes.
* Detección de objetos (por ejemplo, en sistemas de cámaras inteligentes).
* Reconocimiento facial.
* Sistemas de visión por computadora en dispositivos móviles y embebidos.