

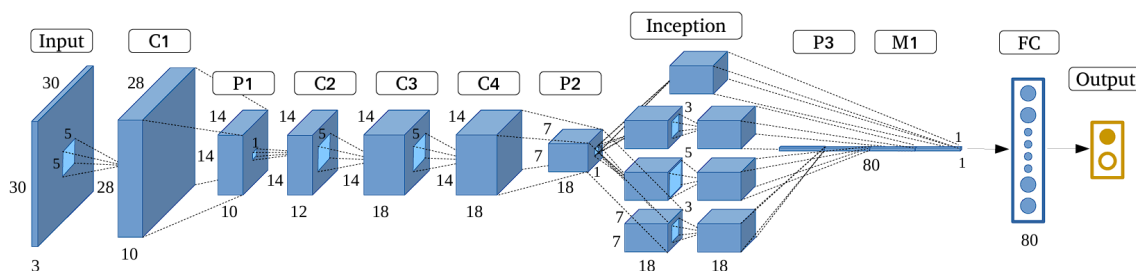
## Deep Learning Laboratorio 8

Task 2 - Teoría Responda claramente y con una extensión adecuada las siguientes preguntas:

1. Investigue e indique en qué casos son útiles las siguientes arquitecturas, agregue imagenes si esto le ayuda a una mejor comprensión
  - a. GoogleNet (Inception)

Esta es bastante útil si se necesita un modelo eficiente y potente con un alto rendimiento en la clasificación de imágenes y detección de objetos.

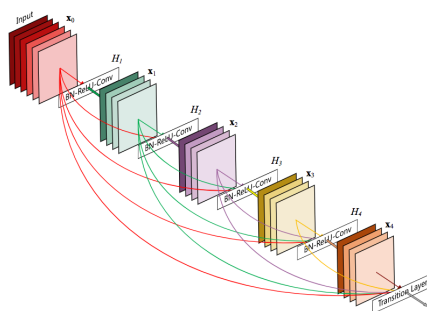
Este utiliza módulos de Inception, lo que permite el procesamiento simultáneo de convoluciones de diferentes tamaños y max-pooling, reduciendo el costo computacional. Esta se puede utilizar en la clasificación de imágenes, reconocimiento de objetos, segmentación, entre otras.



- b. DenseNet (Densely Connected Convolutional Networks)

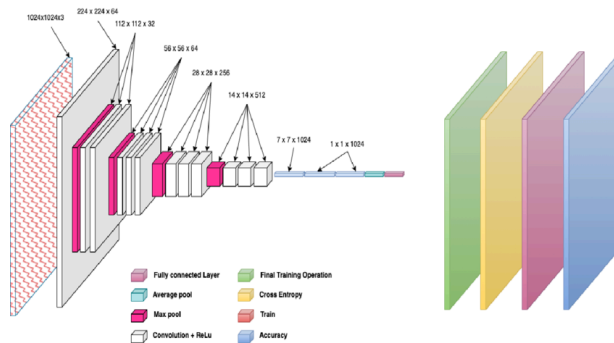
Esta es útil cuando se necesita mejorar el flujo de gradiente en redes profundas, cada capa recibe como entrada las salidas de todas las capas anteriores, promoviendo el reutilizamiento de características y reduciendo el número de parámetros con las arquitecturas tradicionales.

Se utiliza en la clasificación de imágenes, detección de anomalías, problemas médicos, entre otras.



### c. MobileNet

Esta está diseñada para aplicaciones móviles o dispositivos con recursos limitados, ya que este es bastante liviana, usa convoluciones separables en profundidad, lo que reduce el número de operaciones y parámetros sin comprometer mucho el rendimiento. Se utiliza en reconocimiento facial, clasificación de imágenes en teléfonos inteligentes, aplicaciones de realidad aumentada, entre otros.



### d. EfficientNet

Este es un modelo bastante preciso y eficiente, ya que introduce una metodología para escalar modelos de manera eficiente mediante la combinación de la profundidad, la resolución y el ancho de las capas de una manera balanceada. Se utiliza en clasificación de imágenes, detección de objetos, aplicaciones industriales con una gran demanda de precisión y eficiencia.

#### EfficientNet Architecture



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

La arquitectura de transformers se adaptó para la visión por medio del modelo "Vision Transformer". En lugar de procesar secuencias de palabras como en NLP, ViT divide una imagen en parches los cuales son similares a los tokens y trata cada parche como una entrada secuencial. Luego, aplica autoatención para capturar relaciones entre los diferentes parches.