Redes convolucionales avanzadas II

Visión por Computador, curso 2024-2025

Silvia Martín Suazo, silvia.martin@u-tad.com

21 de noviembre de 2024

U-tad | Centro Universitario de Tecnología y Arte Digital

Residual Network

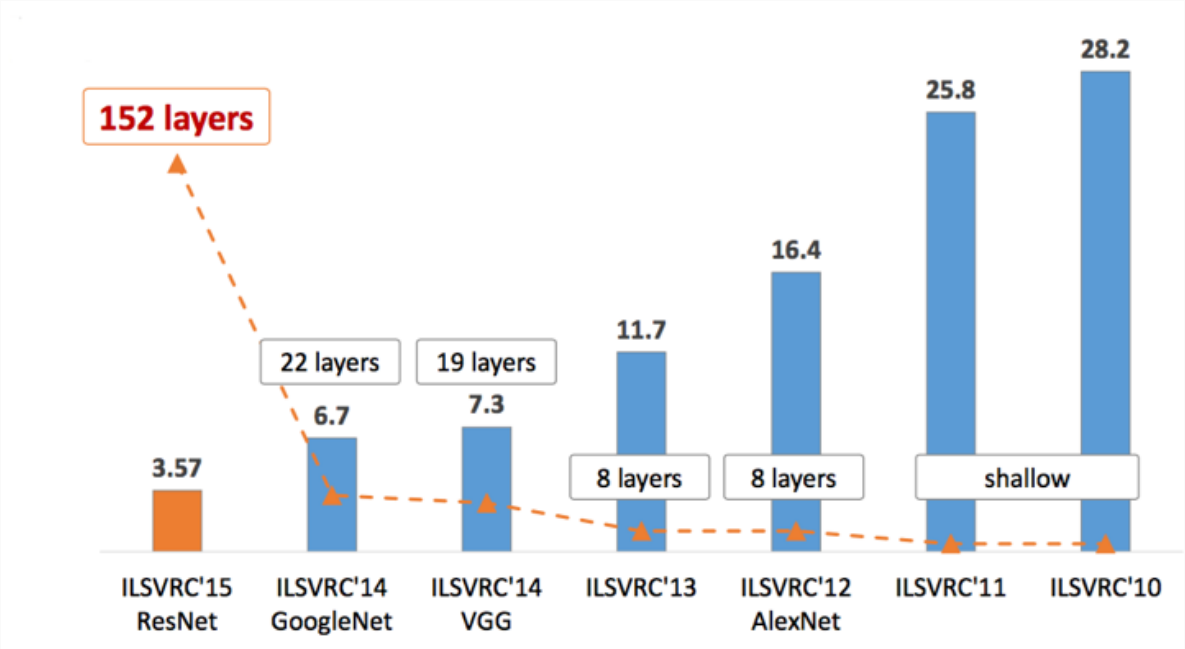
Después del desarrollo de las redes AlexNet[1], Visual Geometry Group (VGG)[2] e Inception[3] la investigación en Convolutional Neural Networks (CNNs) se vió especialmente influida por la profundidad de las redes como factor diferenciador.

Dentro de este afán por más capas la arquitectura Residual Network, también conocida como ResNet, supuso el mayor avance desde la publicación de AlexNet.

ResNet

Dentro de la competición ilsvrc, después de los grandes avances de AlexNet, el siguiente hito más importante fue la introducción de la ResNet.

La arquitectura ResNet consigue obtener una puntuación del Top-5 error de 3.57%, superando incluso al ser humano, con puntuación de 5%.

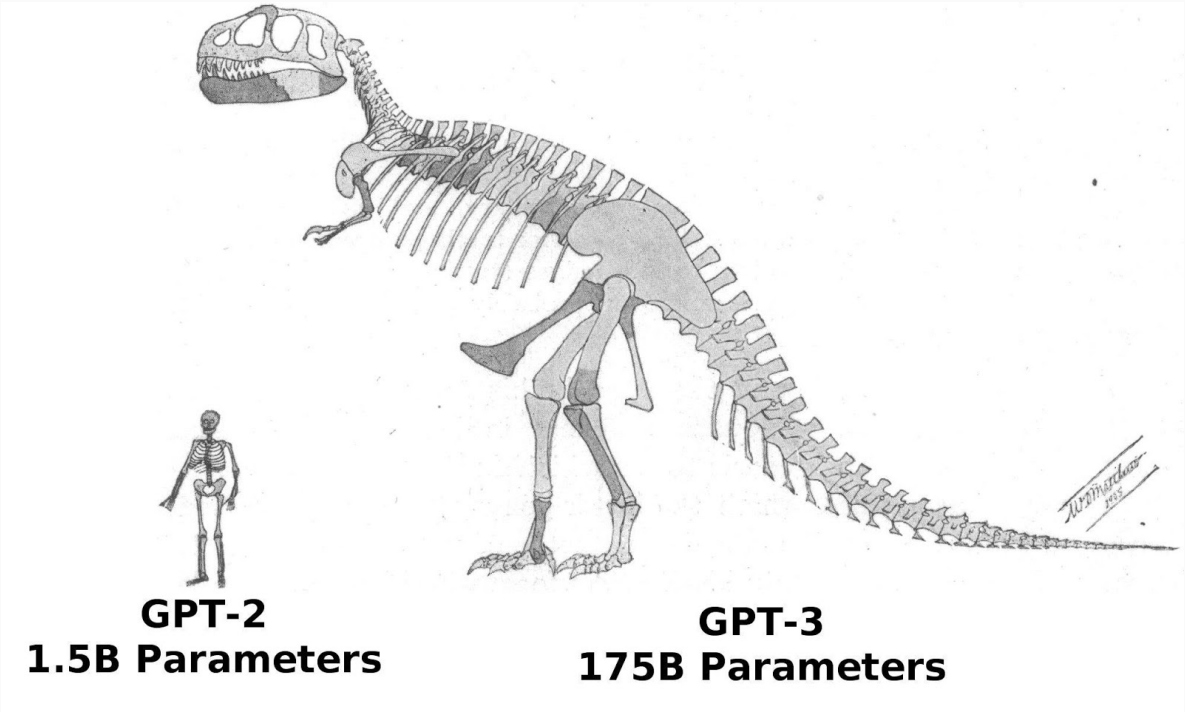


[4]

ResNet

El principal avance que propone la arquitectura ResNet es poder aumentar la profundidad de las redes a cientos e incluso miles de capas sin comprometer el rendimiento de las mismas.

Esta tendencia a aumentar las dimensiones de las redes continua hoy en día con casos como GPT-3[5].



[6]

ResNet

Anteriormente a la publicación de la ResNet existían ciertos mecanismos para evitar el problema del gradient vanishing, como por ejemplo, añadir una pérdida intermedia. Pero ninguna de las soluciones parecía solucionar el problema definitivamente.

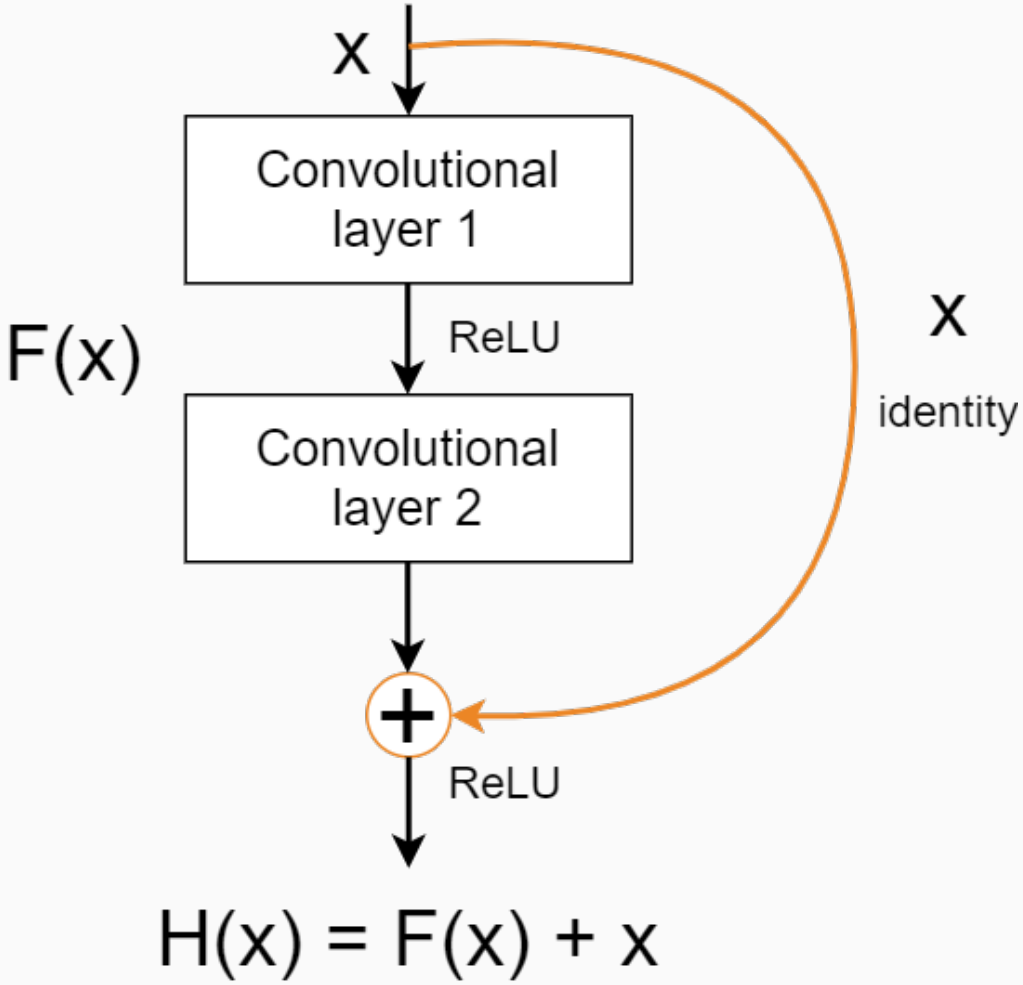
La principal idea que propone la arquitectura es la introducción de los saltos entre capas, con las conocidas como skip connections.

Con ello se pretende:

- Aumentar el número de capas.
- Que la red no pueda tener mayor error con menor número de capas.

ResNet: El bloque residual

El bloque residual de la ResNet introduce las skip connections en el conocido como "identity shortcut connection".



ResNet: El bloque residual

El bloque residual se compone por una skip connection a través de la cual se realiza un bypass a la información de entrada (x). Por otra parte la información procesada por las capas convolucionales ($F(x)$) es transmitida de manera habitual.

Ambas salidas son combinadas a través de una suma elemento a elemento de sus componentes.

Esta nueva salida se define como:

$$H(x) = F(x) + x \tag{1}$$

ResNet: El bloque residual

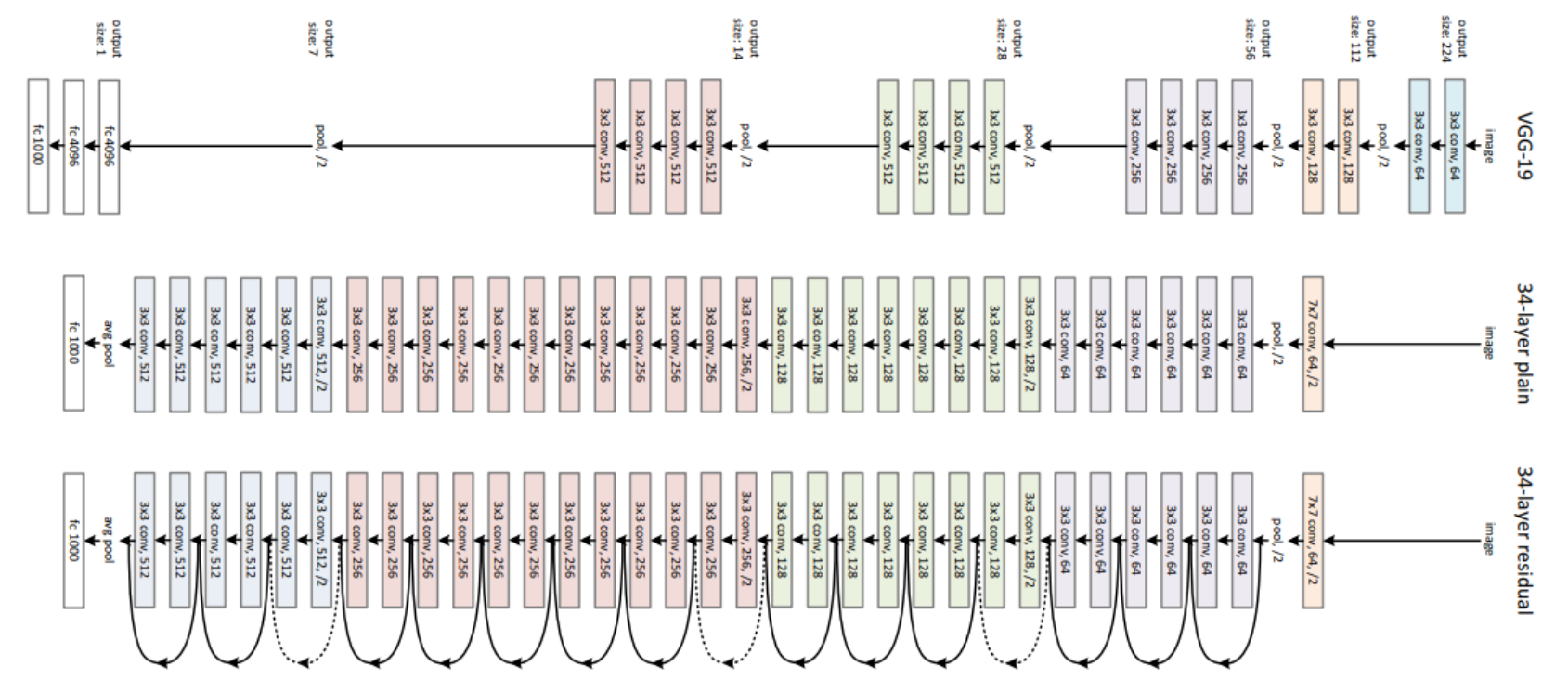
El bloque residual obtiene su nombre ya que se encarga de aprender el mapeo residual $F(x) = H(x) - x$, a través del cual extrae información relevante de la imágen recibida.

Al mismo tiempo la información identidad permite viajar a la información evitando problemas derivados del gradiente.

Esta arquitectura permite que si alguna capa daña el rendimiento de la red esta pueda ser saltada a través de la identidad. Esto funciona como un mecanismo de regularización.

ResNet: La arquitectura

Estos nuevos bloques se utilizan para generar una arquitectura de mayor profundidad.

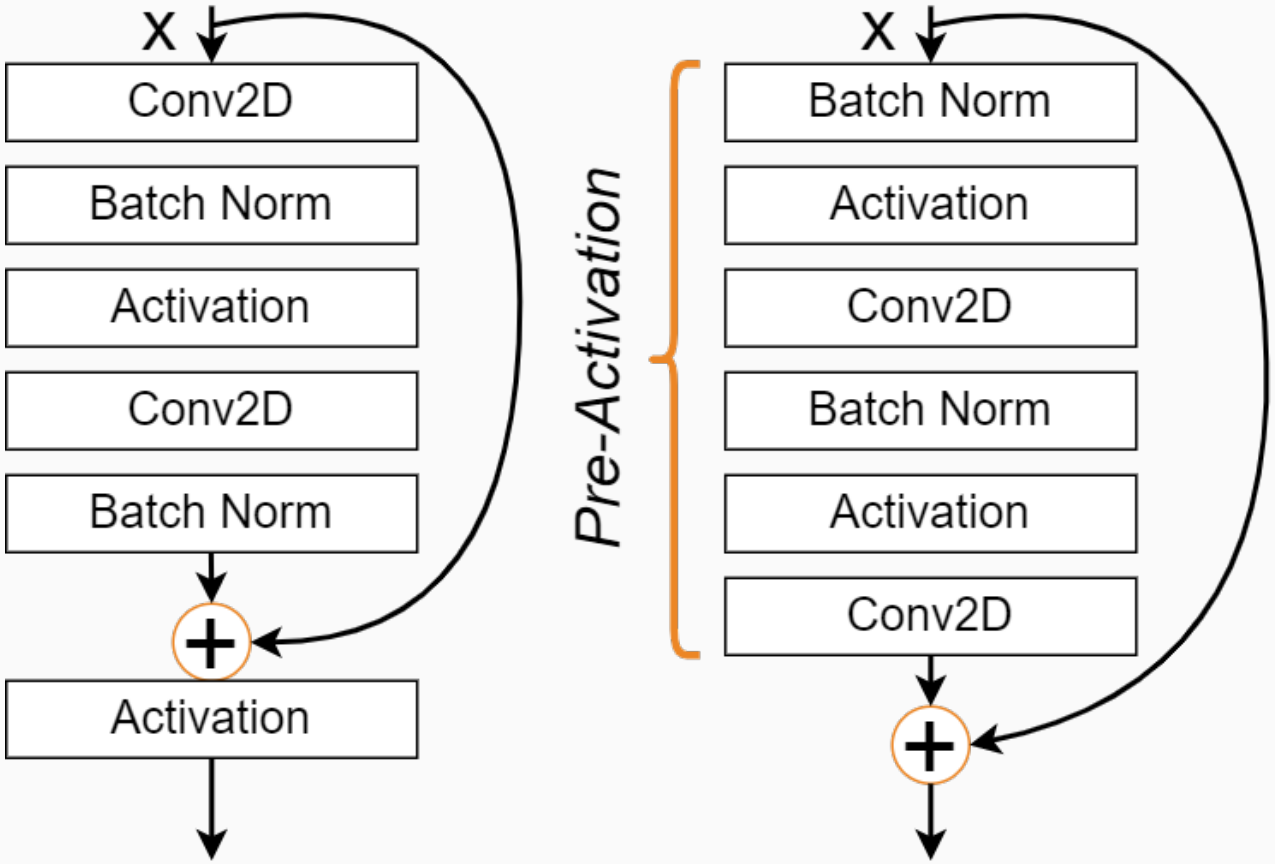


[7]

Variantes de la ResNet

Siguiendo la misma intuición, existen distintas variantes del bloque residual.

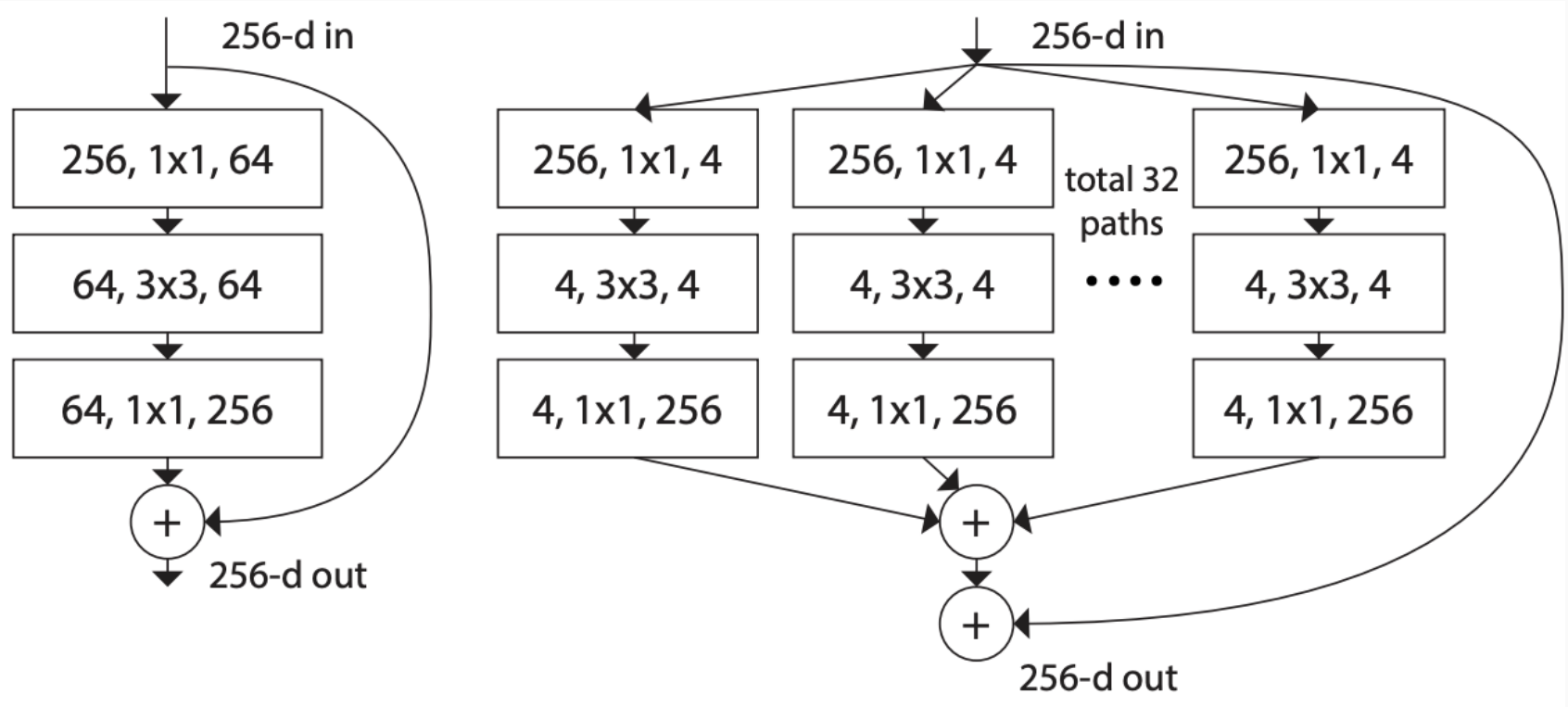
Por ejemplo la variación con pre-activación de las capas permite a los gradientes viajar a través de los "atajos" de manera libre.



ResNext

La arquitectura ResNext[8] introduce un nuevo hiperparámetro llamado cardinalidad, el cual es usado para controlar la capaidad del modelo.

Esta arquitectura esta fuertemente influenciada por la idea de crecer en amplitud presentada por la arquitectura Inception.

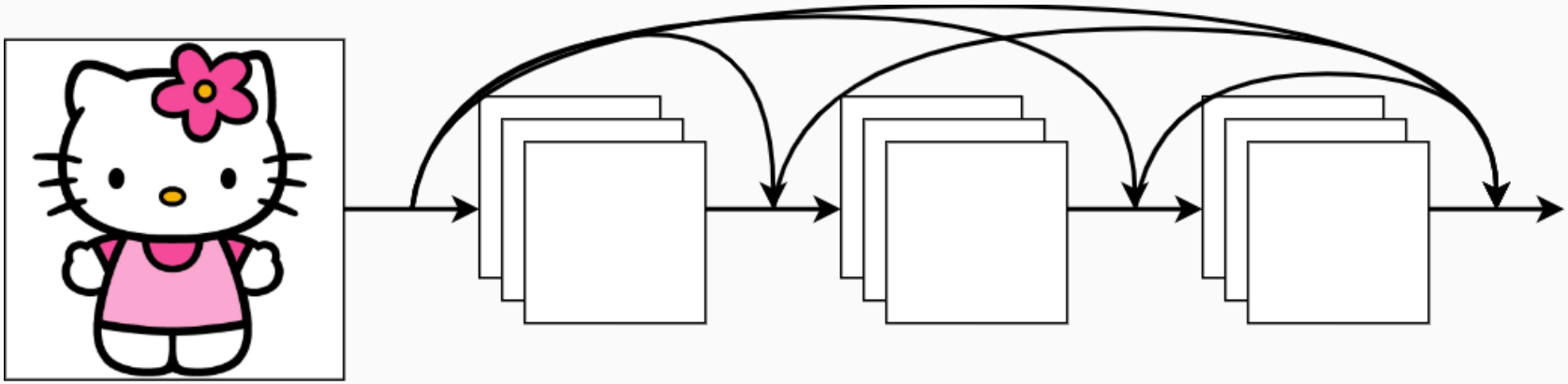


[8]

DenseNet

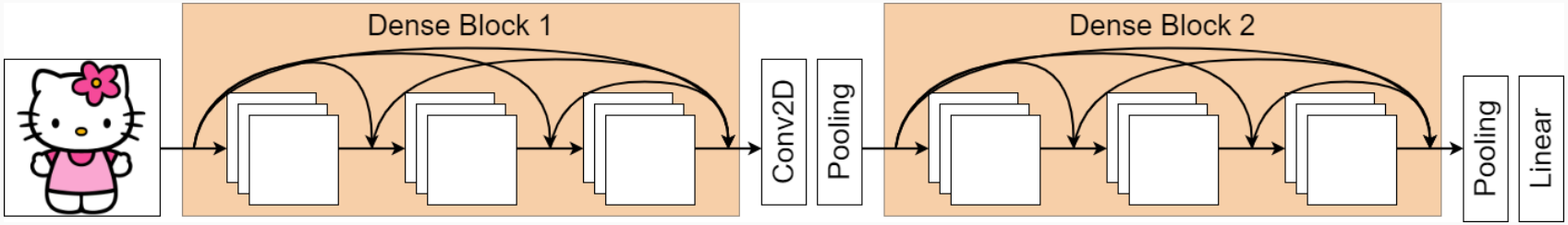
La arquitectura DenseNet[9] explora aún más las posibilidades de las skip connections. Aprovechando las buenas propiedades descritas anteriormente.

La principal aportación es que cada bloque convolucional se conecte a todos los posteriores.



DenseNet

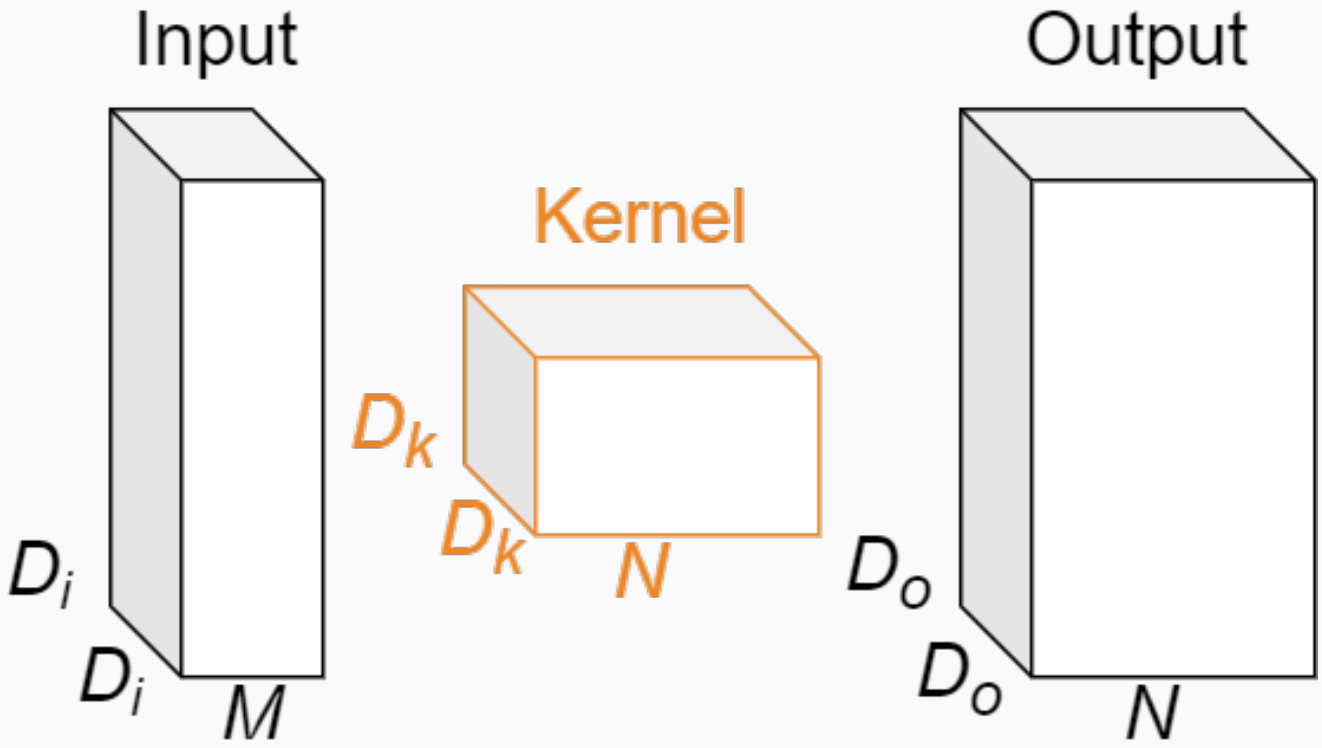
Una de las principales ventajas de esta arquitectura es su eficiencia al utilizar la información extraída de diversas capas al mismo tiempo.



Propiedades de la arquitectura DenseNet

- Transporte del gradiente eficiente: Haciendo uso de las skip connections hacia capas anteriores.
- Eficiencia computacional: El escalado en el número de parámetros es mucho menor que en la ResNet.
- Mayor riqueza en las características recibidas: Cada capa puede obtener información desde todas las precedentes a ella.
- Mayor variedad en las características recibidas: Al recibir información desde todas las capas, el clasificador obtiene características a bajo, medio y alto nivel de la imagen.

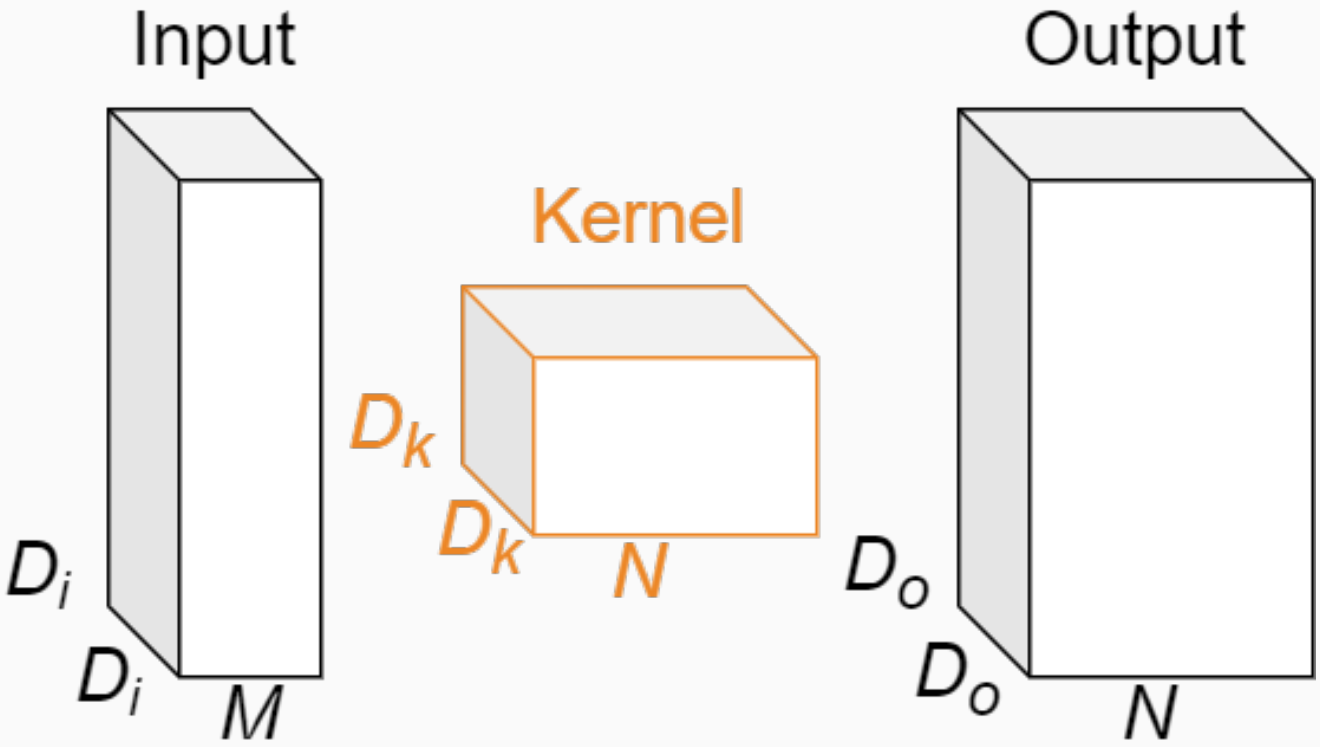
Convoluciones separables



Una convolución normal tiene los siguientes parámetros:

- Datos de entrada: Imagen de dimensiones $D_i x D_i$ y M canales.
- Filtros: De tamaño $D_k x D_k$ y N filtros.
- Datos de salida: De tamaño $D_o x D_o$ y N canales.

Convoluciones separables



Una convolución normal tiene los siguientes parámetros:

- Número de multiplicaciones: $D_k^2 \times M \times D_o^2 \times N$
- Número de parámetros: $D_k^2 \times M \times N$

Convoluciones separables

Las convoluciones separables o depthwise separable convolutions son una especie de convolución cuyo objetivo es obtener mayor eficiencia en cuanto a número de parámetros y operaciones.

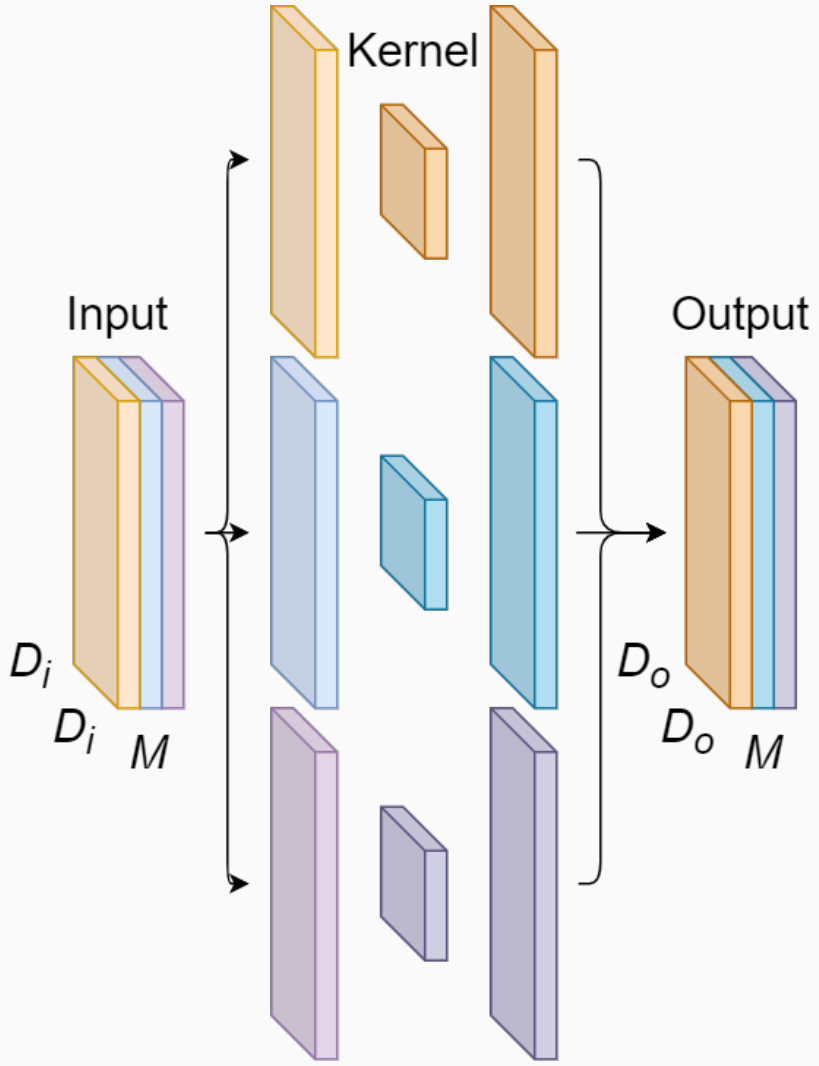
Para ello la operación de convolución se divide en dos pasos:

- Convolución en profundidad (*depthwise convolution*): Introduce un filtro de tamaño $D_k x D_k$ por cada canal de entrada.
- Convolución en punto (*pointwise convolution*): Introduce N filtros de tamaño $1x1$.

Convoluciones en profundidad:

- Número de multiplicaciones:
  $D_k^2 \; x \; D_o^2 \; x \; M$
- Número de parámetros:
  $D_k^2 \; x \; M$

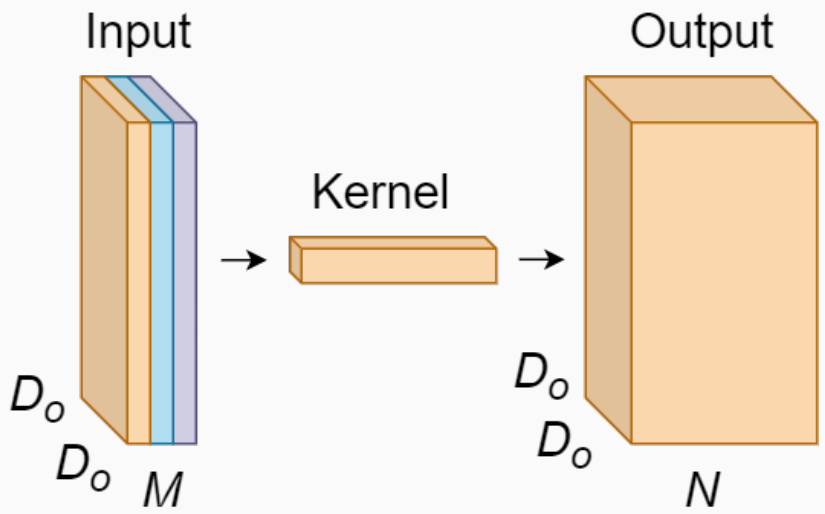
Se utilizan filtros de tamaño nxnx1. De tal manera que cada filtro procesa la información de un canal de entrada.



Convoluciones separables

Convoluciones en punto:

- Número de multiplicaciones:
  $D_o^2 \; x \; M \; x \; N$
- Número de parámetros:
  $M \; x \; N$



Se utilizan filtros de tamaño 1x1xN. De tal manera que la imagen de salida tiene el número de canales deseados.

Convoluciones separables

Por lo tanto el número final de operaciones y parámetros es el siguiente:

- Número de multiplicaciones: $D_o^2 \; x \; M \; x \; (D_k^2 \; + \; N)$

- Número de parámetros: $M \; x \; (D_k^2 \; + \; N)$

Estos números suponen aproximadamente un 10% respecto a las convoluciones estándar.

Convoluciones separables

Al ser convoluciones más eficientes computacionalmente, son generalmente utilizadas en redes enfocadas a aplicaciones que necesitan de esta eficiencia.

- MobileNet: redes utilizadas en aplicaciones móviles, y que necesitan de modelos ligeros y eficientes.
- Xception: es una variante de las arquitecturas Inception vistas anteriormente, que buscan la eficiencia vía convoluciones separables.
- EfficientNet: red neuronal que busca el equilibrio precisión-eficiencia mediante convoluciones separables y coeficientes de escala.
- ShuffleNet: red que contiene bloques *shuffle* que consisten en una convolución en profundidad, una operación *shuffle* y una convolución puntual.

Referencias i

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton.
Imagenet classification with deep convolutional neural networks.
Communications of the ACM, 60(6):84–90, 2017.

[2] Karen Simonyan and Andrew Zisserman.
Very deep convolutional networks for large-scale image recognition.
arXiv preprint arXiv:1409.1556, 2014.

Referencias ii

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich.
Going deeper with convolutions.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[4] Opengenus.
Ilsvrc results image.
[Online; accessed September, 2022].

Referencias iii

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al.
Language models are few-shot learners.
Advances in neural information processing systems, 33:1877–1901, 2020.

[6] Exxact.
Gpt-3 size image.
[Online; accessed September, 2022].

[7] Geeks for Geeks.
Comparison between resnet and vgg image.
[Online; accessed September, 2022].

Referencias iv

[8] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He.
Aggregated residual transformations for deep neural networks.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger.
Densely connected convolutional networks.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.