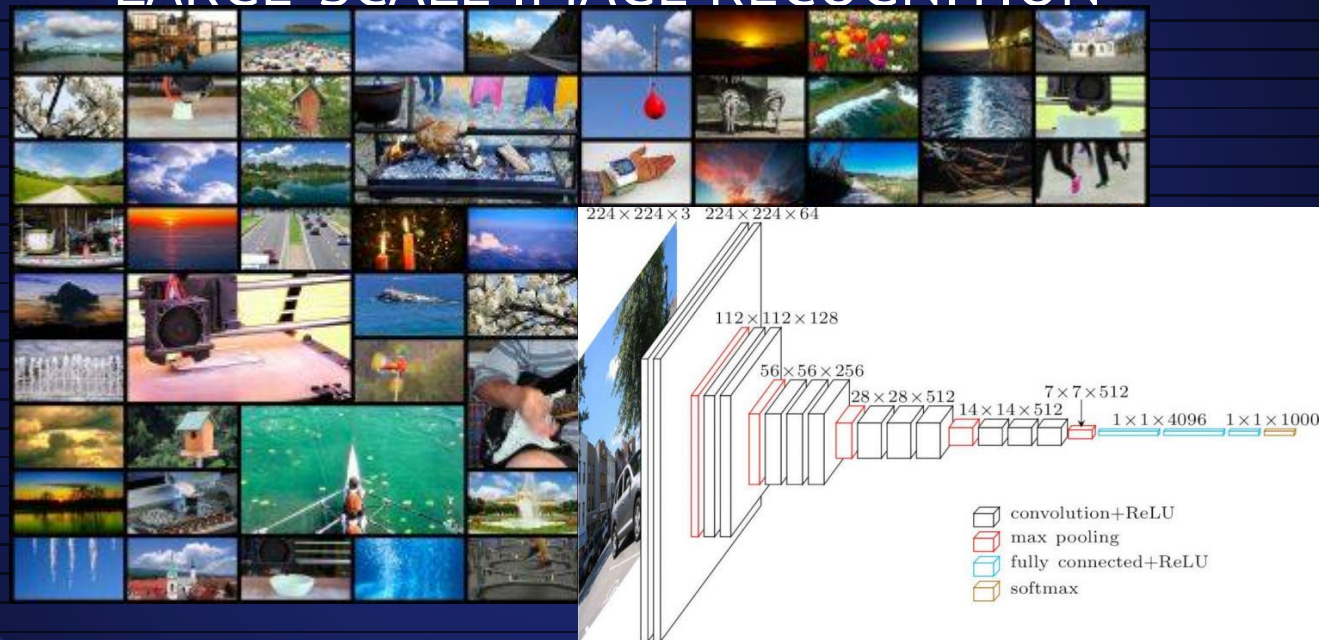


VGG - 16

Chipana Perez, Gabriela Angel
Vizcarra Vargas, Piero Emiliano
Hilares Angelo, Maryori Lizeth

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION



Karen Simonyan y Andrew Zisserman

Artículo de presentación

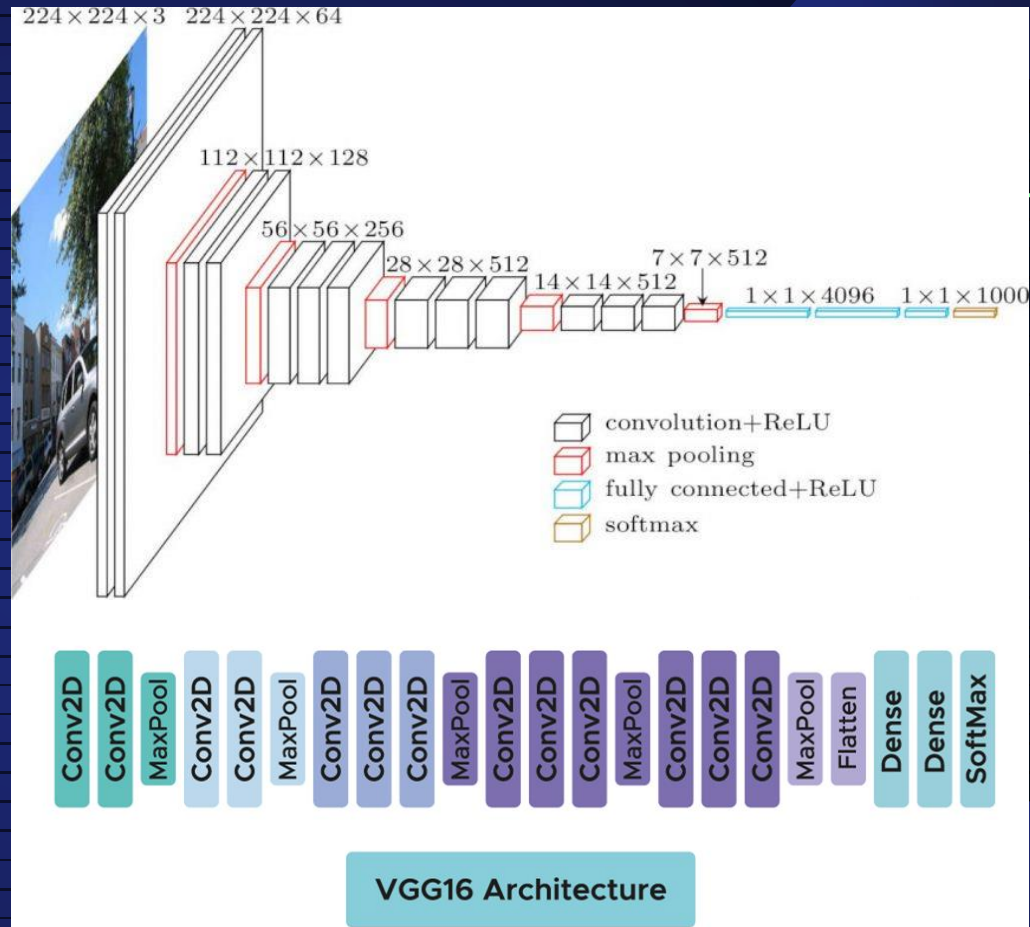
Very deep convolutional networks for
large-scale image recognition

Autores:
Karen Simonyan
Andrew Zisserman

- CNN alto rendimiento en el reconocimiento de imágenes a gran escala.
- Menciona la dificultad de entrenar redes más profundas y la necesidad de explorar arquitecturas que puedan manejar esta complejidad
- Objetivo: Presentar una arquitectura de CNN VGG-16 (y también VGG-19), que tiene una profundidad significativamente mayor que otras arquitecturas.
El objetivo es demostrar que una mayor profundidad de la red puede mejorar el rendimiento en la clasificación de imágenes en el conjunto de datos ImageNet.

Arquitectura/estructura

- **Capas de entrada:** imágenes de tamaño 224×224
- **Capas convolucionales:** filtros de tamaño 3×3 con un paso de 1 píxel .
Aplican la función de activación ReLU después de cada convolución.
Objetivo: extraer características relevantes de las imágenes en diferentes niveles de abstracción.
- **Capas de agrupación (pooling):** Se aplica cada 2 capas convolucionales, (Max Pooling) con un tamaño de ventana de 2×2 y un paso de 2 píxeles.
Objetivo: Reducir la dimensionalidad y ayuda a obtener una representación más compacta de las características extraídas.

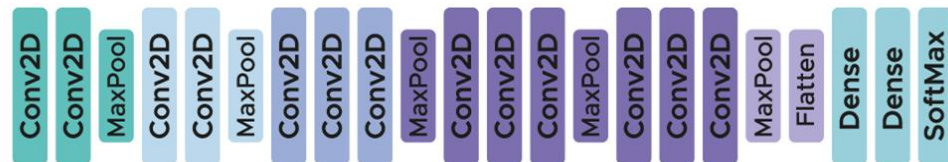
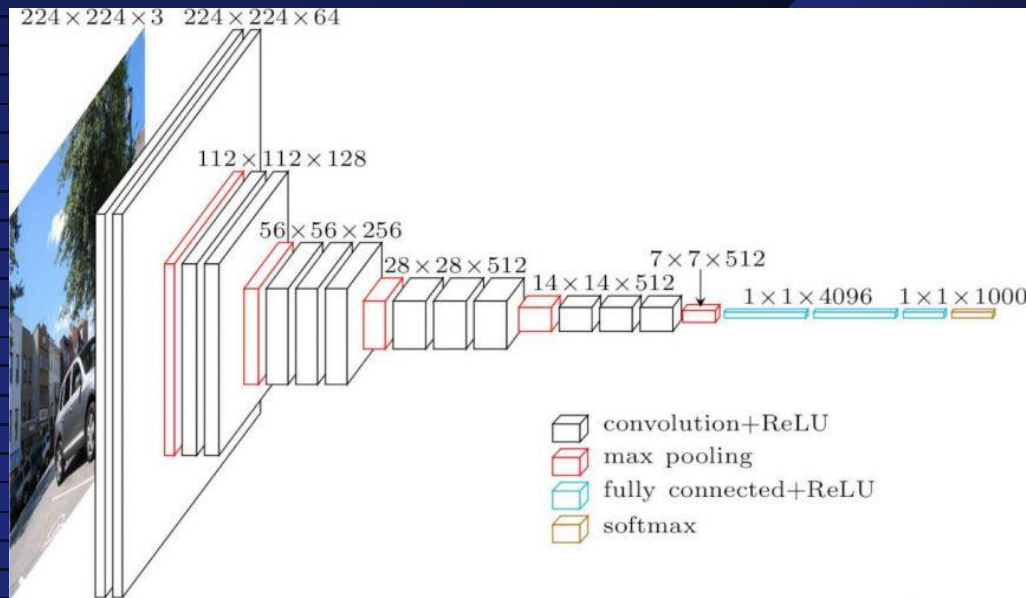


Arquitectura/estructura

- **Capas totalmente conectadas:** Cada capa con 4096 unidades.

Objetivo: transformar las características extraídas en una salida clasificada. Después de cada capa totalmente conectada, se aplica la función de activación ReLU, excepto en la última capa.

- **Capa de salida:** La capa de salida consiste en una capa totalmente conectada con 1000 unidades, correspondientes a las 1000 clases diferentes en el conjunto de datos ImageNet. La función de activación utilizada en esta capa es la función Softmax, que asigna probabilidades a cada clase.



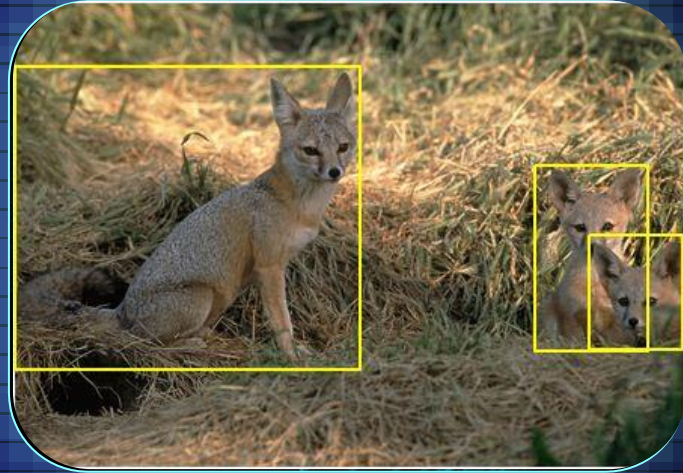
VGG16 Architecture

Datos

ILSVRC

Contiene los datos de imagen y la realidad del terreno para los conjuntos de entrenamiento y validación, y los datos de imagen para el conjunto de prueba.

Las anotaciones se ordenan por sus synsets (por ejemplo, "gato persa", "bicicleta de montaña" o "perro caliente") como su wnid. Estas identificaciones se ven como n00141669. El nombre de cada imagen tiene una correspondencia directa con el nombre del archivo de anotación.



ImageNet Large-Scale Visual
Recognition Challenge

Table 7: **Comparison with the state of the art in ILSVRC classification.** Our method is denoted as “VGG”. Only the results obtained without outside training data are reported.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	23.7	6.8	6.8
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-	7.9	
GoogLeNet (Szegedy et al., 2014) (7 nets)	-	6.7	
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	27.9	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-

Áreas de aplicación

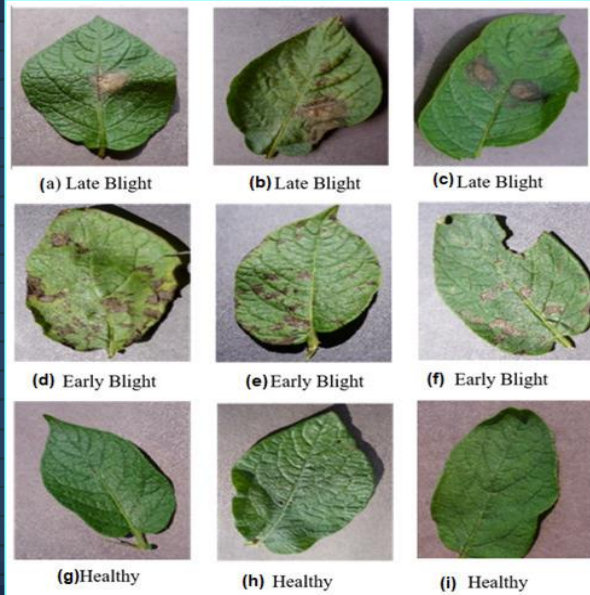
Clasificación de imágenes

VGG-16 ha sido ampliamente utilizada para tareas de clasificación de imágenes en diferentes dominios, como reconocimiento de objetos, detección de enfermedades en imágenes médicas, clasificación de escenas, etc.



Predicción de edad ósea con red
basada en VGG-16 y transfer
learning

Áreas de aplicación



Detección de objetos

Junto con técnicas de detección como R-CNN, Faster R-CNN y YOLO, la VGG-16 se ha utilizado para la detección de objetos en imágenes, lo que implica identificar la presencia y ubicación de objetos específicos en una imagen.

Application of convolutional neural networks for detection of the late blight

Phytophthora infestans in potato *Solanum tuberosum*

Áreas de aplicación

Recuperación de información visual

La VGG-16 se ha utilizado en aplicaciones de recuperación de información visual, donde el objetivo es buscar imágenes similares en grandes bases de datos de imágenes. La extracción de características de la VGG-16 permite representar las imágenes de manera compacta y compararlas eficientemente.

Segmentación semántica

La VGG-16 se ha utilizado en aplicaciones de segmentación semántica, donde el objetivo es asignar una etiqueta semántica a cada píxel de una imagen. Esta tarea es útil en aplicaciones como la conducción autónoma, el etiquetado automático de imágenes y la realidad aumentada.

The background is a dark blue gradient with various abstract digital elements. There are several green lines of different lengths and orientations, some ending in small circles, resembling circuit traces. There are also clusters of small white dots. On the left and right sides, there are blue arrow-like shapes pointing inwards. The word "Demo" is centered in a large, bold, white font.

Demo

Conclusiones

1. VGG-16 es una arquitectura de red convolucional popular y ampliamente utilizada en el campo de la visión por computadora. Fue desarrollada por el grupo de investigación Visual Geometry Group en la Universidad de Oxford.
2. La arquitectura VGG-16 se caracteriza por su simplicidad y eficacia. Está compuesta por 16 capas, incluyendo 13 capas convolucionales y 3 capas completamente conectadas.
3. VGG-16 ha demostrado un rendimiento impresionante en tareas de clasificación de imágenes, logrando resultados cercanos al estado del arte en conjuntos de datos desafiantes como ImageNet.
4. La estructura de VGG-16, con capas convolucionales en cascada, permite aprender características visuales jerárquicas de una imagen, desde características de bajo nivel (como bordes y texturas) hasta características de alto nivel (como formas y objetos).

Conclusiones

5. La capacidad de transferencia de aprendizaje de VGG-16 es una de sus características más destacadas. Al pre-entrenar la red en grandes conjuntos de datos, se pueden transferir los conocimientos aprendidos a tareas específicas con conjuntos de datos más pequeños, ahorrando tiempo y recursos de entrenamiento.
6. VGG-16 se ha utilizado en una variedad de aplicaciones, como clasificación de objetos, detección de objetos, segmentación semántica, recuperación de información visual, entre otras. Su capacidad para aprender características visuales complejas la hace valiosa en diversos problemas de visión por computadora.
7. Aunque VGG-16 ha sido una arquitectura influyente, también tiene algunas limitaciones. Es una red profunda y relativamente pesada, lo que puede dificultar su implementación en dispositivos con recursos limitados. Además, puede requerir conjuntos de datos grandes para un entrenamiento efectivo.

Conclusiones

En general, la arquitectura VGG-16 ha dejado una huella significativa en el campo de la visión por computadora. Su simplicidad, eficacia y capacidad de transferencia de aprendizaje la han convertido en una opción popular para muchas aplicaciones de procesamiento de imágenes y ha sentado las bases para el desarrollo de arquitecturas más avanzadas.