



Optimización en reconocimiento de imágenes para clasificación de productos

Redes Neuronales Profundas

Gonzalo Juarros, Catalina Nehmad Alché, Miguel Pérez Centeno, Joaquín
30 de noviembre de 2017

Departamento de Computación, Universidad de Buenos Aires

1. Objetivo
2. Problema
3. La red
4. Experimentación

Objetivo

Objetivo

Dada una red neuronal convolucional y un dataset, queremos determinar si hay alguna diferencia en la precisión de la clasificación usando estos dos métodos:

- Entrenar la red capa por capa con ese dataset y que aprenda los pesos desde cero (*fresh start*)
- Entrenar todas las capas en base a pesos ya aprendidos con otro dataset (*fine tuning*)
- Tomar los pesos aprendidos por la red en base a otro dataset y entrenar sólo la última capa (*feature extraction*)

Capa por capa

Es lo que vimos durante la materia: en base al dataset, la red aprende *features* de bajo nivel en las primeras capas y de alto nivel en las últimas. Ya conocemos muchísimos ejemplos de redes entrenadas de esta manera que dan **muy buenos resultados**.

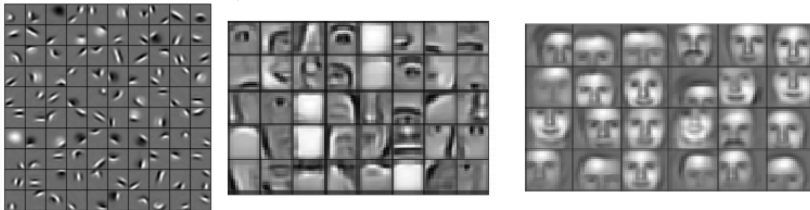


Figura 1: Features de distintos niveles de una red convolucional ¹

¹Andrew Ng et al, *Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks*. *Communications of the ACM*, vol. 54 (2011)

Pero...

Las redes convolucionales **tardan mucho en entrenarse**, ya que la operación de convolución es **computacionalmente costosa**. Con cada capa convolucional que agregamos, el tiempo de entrenamiento aumenta considerablemente.

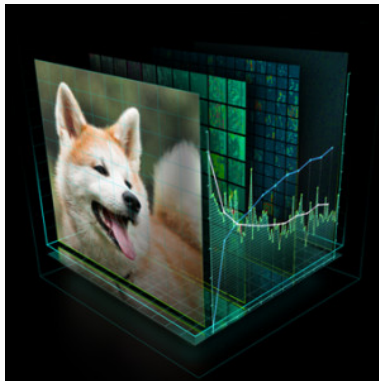


Figura 2: Este perro **no da más** de entrenar redes convolucionales.

Últimas capas - feature extraction

Podemos aprovechar lo que la red **ya aprendió** y usarlo para un nuevo problema de clasificación.

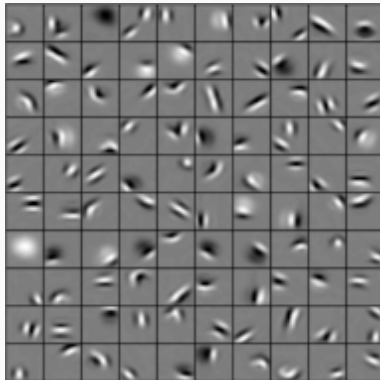


Figura 3: Los features detectados por las primeras capas de una red convolucional no son específicos de alguna clase particular de objetos.

Problema

Clasificación de objetos

Nuestro dataset consiste de imágenes de **bolígrafos**, **bolsos** y **dispositivos tecnológicos** que queremos clasificar adecuadamente.



Clasificación de objetos

Los *features* que vimos anteriormente podrían perfectamente formar parte de una imagen de un bolígrafo, un bolso o un dispositivo tecnológico, aunque hayan sido aprendidas en base a imágenes de otras categorías. ¿Podemos usarlas para aprender a identificar estos objetos? ¿Cambiará la precisión de la red si usamos *feature extraction* en vez de entrenar todas las capas?



- Los sitios de venta online actualmente **no clasifican según imágenes.**
- Poder categorizar los productos **sólo en base a las imágenes**, casi sin intervención, **sería mucho más práctico.**

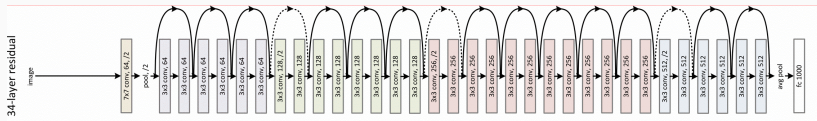
El problema de clasificación no es nuevo, pero tengamos en cuenta que sería prácticamente **imposible** entrenar una red neuronal desde cero para que reconozca exactamente los productos que necesitamos.

¡Para esto podemos usar **feature extraction**! ²

²¡Y si llama en los próximos 5 minutos, obtendrá una ResNet-18-Extractorator DE REGALO!

La red

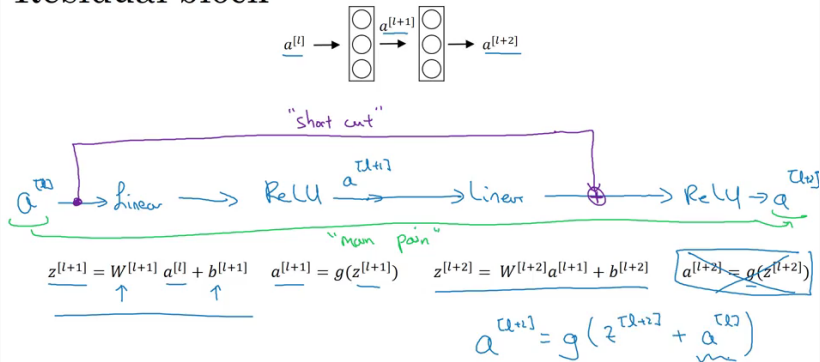
ResNet-18



- **Residual Network**
- 34 capas (es muchísimo)
- *Skip connection* para solucionar el problema del gradiente que se desvanece o explota
- Esta técnica permite entrenar redes mucho más profundas

Skip connection

Residual block



[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng ³

³Andrew Ng, *Convolutional Neural Networks*, Coursera, 2017

Experimentación

Resultados de la clasificación

predicted: boligrafos



predicted: marroquineria



predicted: tecnologia



predicted: marroquineria



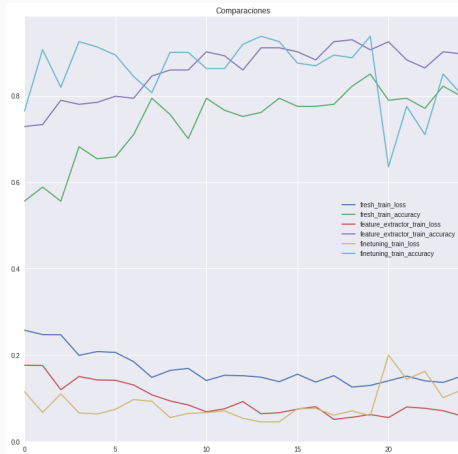
predicted: marroquineria



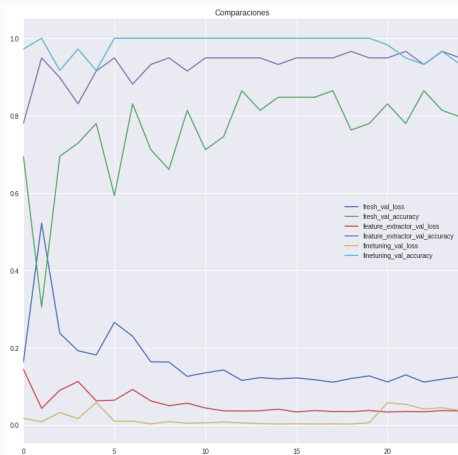
predicted: marroquineria



Evolución de accuracy y loss - entrenamiento



Evolución de accuracy y loss - validación



Tiempos de entrenamiento

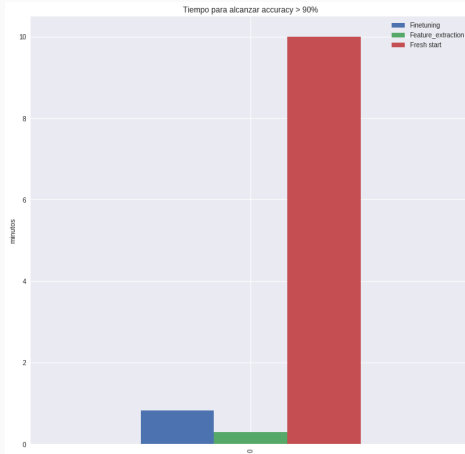


Figura 4: Tiempo para alcanzar accuracy > 90 %

¿Preguntas?