

Una guía intuitiva de redes neuronales convolucionales

Carro Barallobre, Icí
Álvarez Crespo, Lucía María (colaboradora)

Universidade da Coruña, 2019

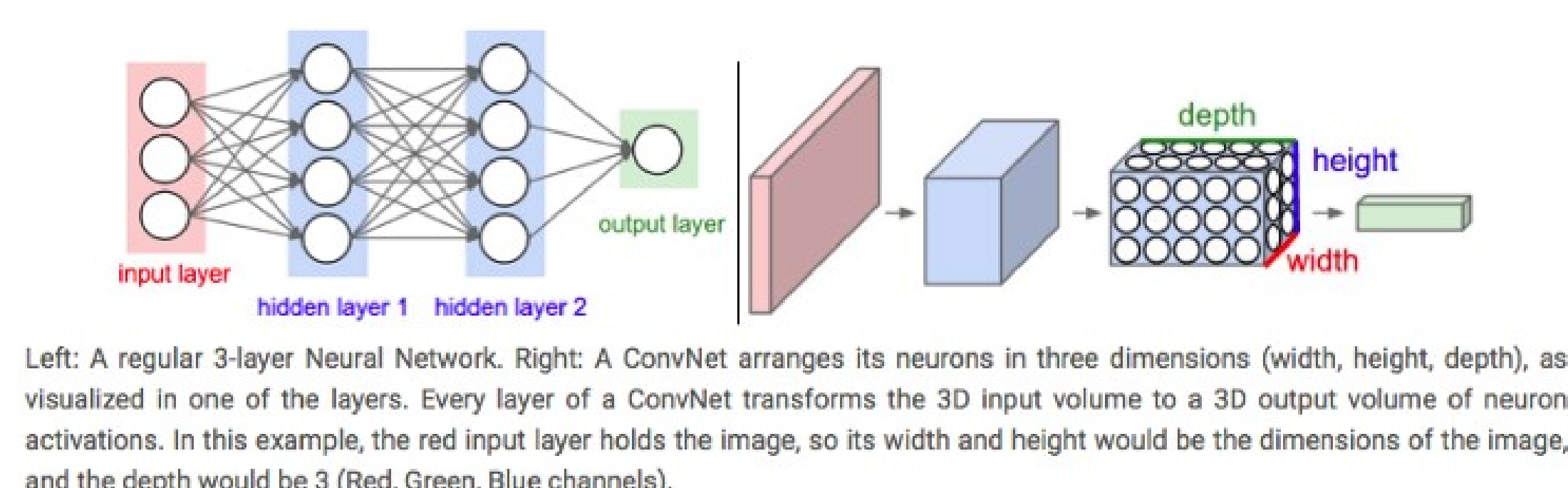
CNN y su analogía con el cerebro

Constantemente estamos analizando el mundo que nos rodea: vemos, etiquetamos, hacemos predicciones y reconocemos patrones en función de lo que hemos aprendido. Análogamente, necesitamos mostrar imágenes a un algoritmo antes de que sea capaz de generalizar la entrada y hacer predicciones para imágenes que nunca antes había visto. Para enseñar a un algoritmo a reconocer objetos en imágenes, usamos un tipo específico de RNA: las redes neuronales convolucionales (CNN), inspiradas en el cerebro. En el cerebro hay 2 tipos básicos de células de neuronas visuales que actúan de manera diferente: células simples que se activan cuando identifican formas básicas y, células complejas que tienen campos receptivos más grandes y su salida no es sensible a la posición específica en el campo, por lo que continúan respondiendo a un cierto estímulo aunque cambie su posición absoluta en la retina.

Arquitectura (RNA convencional vs CNN)

En las CNN las capas están organizadas en tres dimensiones: ancho, alto y profundo. Además, las neuronas en una capa no se conectan a todas las neuronas en la siguiente capa sino solo a una pequeña región de la misma y, la salida final se reducirá a un solo vector de puntajes de probabilidad, organizado a lo largo de la dimensión de profundidad. Las CNN tienen dos componentes:

- Parte de extracción de características (capa oculta): en esta parte, la red realizará una serie de convoluciones y operaciones de agrupación durante las cuales se detectan las características. En una imagen de una cebra: esta es la parte donde la red reconocería sus franjas, dos orejas y cuatro patas.
- La parte de clasificación: aquí las capas totalmente conectadas servirán como un clasificador por encima de estas características extraídas. Asignarán una probabilidad para que el objeto en la imagen sea lo que el algoritmo predice que es.



Extracción de características

Las CNN tienen la particularidad de usar aprendizaje profundo (Deep Learning) para la extracción de características; esto es, el propio algoritmo decidirá qué detalles de la imagen considera más relevantes a la hora de diferenciar los distintos objetos presentados. Es interesante no solo porque evita la programación de técnicas basadas en visión artificial, sino también porque sus resultados suelen ser de alta fiabilidad.

Contacto

- lucia.maria.alvarez.crespo@udc.es
- icia.carro.barallobre@udc.es



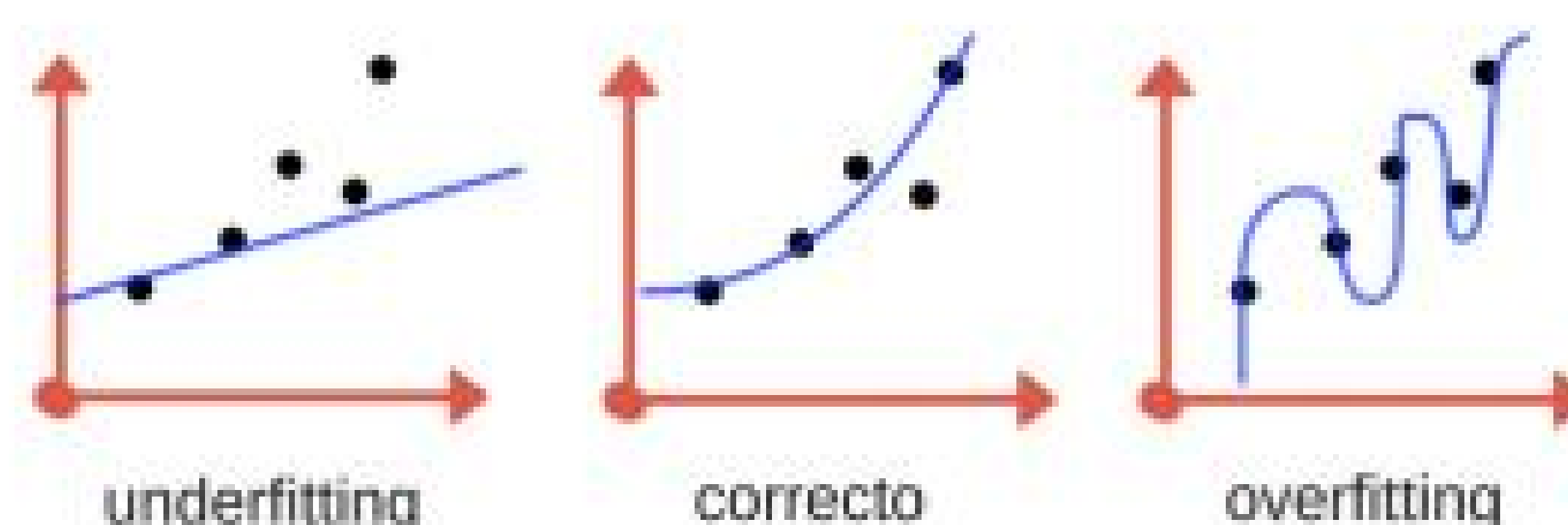
UNIVERSIDADE DA CORUÑA

Clasificación

Después de la convolución y la agrupación de capas, nuestra parte de clasificación consta de unas pocas capas totalmente conectadas. Sin embargo, estas capas totalmente conectadas solo pueden aceptar datos de una dimensión, por lo que debemos convertir nuestro volumen 3D en un vector 1D. Las últimas capas de un NN convolucional son capas totalmente conectadas: como en una RNA convencional.

Entrenamiento

Como ya hemos comentado, al igual que a un niño debemos mostrar ejemplos y corregirlos para que puedan aprender a diferenciar y etiquetar objetos. Aquí se presentan un problema: ¿Cuándo dejamos de entrenar a nuestra red neuronal? Parece lógico pensar que cuantos menos ejemplos, menos capacitada estará para responder a la entrada que le pasemos y cuantos más, mejor sabrá hacerlo: pero esto no es así. Si bien es cierto que con pocos ejemplos (underfitt) no clasificaremos correctamente, cuando un sistema se entrena demasiado o se entrena con datos extraños (overfitt), se ajusta a unas características muy específicas de los datos de entrenamiento que no tienen relación casual con la función objetivo, lo que lleva a que el éxito al responder usando como entrada las muestras de entrenamiento se incrementa pero, con muestras nuevas muestras empeora. Debemos encontrar un punto medio en el aprendizaje de nuestro modelo en el que no estemos incurriendo en underfitting y tampoco en overfitting. A veces esto puede resultar una tarea muy difícil.



Conclusión

En resumen, las CNN son especialmente útiles para la clasificación y reconocimiento de imágenes. Contienen dos partes: una parte de extracción de características y una parte de clasificación. La principal técnica especial en las CNN es la convolución: donde un filtro se desliza sobre la entrada y combina el valor de entrada más el valor del filtro en el mapa de características.

Referencias

- [1] D. Cornelisse. An intuitive guide to convolutional neural networks, Apr 2018. URL <https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>.
- [2] A. Dertat. Applied deep learning - part 4: Convolutional neural networks, Nov 2017. URL <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>.