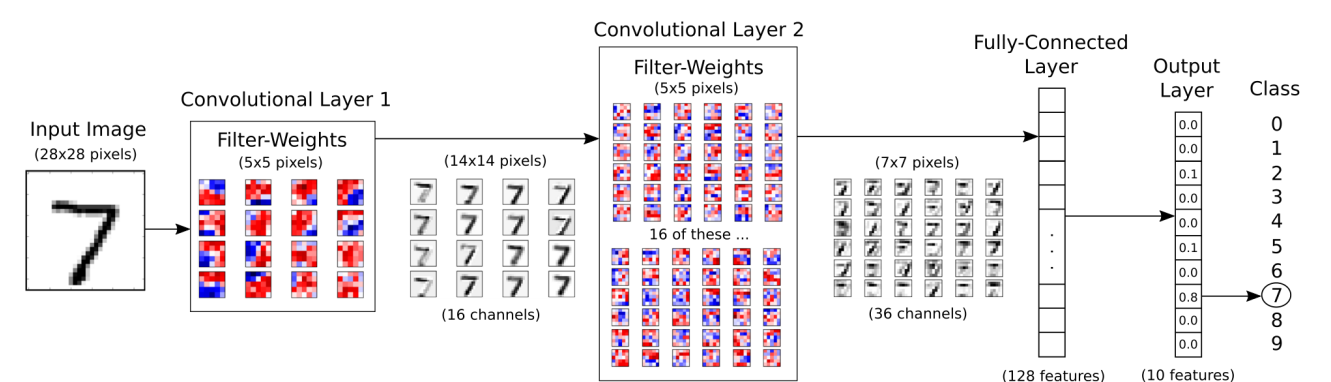


# SISTEMA DE IDENTIFICACIÓN DE LENGUAJE DE SEÑAS USANDO REDES NEURONALES CONVOLUCIONALES (CNN)

POR ISAAC MENCHACA

## INTRODUCCIÓN

Las redes neuronales convolucionales (CNN) son herramientas ampliamente utilizadas para el aprendizaje profundo. Son especialmente adecuados para imágenes como entradas, aunque también se utilizan para otras aplicaciones como texto, señales y otras respuestas continuas. Una CNN consta de varias capas, como **capas convolucionales**, capas de **agrupación máxima** o **agrupación promedio** y capas **completamente conectadas**.



## DISEÑO DE LA RED NEURONAL

La arquitectura de una CNN puede variar según los tipos y la cantidad de capas incluidas. Los tipos y la cantidad de capas incluidas dependen de la aplicación o los datos en particular. Por ejemplo, si tiene respuestas categóricas, debe tener una función de clasificación y una capa de clasificación, mientras que si su respuesta es continua, debe tener una capa de regresión al final de la red. Una red más pequeña con solo una o dos capas convolucionales podría ser suficiente para aprender una pequeña cantidad de datos de imagen en escala de grises. Por otro lado, para datos más complejos con millones de imágenes en color, es posible que necesite una red más complicada con **múltiples capas convolucionales** y **completamente conectadas**.

En el diseño se usó una red neuronal multicapa con una capa final de clasificación completamente conectada. La estructura de la red está conformada por una capa de entrada de 200x200x3, 4 capas intermedias y una capa de salida de 24 clasificaciones (26 letras del abecedario inglés, menos la J y Z porque requieren movimientos)

## ETAPA DE ENTRENAMIENTO

En la etapa de entrenamiento se utiliza un algoritmo de descenso de gradiente estocástico con impulso (sgdm). Es un método que ayuda a acelerar los vectores de gradientes en las direcciones correctas, lo que conduce a una convergencia más rápida.

Se utilizó una base de datos de la página kaggle. El dataset consta de 3000 fotografías 200x200 por cada letra del abecedario. En total son 72000 imágenes. De este total se tomó el 70% para entrenar y el restante para validar el entrenamiento.

## RESULTADOS Y CONCLUSIONES

Hubo una gran diferencia al hacer el entrenamiento con una cantidad mayor de datos. La precisión aumentó considerablemente. Como futuro trabajo se planea mejorar el preprocesamiento de los datos introducidos a la red, con el fin de mejorar la precisión.

