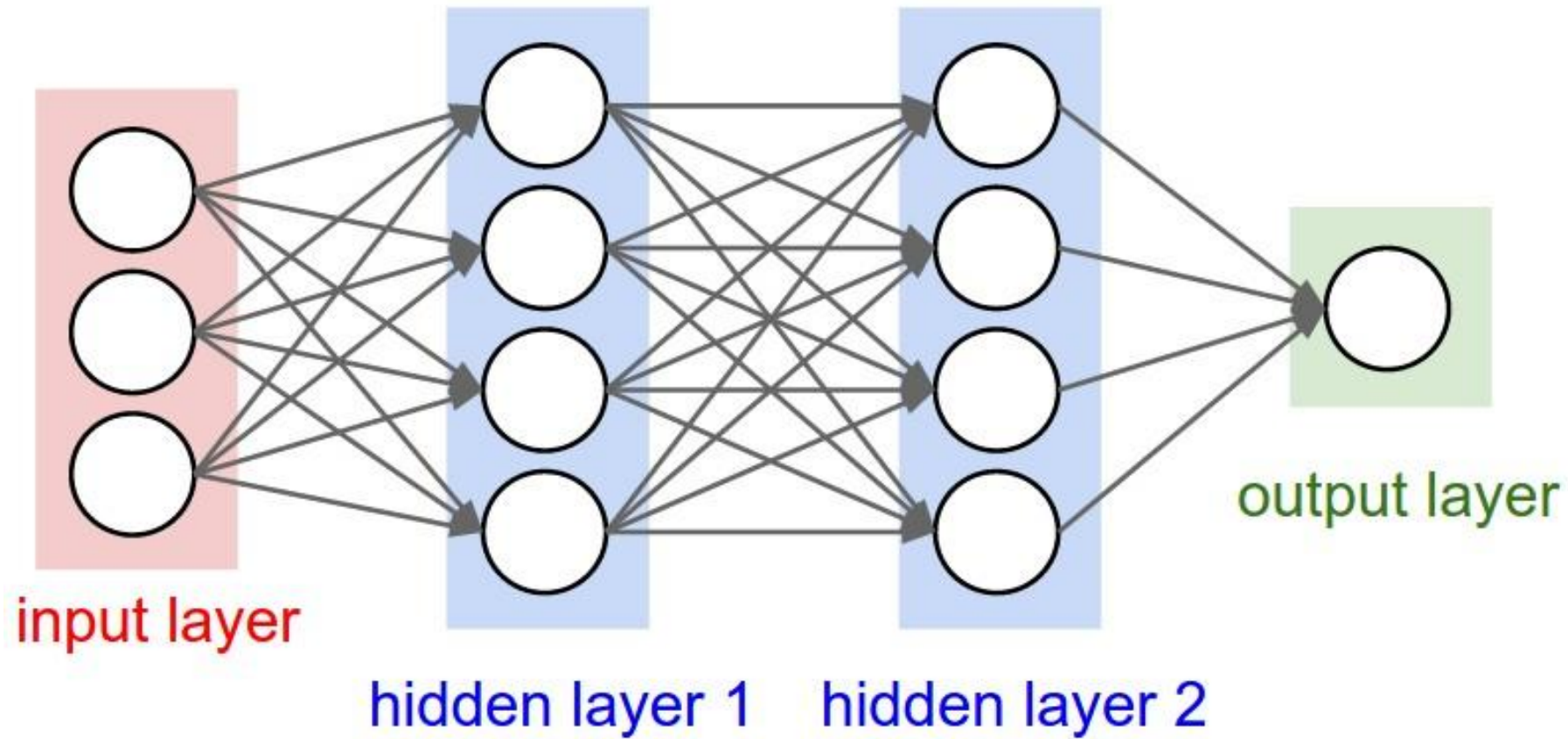# Convolutional Neural Networks (CNN)

E. Milgo

Introduction

- A Convolutional Neural Network (CNN) is a type of deep neural network primarily used for analyzing visual data.

- CNN can extract higher representations of the image.

- It is particularly effective for tasks like image classification, object detection, and segmentation.

- In classical image classification you define the image features. CNN takes the image's raw pixel data, trains the model and then extracts the features for better classification.
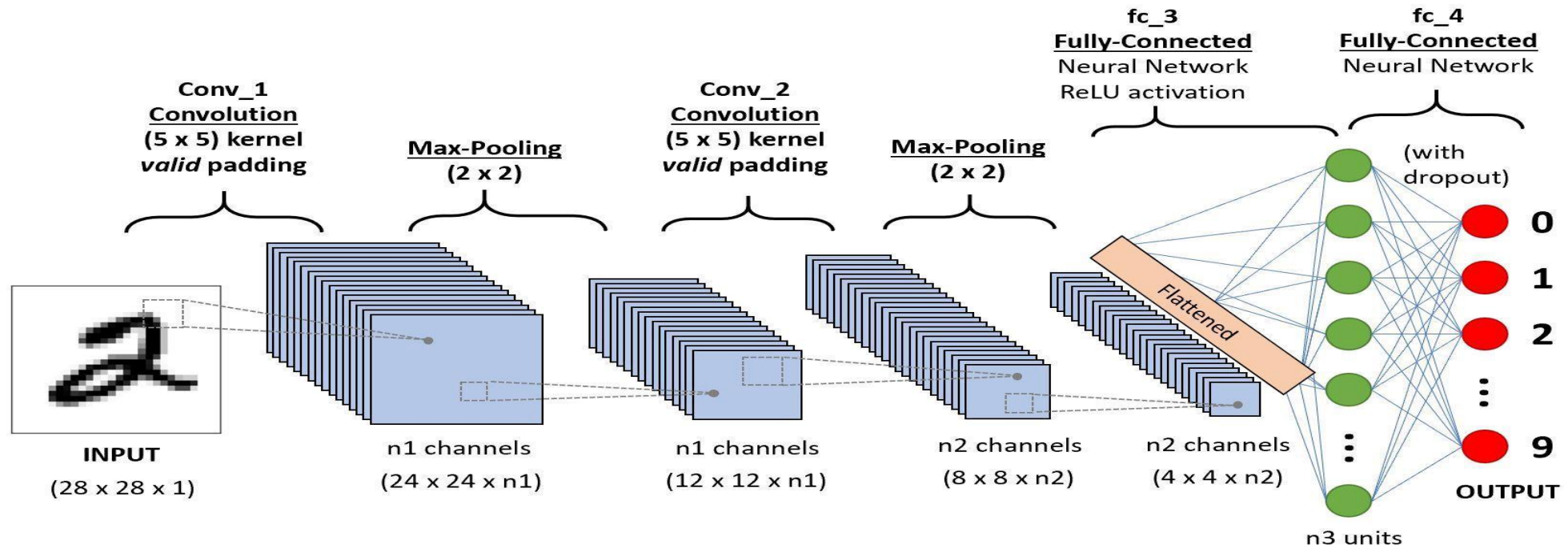
Neural Network

input layer

hidden layer 1   hidden layer 2

output layer

Convolutional Neural Networks (CNNs)
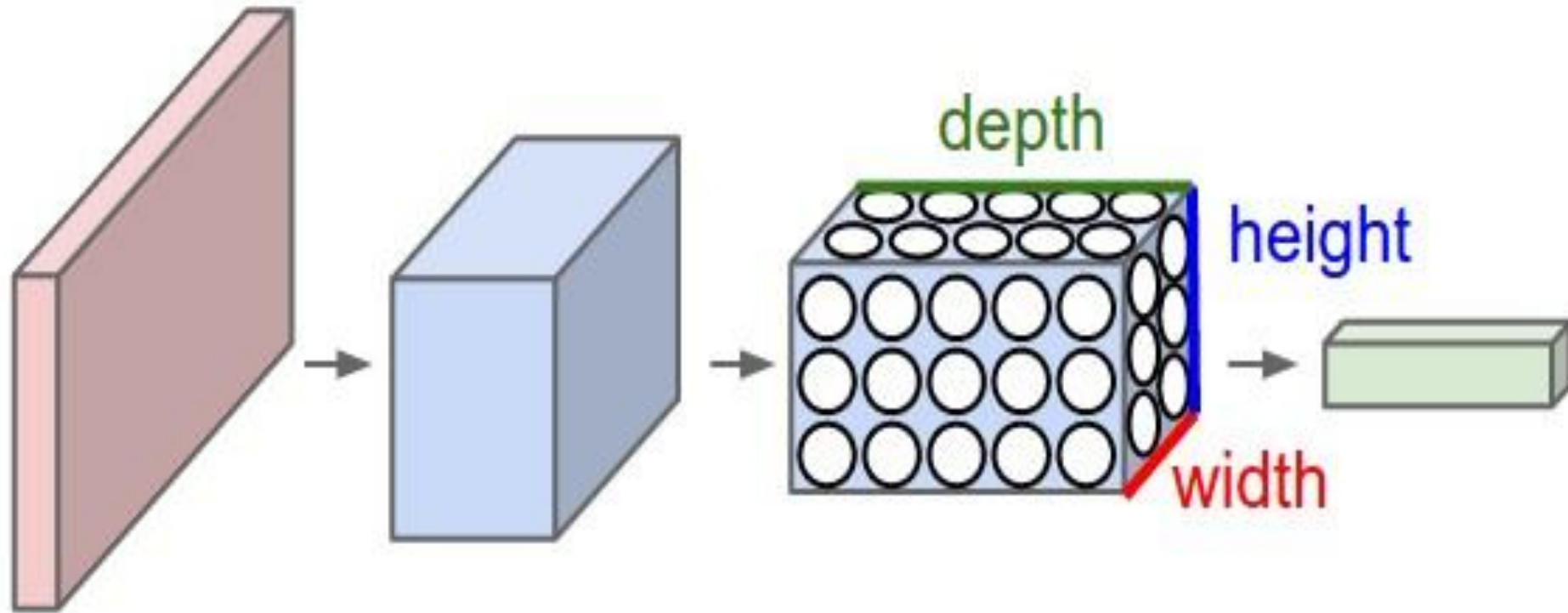
How does it work

Layers in CNN

- Input Layer: Image input (e.g., 28x28 pixels, grayscale or RGB).

- **Convolutional Layer** : Applies filters (kernels) to detect features like edges or textures.

- Activation Function (ReLU): Introduces non-linearity.

- **Pooling Layer (Max Pooling)**: Downsamples the image to reduce dimensions and computation.

- **Fully Connected Layers (Dense)** : Flattened features are passed through dense layers for classification.

- Output Layer: Final predictions

Convolutional Layer

- A Convolutional Layer applies a set of filters (also called kernels) to the input (e.g., an image or feature map) to extract features like **edges**, **corners**, **textures**, or more complex patterns.

How It Works

- A filter (usually a small matrix like 3x3 or 5x5) slides over the input data.

- At each position, it performs element-wise multiplication and sums the result → this is called the convolution operation.

- The output is a feature map (also called an activation map).

"Convolution"

Filter 3x3

Output 4x4

Original image 6x6

Result of the element-wise product and sum of the filter matrix and the orginal image
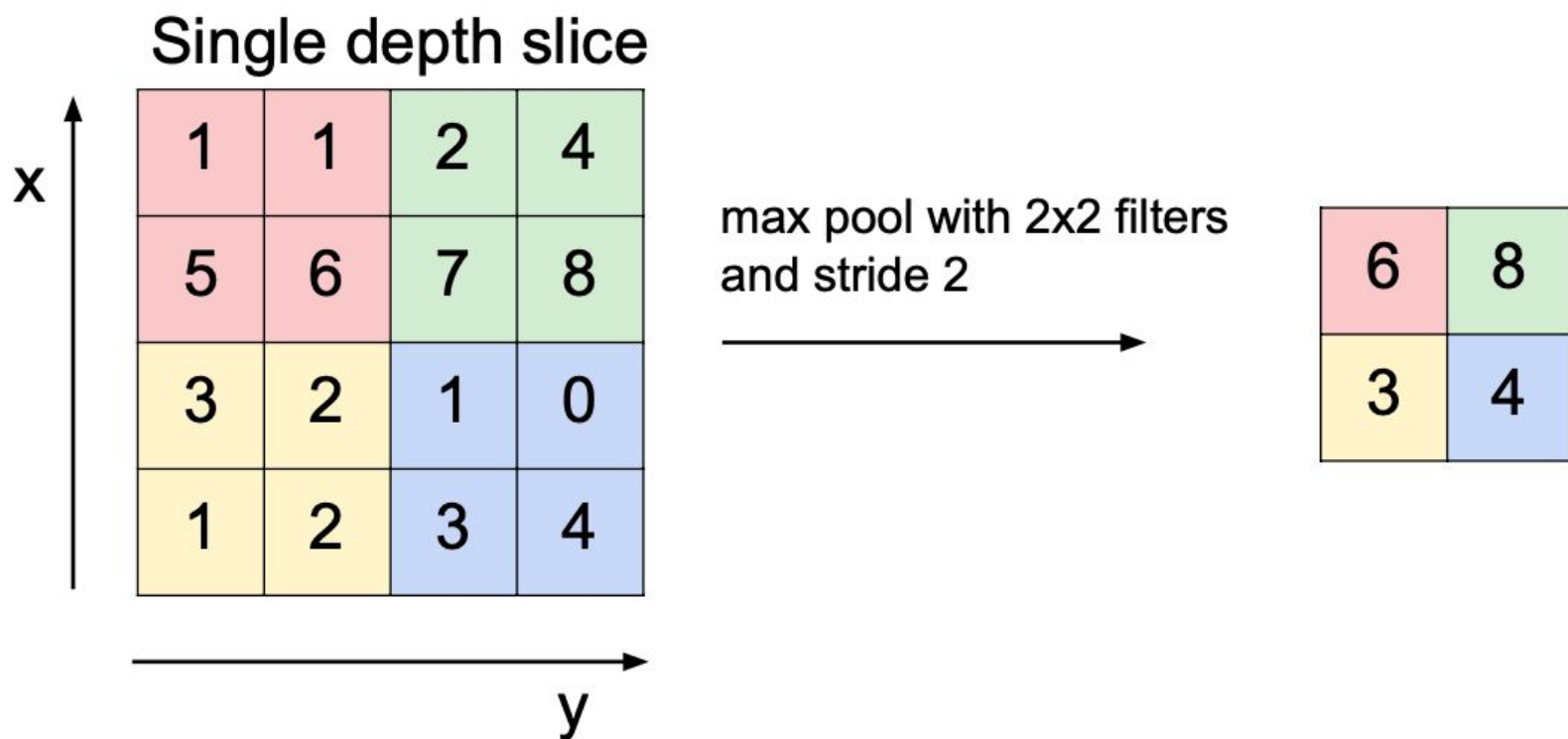
maxPooling Layer

- Max Pooling is a downsampling technique used in Convolutional Neural Networks (CNNs).

- It reduces the spatial dimensions (width and height) of the input feature map while retaining the most important information.
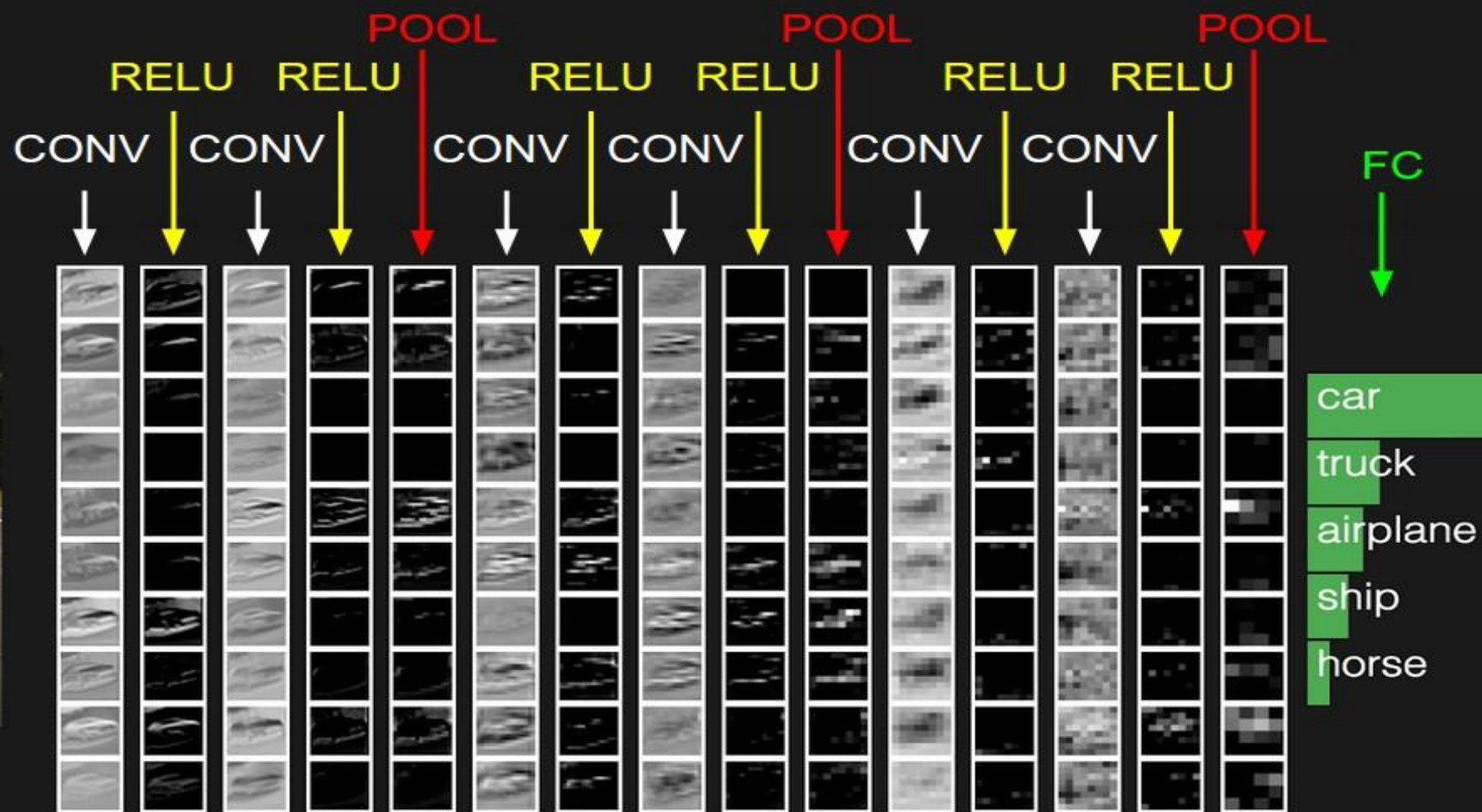
How it works

- A window (kernel) slides over the input feature map.
- For each window, it outputs the maximum value within that window.
- Common window size: 2x2 with stride 2 (i.e., non-overlapping).

MAX POOLING

Single depth slice

Fully Connected Layer (FC)

- A Fully Connected Layer (FC) is a standard neural network layer where every neuron is connected to every neuron in the previous layer.

- It's used at the end of the CNN, after the convolutional and pooling layers, to:
  - Flatten the spatial features
  - Interpret them
  - Produce the final output, like class scores

RELU RELU POOL RELU RELU POOL RELU RELU POOL

CONV CONV CONV CONV CONV CONV FC

car
truck
airplane
ship
horse

Application of CNN

- Image classification (LeNet, InceptionNet, ResNet)
- Image segmentation (UNet, FCNN, RCNN)
- Image Generation(GAN)