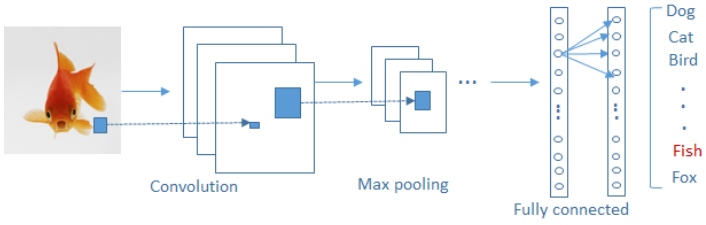
Deep Learning

CNN: Introduction

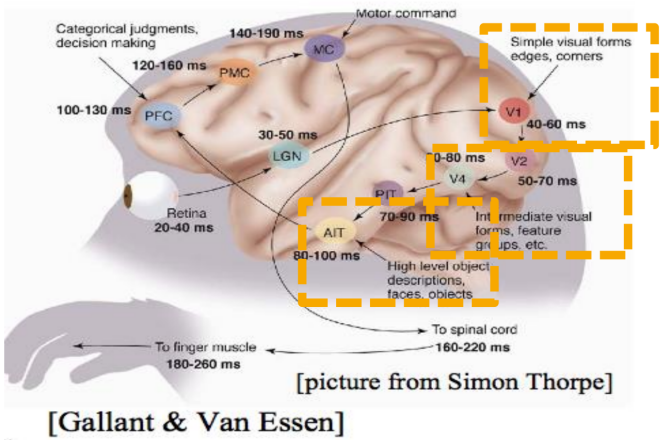


Learning goals

- Introduction to CNNs
- Applications of CNNs
- A First Glimpse to the Architecture

CONVOLUTIONAL NEURAL NETWORKS

- Convolutional Neural Networks (CNN, or ConvNet) are a powerful family of neural networks that are inspired by biological processes in which the connectivity pattern between neurons resembles the organization of the mamel visual cortex.



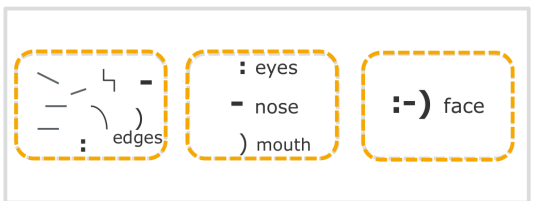


Figure: The ventral (recognition) pathway in the visual cortex has multiple stage: Retina - LGN - V1 - V2 - V4 - PIT - AIT etc., which consist of lots of intermediate representations.

CONVOLUTIONAL NEURAL NETWORKS

- Since 2012, by the success of CNNs model in ILSVRC compition, they are popular in many fields such as computer vision.
- Common applications of CNN-based architectures in computer vision are:
  - Image classification.
  - Object detection / localization.
  - Semantic segmentation.
- CNNs are widely applied in other domains such as natural language processing (NLP), audio, and time-series data.
- Basic idea: a CNN automatically extracts visual, or, more generally, spatial features from an input data such that it is able to make the optimal prediction based on the extracted features.
- It contains different building blocks and components.

CNNS - WHAT FOR?

Figure: All Tesla cars being produced now have full self-driving hardware and customers can purchase one today (source Tesla website). Twenty-nine states out of fifty states in the U.S. have enacted legislation related to autonomous vehicles. A convolutional neural network is used to map raw pixels from a single front-facing camera directly into steering commands. The system learns to drive in traffic on local roads with or without lane markings as well as on highways.

CNNS - WHAT FOR?

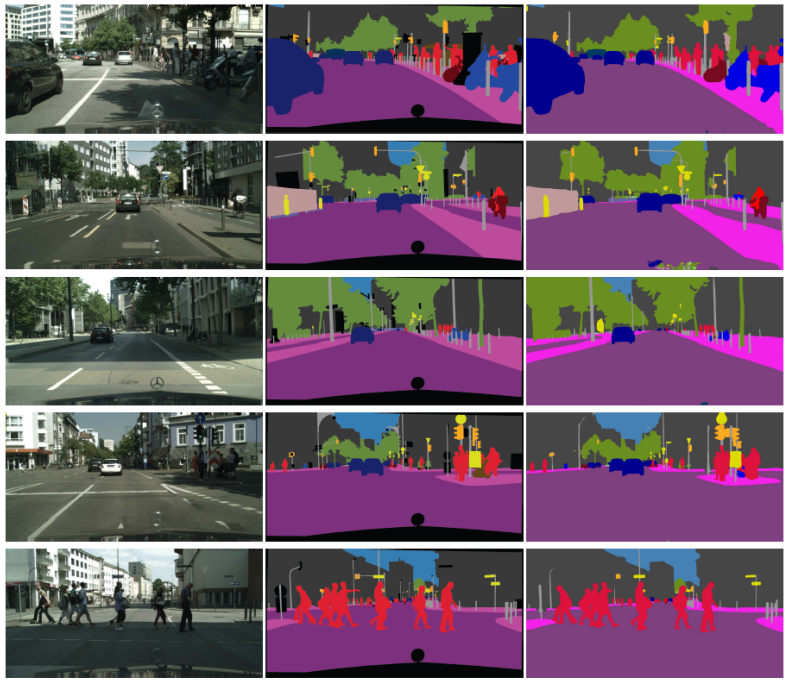


Figure: Given an input image, first CNN is used to get the feature map of the last convolutional layer, then a pyramid parsing module is applied to harvest different sub-region representations, followed by upsampling and concatenation layers to form the final feature representation, which carries both local and global context information. Finally, the representation is fed into a convolution layer to get the final per-pixel prediction. (Source: pyramid scene parsing network,by Zhao et. al, CVPR 2017)

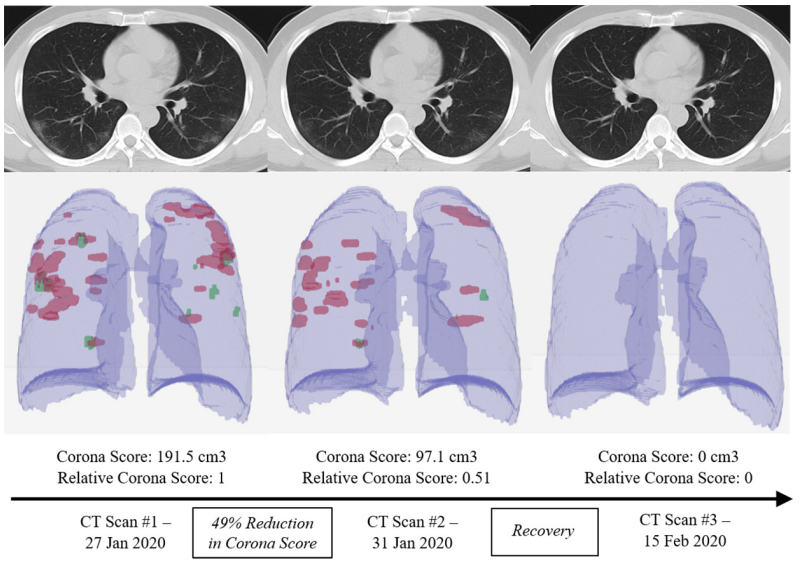
CNNS - WHAT FOR?

Figure: Road segmentation (Mnih Volodymyr (2013)). Aerial images and possibly outdated map pixels are segmented.

CNNS - WHAT FOR?

CNN for personalized medicine

- Examples
  are: Tracking, diagnosis and localization of Covid-19 patients.
- CNN
  based method (RADLogists) for personalized Covid-19 detection: three CT scans from a single coronavirus patient diagnosed by RADLogists.



CNNS - WHAT FOR?

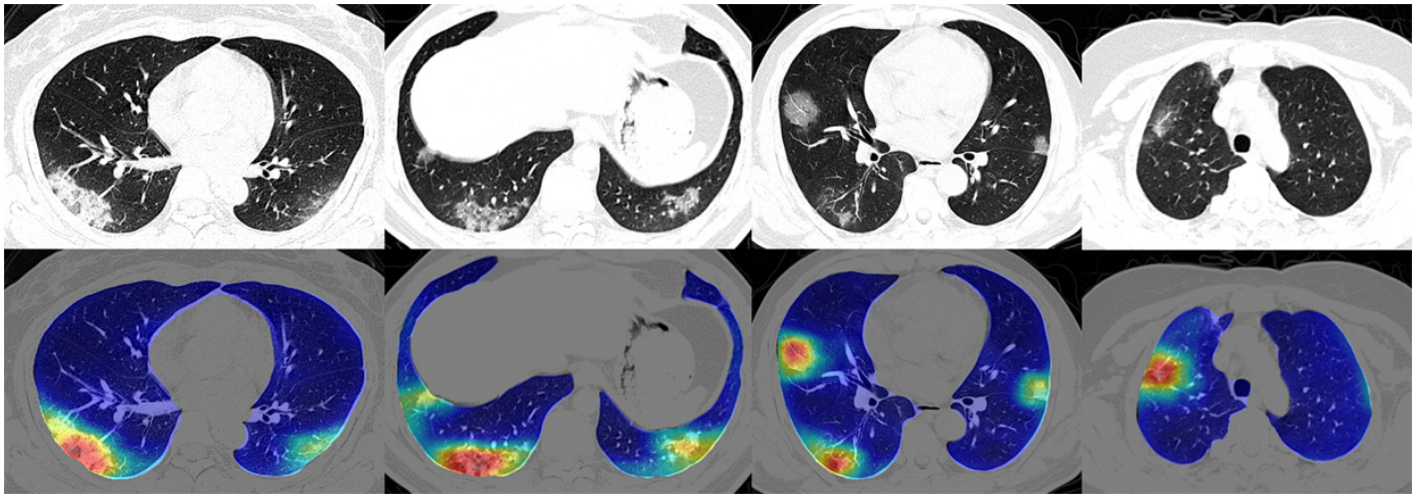


Figure: Four COVID-19 lung CT scans at the top with corresponding colored maps showing coronavirus abnormalities at the bottom (Source: IEEE Spectrum).

CNNS - WHAT FOR?

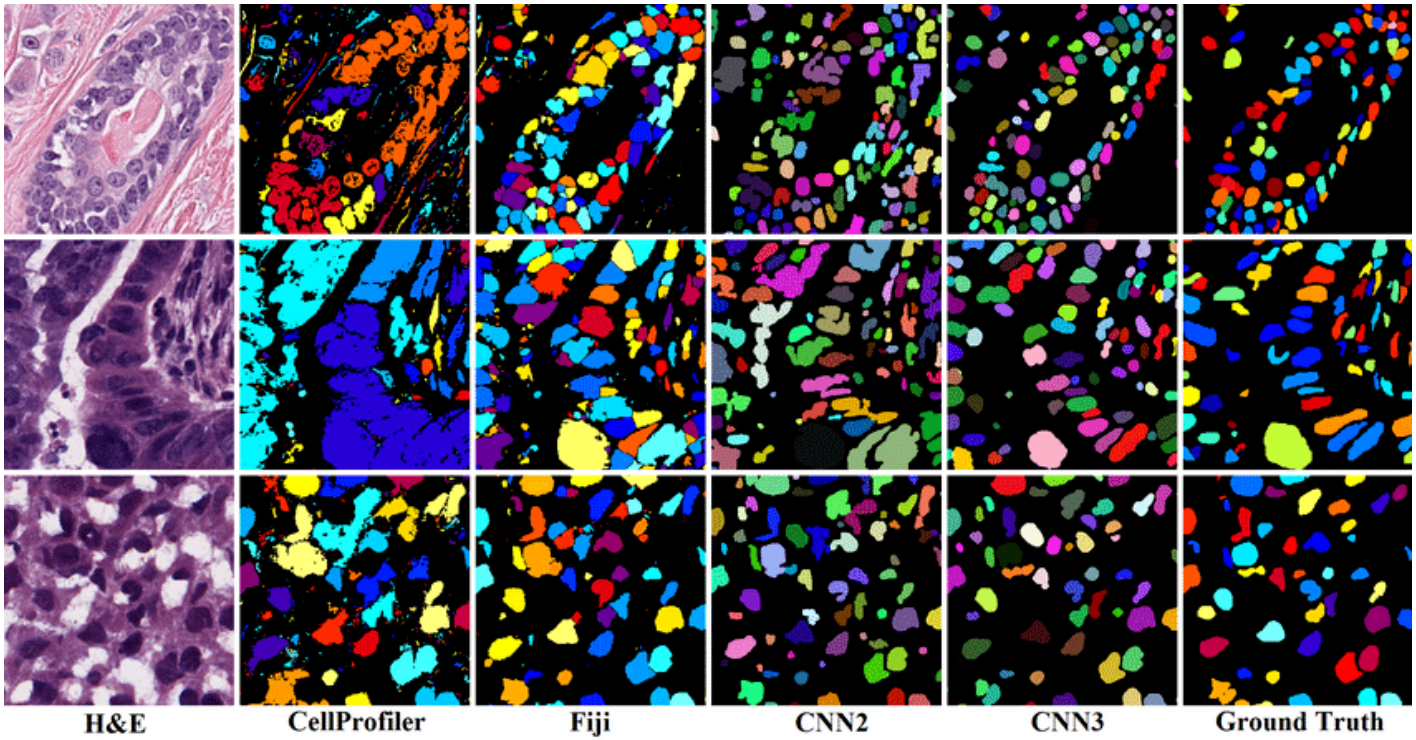


Figure: Various analyses in computational pathology are possible. For example, nuclear segmentation in digital microscopic tissue images enable extraction of high-quality features for nuclear morphometrics (source: Kummar et. al. IEEE Transaction Medical Imaging).

CNNS - WHAT FOR?

Figure: Image Colorization is another interesting application of CNN in computer vision (Zhang et al. (2016)). Given a grayscale photo as the input (top row), this network solves the problem of hallucinating a plausible color version of the photo (bottom row, i.e. the prediction of the network).

CNNS - WHAT FOR?

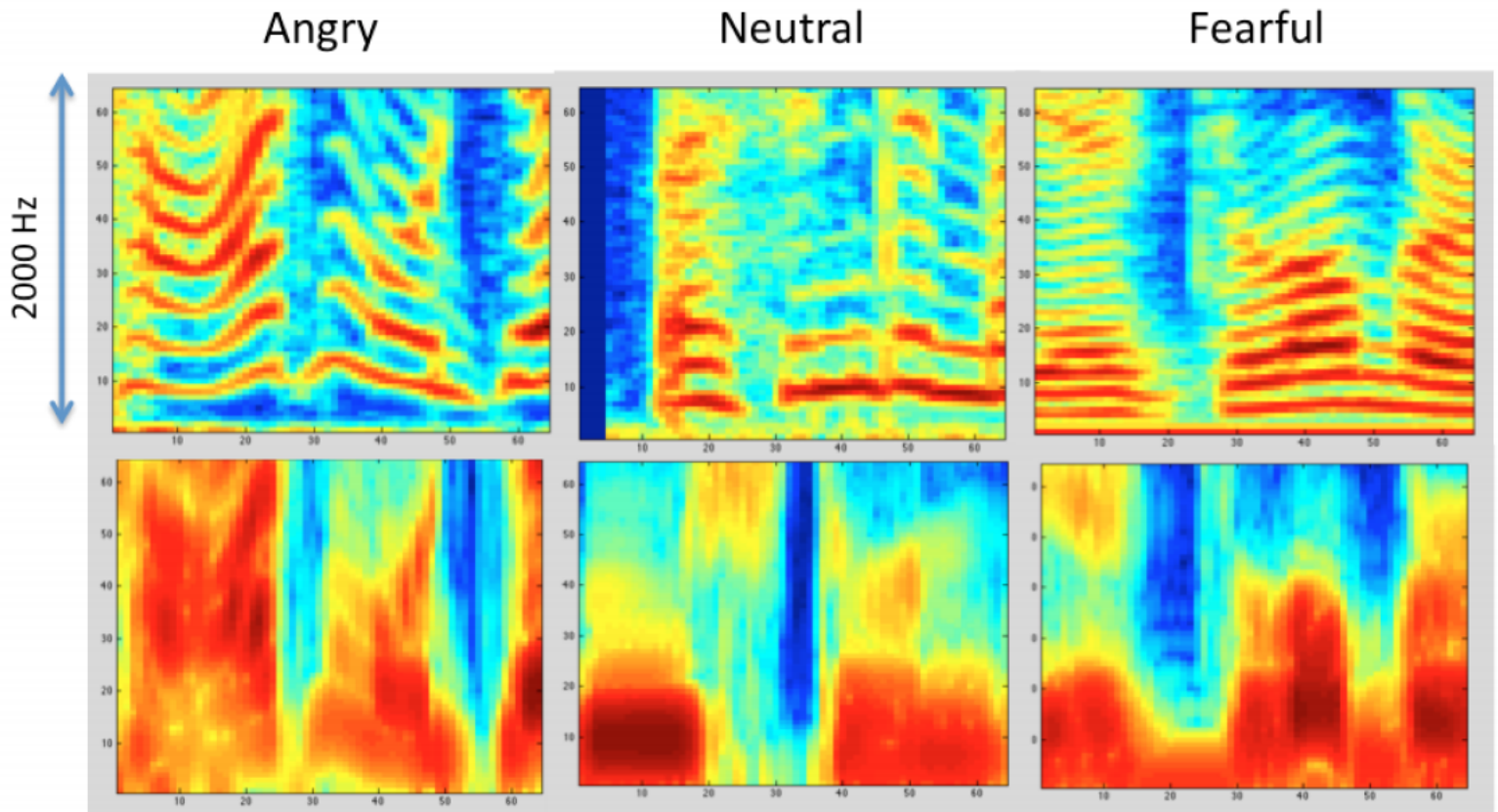
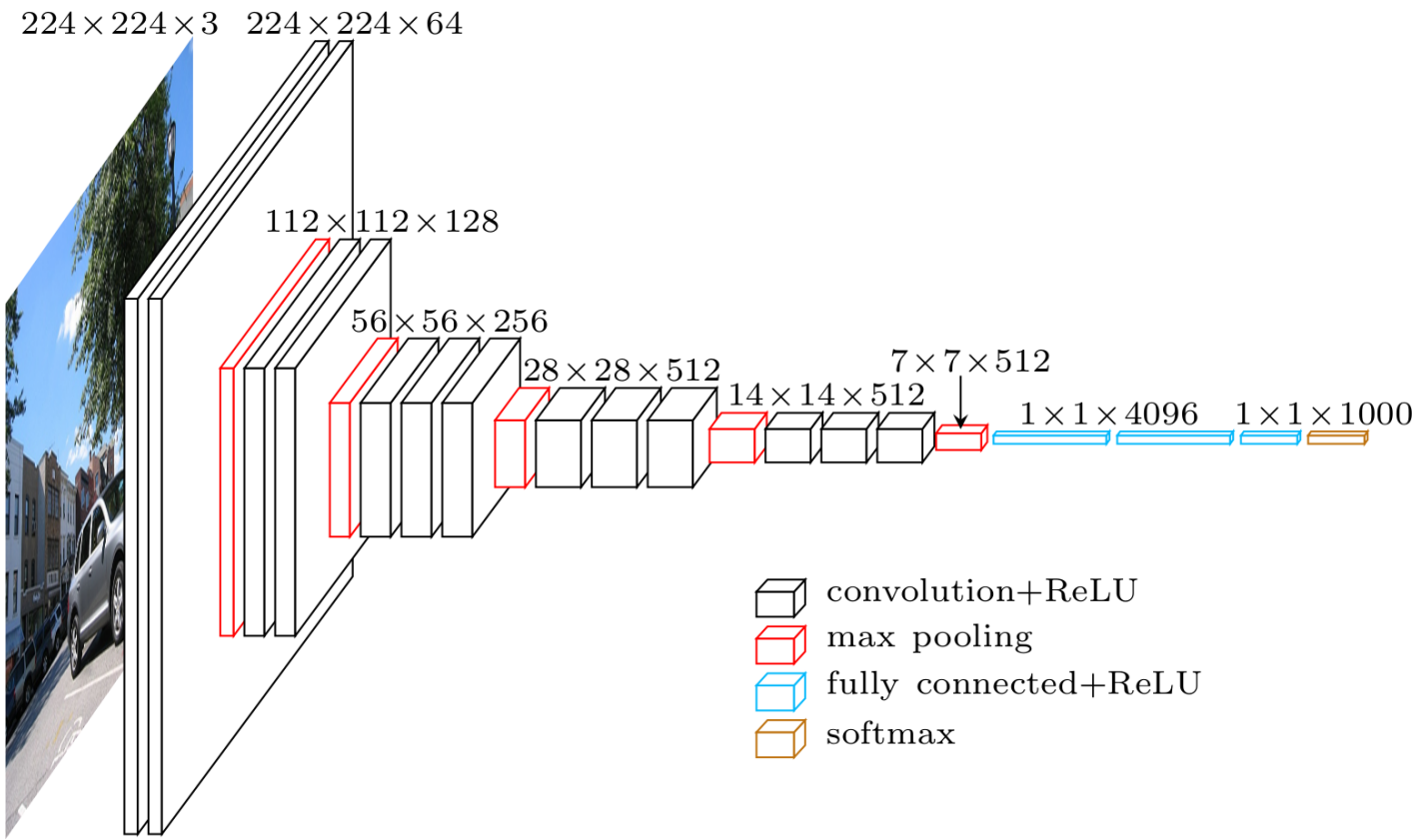


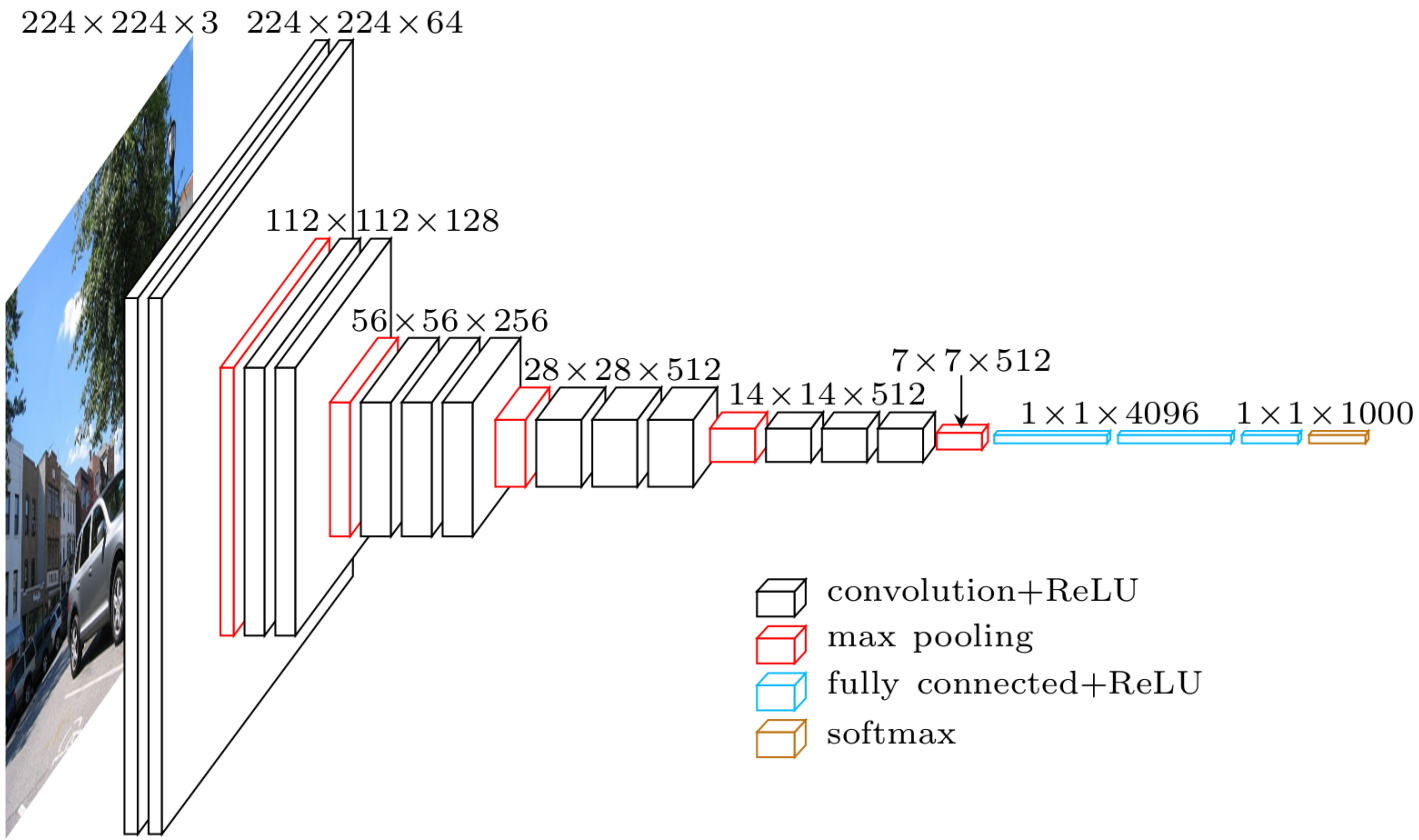
Figure: Speech recognition (Anand & Verma (2015)). Convolutional neural network is used to learn features from the audio data in order to classify emotions.

CNNS - A FIRST GLIMPSE



- **Input layer** takes input data (e.g. image, audio).
- **Convolution layers** extract feature maps from the previous layers.
- **Pooling layers** reduce the dimensionality of feature maps and filter meaningful features.

CNNS - A FIRST GLIMPSE



- **Fully connected layers** connect feature map elements to the output neurons.
- **Softmax** converts output values to probability scores.