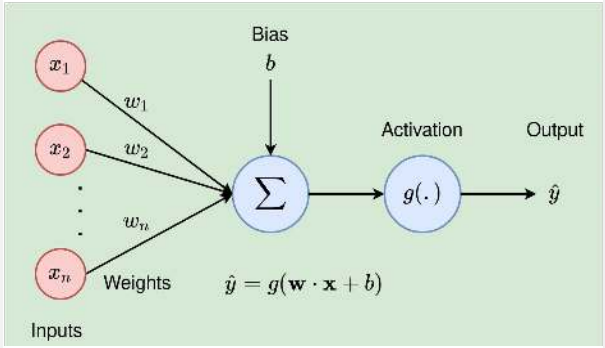
RCIF Applied AI Seminar Series: Training Segmentation Models with Monai and MedSAM

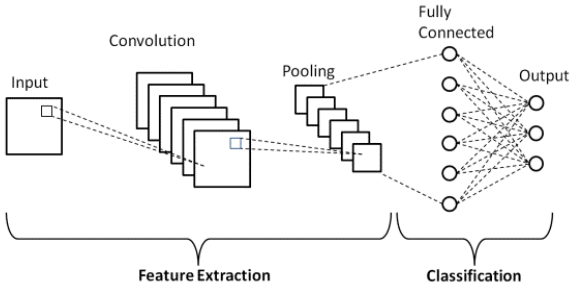
Basics of Deep Learning

- Anatomy of a neural net:
  - Input Layer: The entry point for data into the neural network, where each neuron represents a feature of the input dataset.
  - Weights: Parameters that are learned during the training process. They amplify or dampen the input signal, influencing the network's output.
  - Optimizer: Algorithms that adjust the weights of connections within the network to minimize the loss function. Examples include SGD, Adam, and RMSprop.
  - Activation Function: Non-linear functions that determine the output of a neural processing unit, ReLU, Sigmoid, and Tanh, adding non-linearity to enable the network to learn complex patterns.
  - Loss Function: A metric that measures the difference between the network's prediction and the actual target values, i.e. MSE for regression and Cross-Entropy for classification tasks.
  - Output Layer: Produces the model's predictions, tailored to the specific type of problem being solved, such as classification or regression.



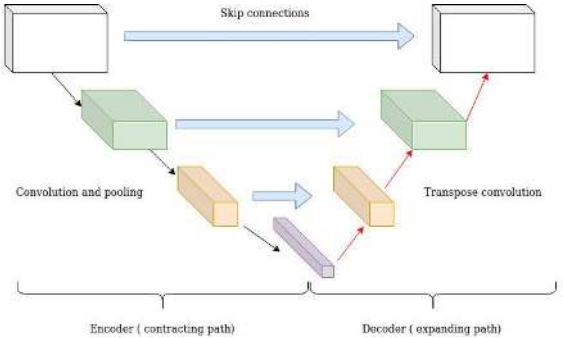
CNN Architecture

- Developed by Yann LeCun for Zip Code Recognition, AT&T Bell Labs1989
- Input Layer: Holds image pixel values (e.g., 224×224×3 for RGB).
- Convolutional Layer: Uses filters for spatial feature detection; generates 2D activation maps.
- Activation Function: Adds non-linearity (e.g., ReLU nullifies negative values).
- Pooling Layer: Reduces spatial size to lessen parameters and computation.
- Fully Connected Layer: Links all neurons to previous layer's activations for reasoning.
- Output Layer: Often a softmax layer for class probability outputs.
- Additional Techniques: Dropout prevents overfitting, batch normalization aids training.
- Feature Hierarchy: Initial layers capture basic features; deeper layers discern complex patterns.



U-Net Architecture

- Developed at University of Freiburg, Germany for biomedical image segmentation.
- Encoder/Decoder:
  - Encoder: Repeated 3×3 convolutions, ReLU activations, 2×2 max pooling, feature channels double post-downsampling.
  - Decoder: Mirrors encoder with up-convolutions, 3×3 convolutions, feature channels halve post-upsampling.
- Skip Connections:
  - Link encoder and decoder layers.
  - Enhance high-resolution feature flow.
- Final Layer:
  - 1×1 convolution for class mapping.
- Key Strengths:
  - Effective with small training sets
  - Precise localization and context integration via skip connections

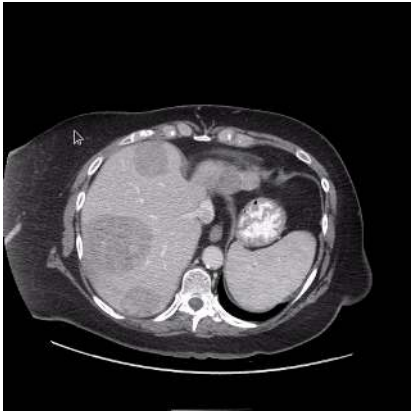


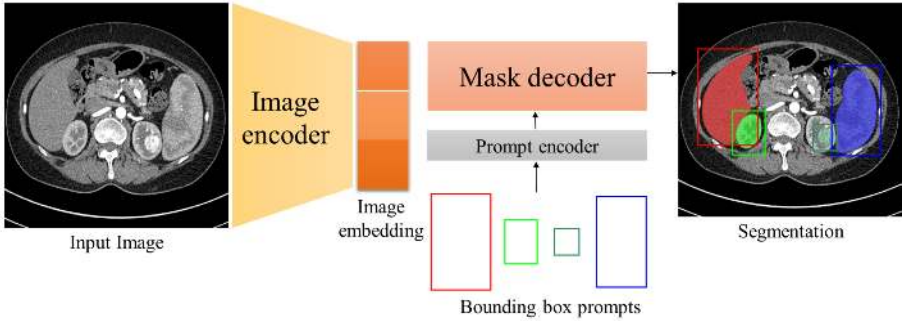
Let's Train a U-Net

U First

Introduction to the MedSAM Model

- MedSAM introduces a universal approach to medical image segmentation across modalities and diseases.
- Utilizes a large dataset for training, showcasing improved accuracy over modality-specific models.
- Expected to enhance diagnostic precision and provide deeper insights for personalized treatment plans.
- Supports three types of prompts: bounding box, point, and text. (Note: The text-prompt function in SAM is not publicly available.)





Let's Checkout MedSAM

Let's a go!