UST KiSTi www.kisti.re.kr

Seminar - Fall 2023

Medical Image Segmentation for Realizing Human Digital Twin

Syed Hasnain Raza Shah

Agenda

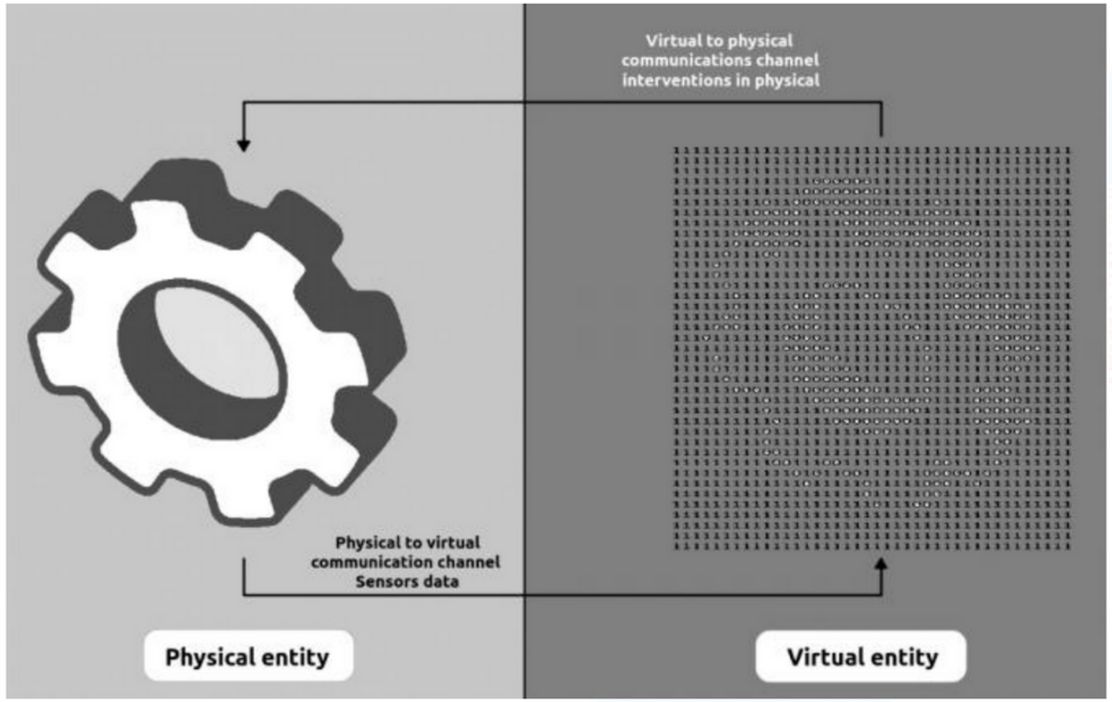
1. Human Digital Twin
2. Medical Image Segmentation
3. Related Work

Human Digital Twin

WHAT IS DIGITAL TWIN?

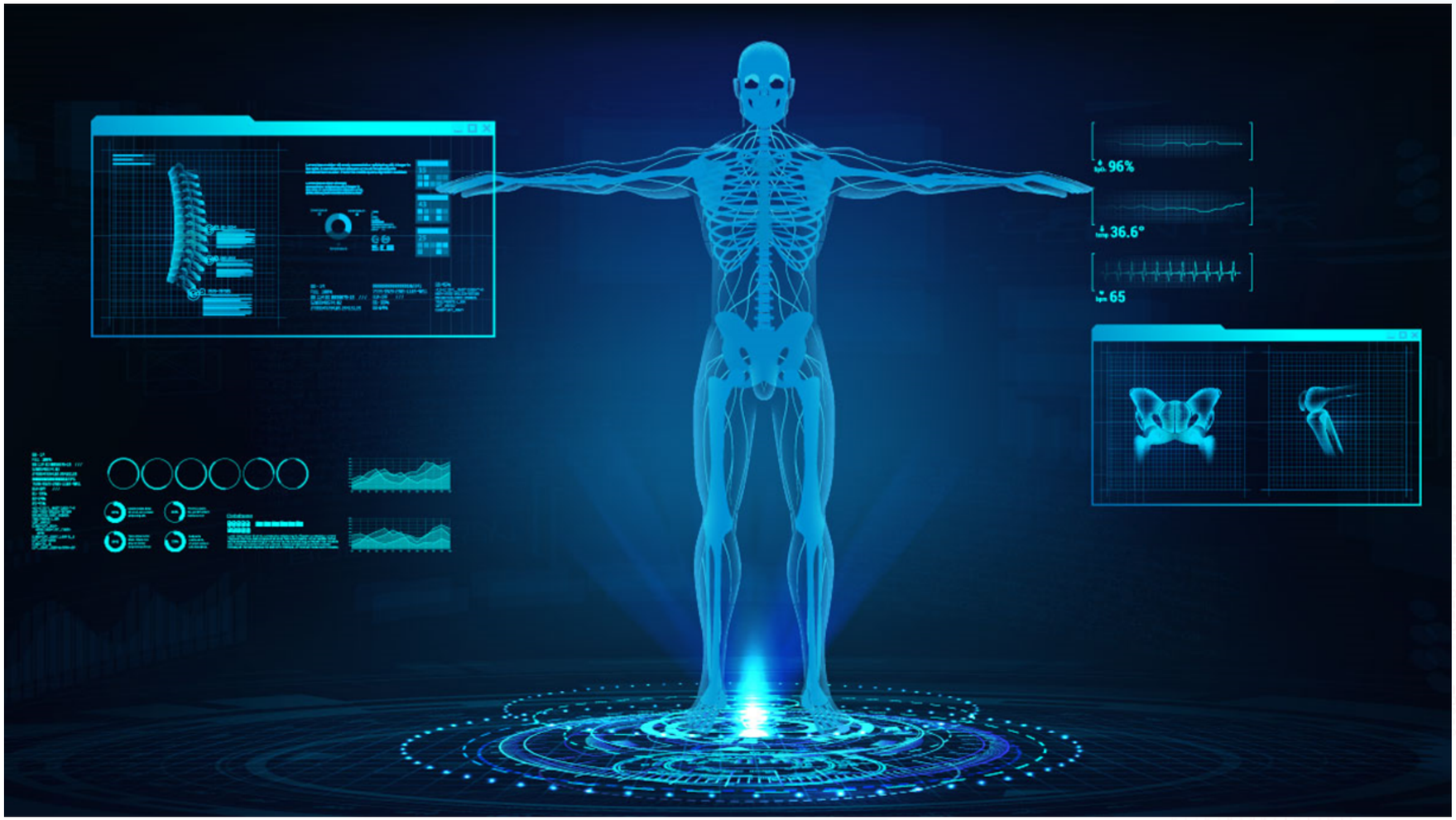
Digital Twin

A virtual replica of a physical system, process, or product, that is periodically updated with data collected from its corresponding physical entity and environment.



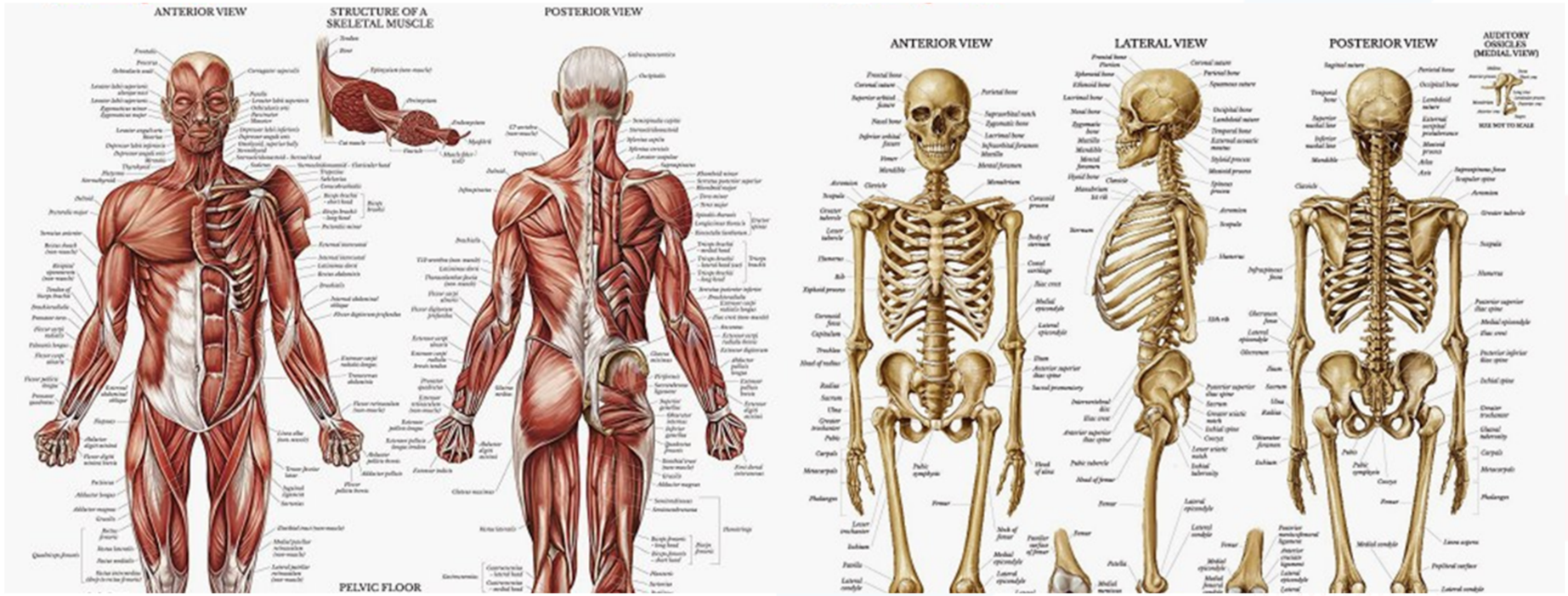
Human Digital Twin

To reproduce human anatomical characteristics and physiological conditions in a digital space, recently received a lot of attention in the medical field.



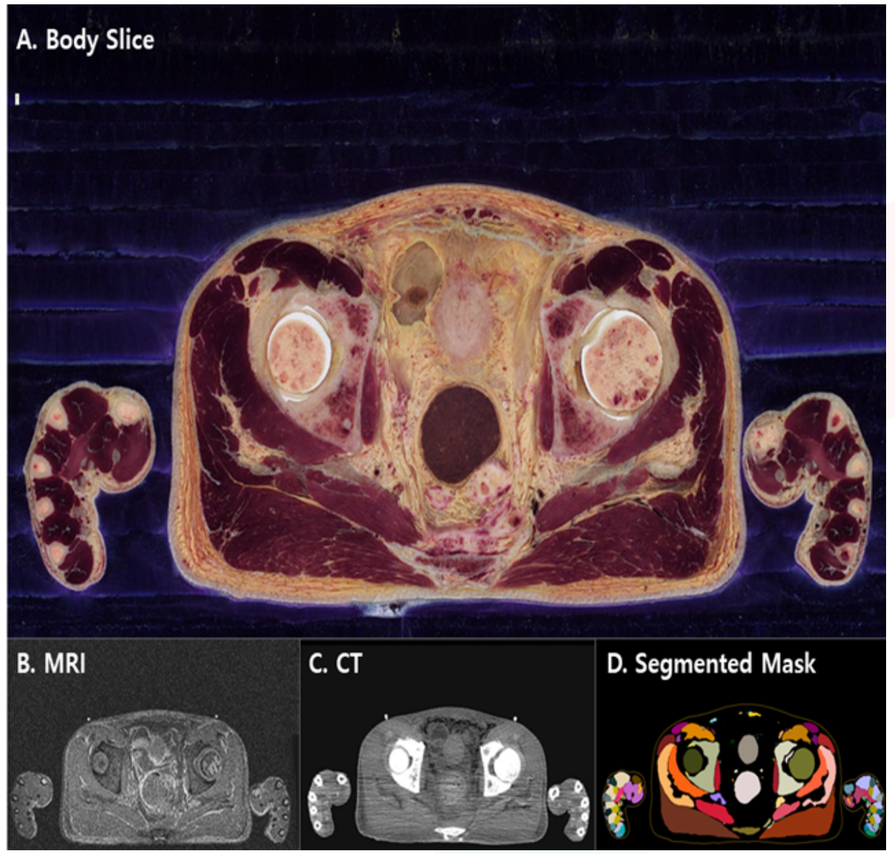
Human Digital Twin

- In order to develop a precise human digital twin, it is necessary to precisely model various elements that make up the human body.
- Among various components, it is important to precisely model the musculoskeletal structure.
  - The musculoskeletal structure composed of bones and muscles is a key component that determines the structure of the human body.
  - Bones play a role in supporting and protecting the human body, and muscles are connected to bones to control the movement of the human body according to contraction and relaxation.

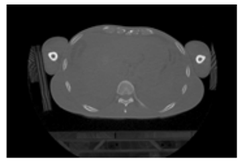


Visible Korean Dataset (KISTI, 2000-2012)

- Consists of full-body scans of a Korean man and a woman, and the corresponding CT, MRI, and segmentation masks for each body element.
- The segmentation mask consists of a total of 13 major categories and 902 subcategories.

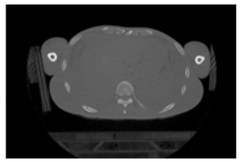


Research Pipeline for Human Digital Twin



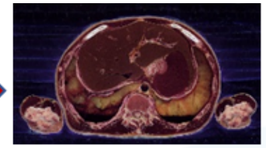
Original Image

SRx4



High resolution Image

Convertion



Body Slice

Segmentation

3d Reconstruction

3d Body

Simulation

Motivation

- Lack of research focusing on Segmentation of Bones and Muscles
- Advancing Healthcare
- Personalized treatment strategies, optimizing outcomes and reducing risks.
- Assisment in surgical planning, leading to safer and more efficient surgical procedures.
- Segmentation-driven digital twins are at the forefront of biomedical research and technology.

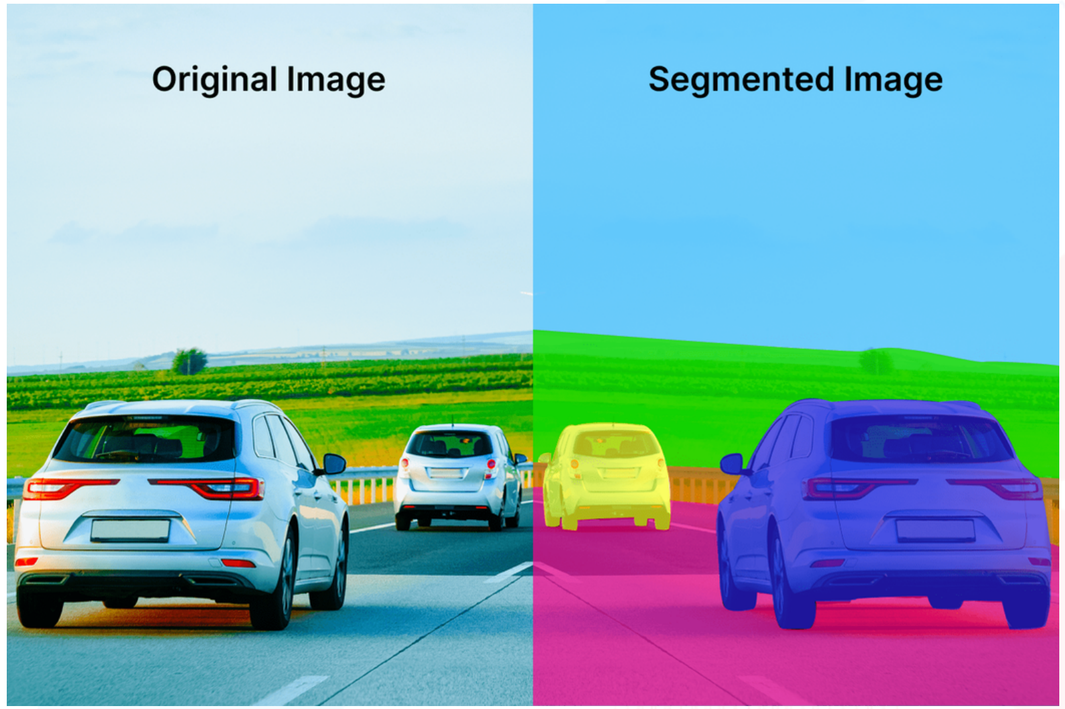
Medical Image Segmentation

Image Segmentation

The process of **dividing an image** into multiple meaningful and homogeneous **regions or objects** based on their **inherent characteristics**, such as color, texture, shape, or brightness

Applications

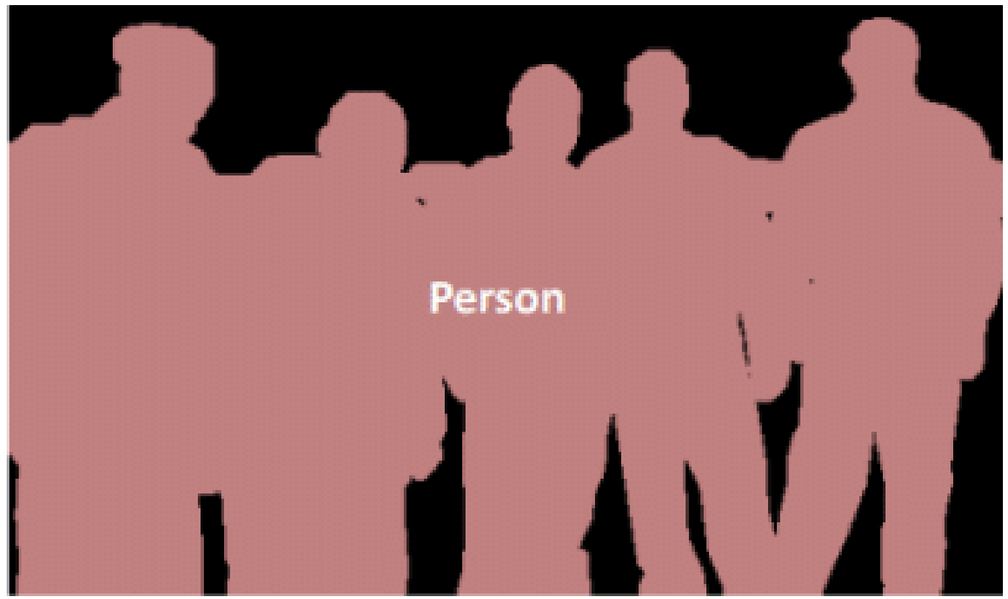
- Medical Images
- Robotics
- Autonomous Vehicles
- etc



Types of Image Segmentation

Semantic Segmentation

Pixel wise classification



Instance Segmentation

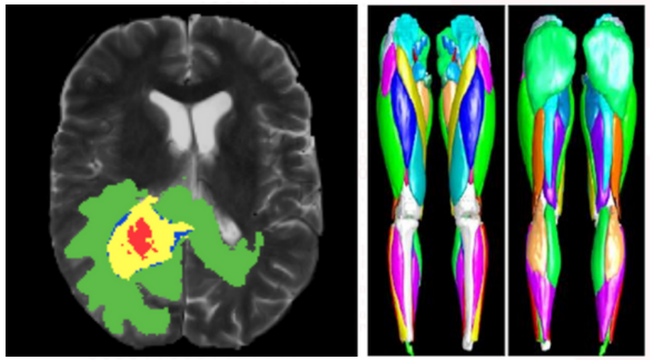
Pixel wise classification and also object instances

Medical Image Segmentation

- Development of medical equipements: **Xrays, CT, MRI and Ultrasound**
- Identifying **regions of interest** and **anatomical structures, abnormalities** in medical images
- Computer Aided **Diagnosis, Treatment Planning, Monitoring** of patients
- Helps in **accurate** diagnosis, improved **efficiency** and healthcare

Some of the popular segmentation tasks for medical are

- **Liver and Liver-tumor Segmentaion**
- **Brain and Brain-tumor Segmentation**
- **Skin lesion segmentation**
- **Prostate segmentation**
- **Lung Segmentation**
- **etc.**



Research Goals

- Understanding State-of-the-art techniques for medical image segmentation
- Aquiring the latest Vision Transformer-based expertise
- Development of segmentation model for bones and muscles on Visible Korean

Related Work

Early Approaches

Edge detection: gradient-based methods, Laplacian-based methods, Canny edge detector

Template matching techniques: template matching, normalized cross-correlation, scale-invariant feature transform (SIFT)

Statistical shape models: principal component analysis (PCA), active shape models (ASM), active appearance models (AAM)

Active contours: snakes, level sets, geodesic active contours, Chan-Vese model

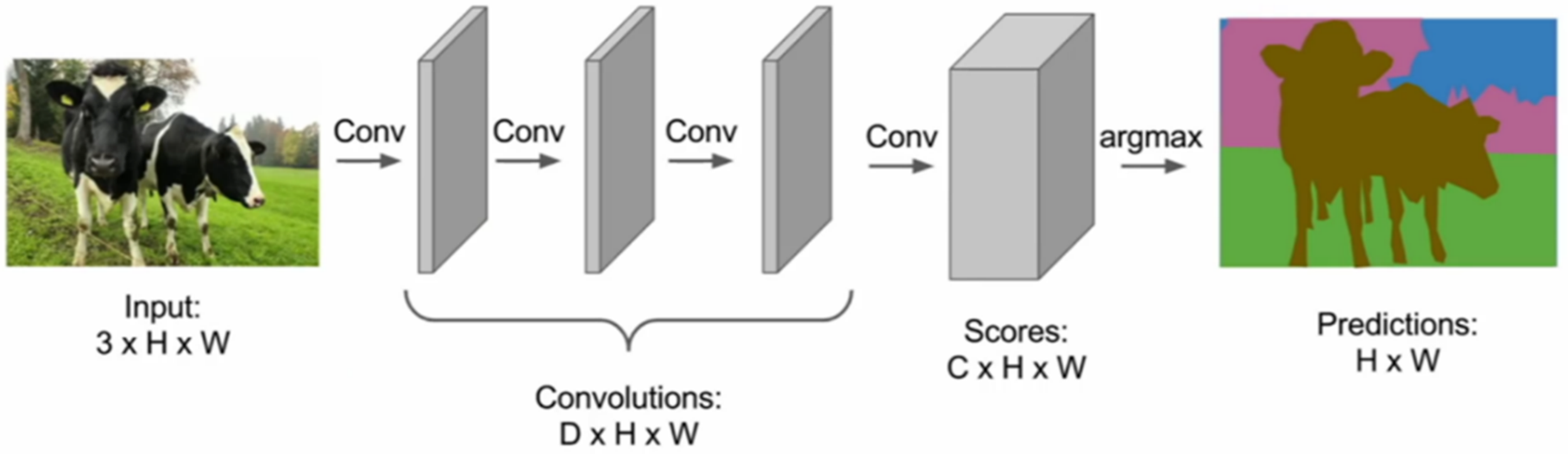
Machine learning: decision trees, support vector machines (SVM), random forests

Problems

- difficult to handle features in medical images than normal RGB images
- due to noisy, blur, variability in image resolution and contrast

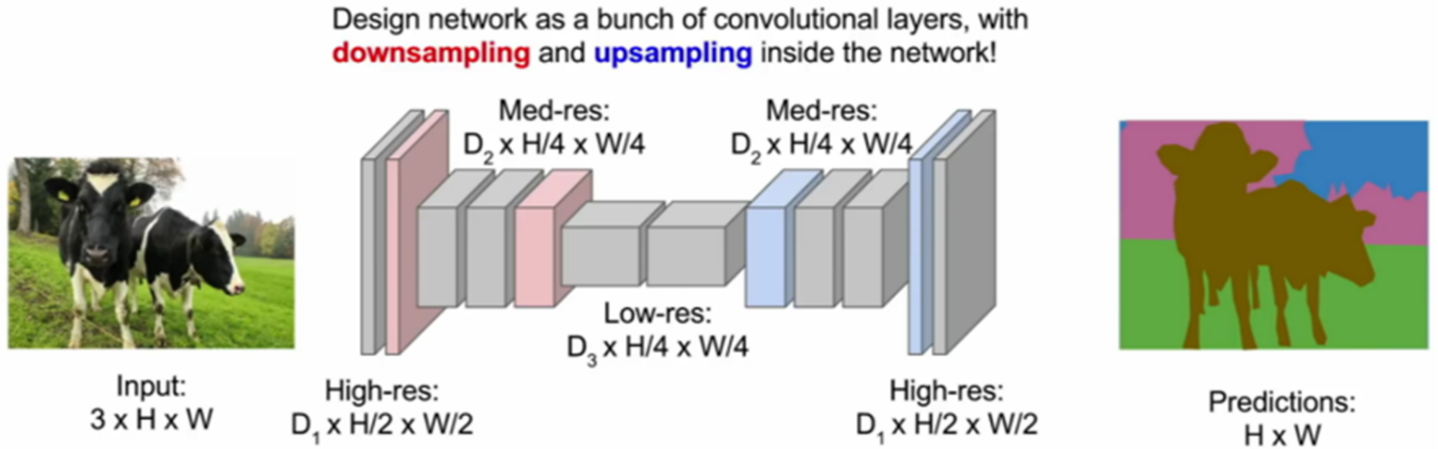
FCN (Fully Convolutional Network) (CVPR,2015)

Design a network as a bunch of convolutional layers
to make predictions for pixels all at once!



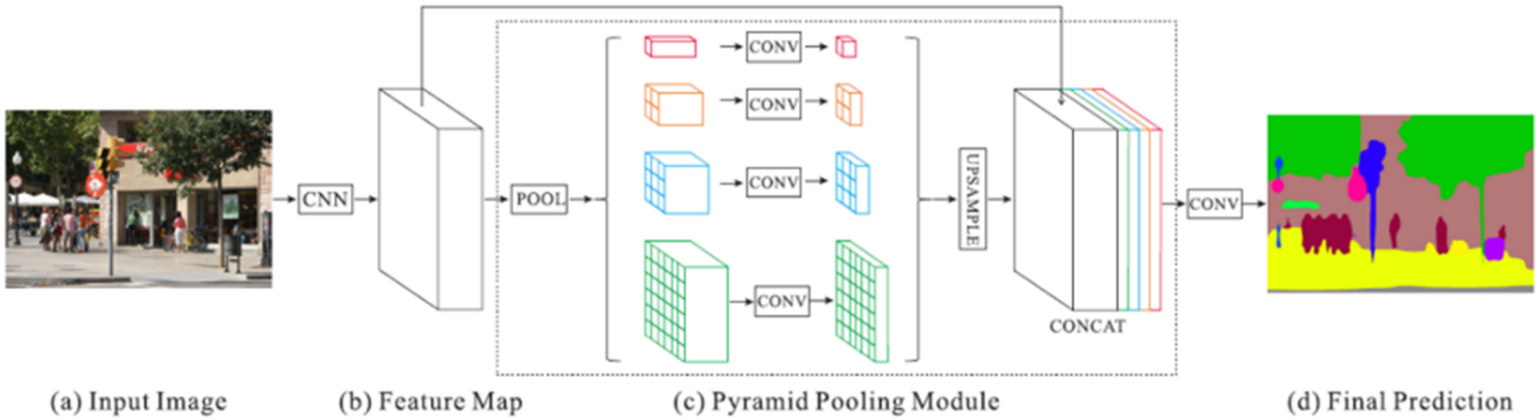
CNN based Image Segmentation Models

SegNet (IEEE transactions on pattern analysis and machine intelligence,2017)



CNN based Image Segmentation Models

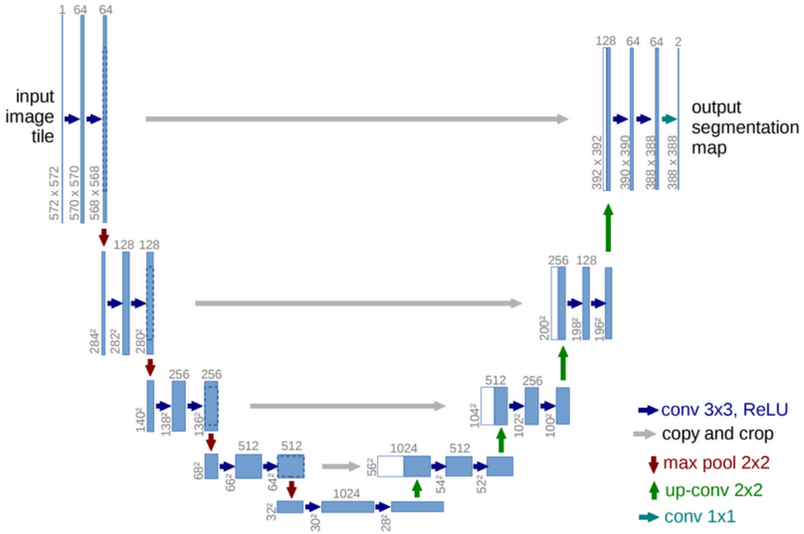
PSPNet (Pyramid Scene Parsing Network) (CVPR, 2017)



(a) Input Image (b) Feature Map (c) Pyramid Pooling Module (d) Final Prediction

CNN based Encoder-Decoder Models

U-Net

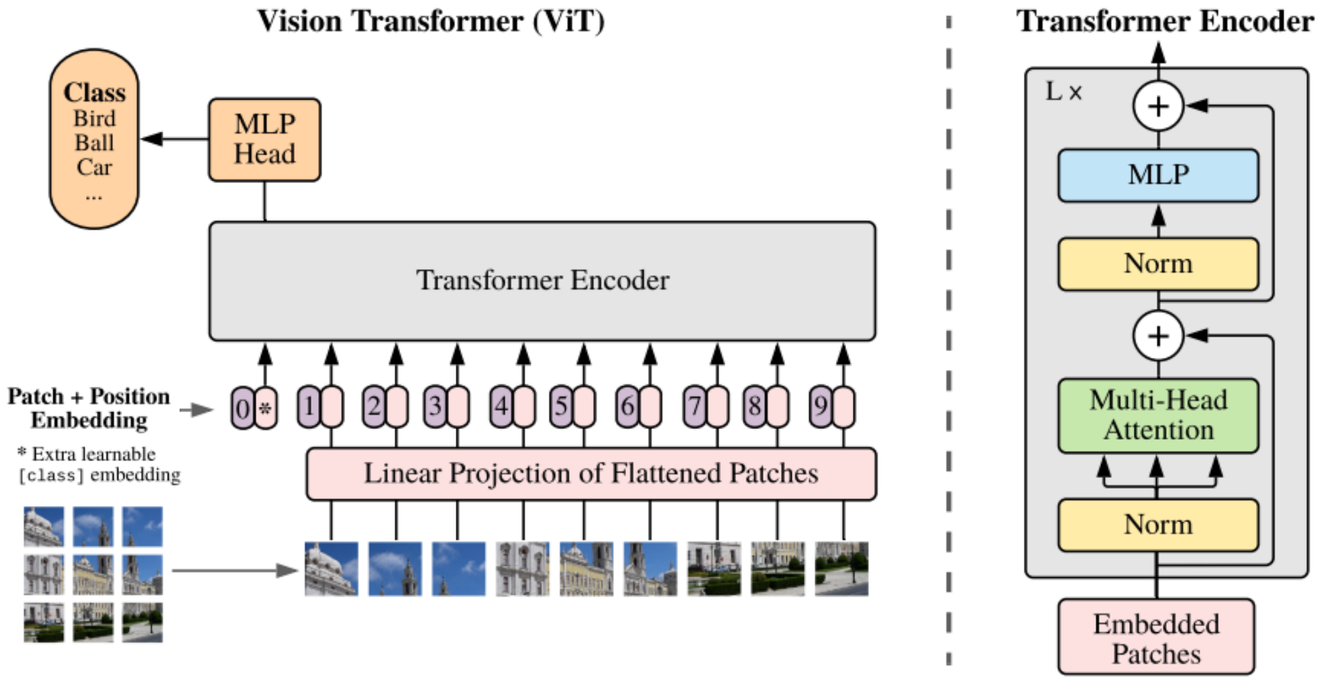


- An encoder-decoder structure
- Combines high level and low level feature maps using skip connections
- perfect for medical image segmantation tasks

Others ---> ***3D-Net,*** ***V-Net***

Attention Mechanism in U-Net ---> ***Attention U-Net*** ***FocusNet*** ***FocusNet++***

Vision Transformer - ViT



- pure transformer based architecture
- divides the image into smaller patches
- uses positional encoding to capture the content and position of each patch

Transformer-based Models

#	Name	Architecture	Pros	Cons
1	ViT	Pure Transformer	• State of art Performance on vision tasks • beats traditional CNN	• Computationally expensive • Data Hungry
2	CvT	Hybrid Transformer-CNN	• computationally efficient compared to ViT • employs all the benefits of CNNs + Transformers	• may be less effective at capturing long-range dependencies • still requires large amount of data
3	CCT	Pure Transformer	• Fewer parameters • works well with even small datasets	• lower performance on large datasets
4	Swin Transformer	Pure Transformer	• enables multi-scale feature learning • good for multiple vision tasks like classification & object detection	• sensitive to window size • require longer training time due to patch merging and splitting

Transformer + Unet based Models

SwinUNet ***TransUNet*** ***LeViT-UNet*** ***DS-TransUnet*** ***nnFormer***

DFormer

- advantages of both models and achieve better segmentation
- replacing encoder or decoder of Unet with a Transformer

Problems

Uses convolution layers as foundation

OR

seeks to integrate Transformer's ability for long range semantic

OR

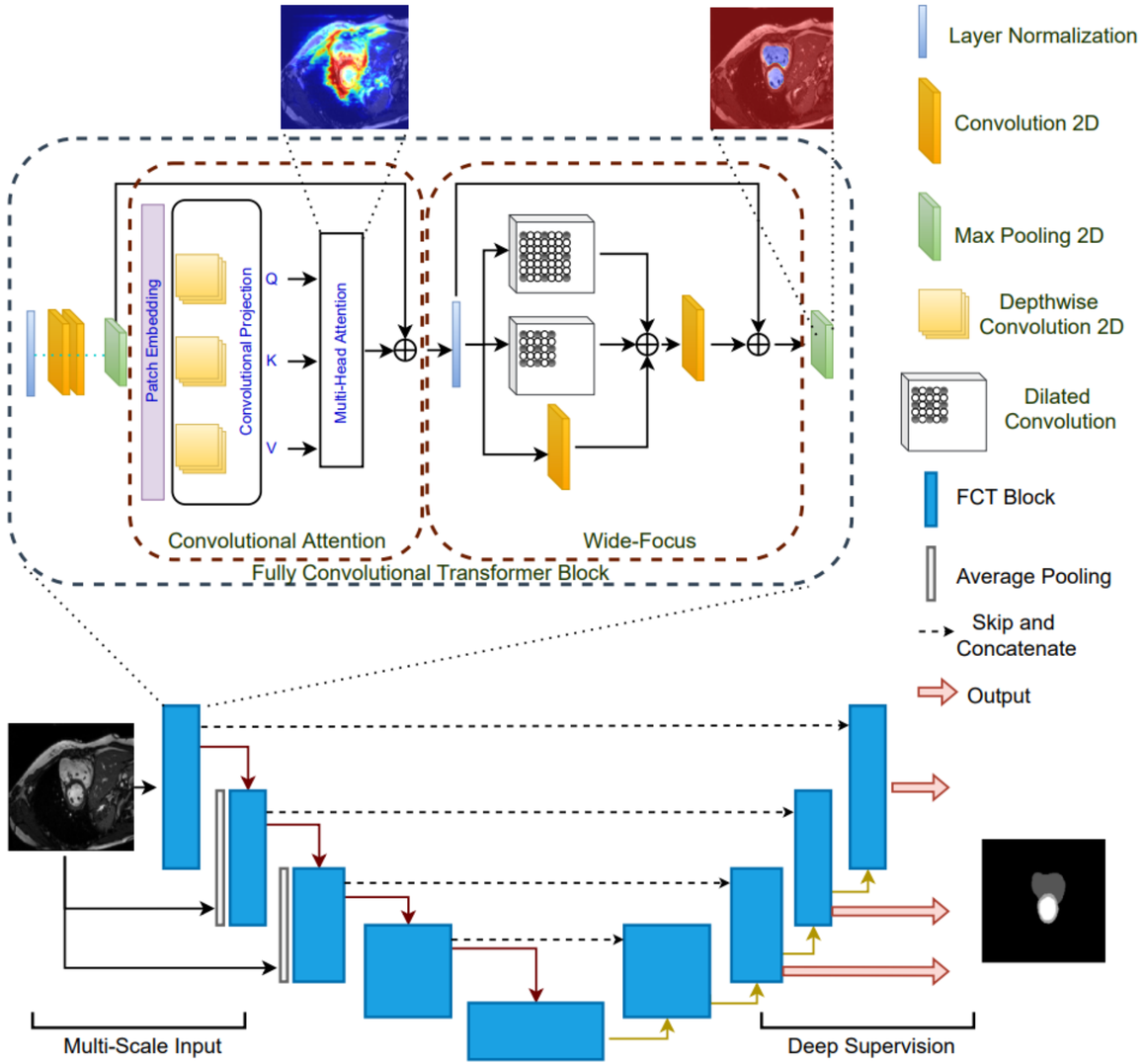
uses pure transformer without modeling local spatial context at a low level feature extraction

Fully Convolutional Transformer

Neither Transformer-CNN hybrid nor Transformer-UNet

Uses *FCT layer* as building block

- consists of convolutional layers followed by Gelu activation function
- convolutional attention module replacing linear projection with Depthwise-Convolutions, removes positional encoding
- wide focus module contains dilated convolutions and convolutional layer for feature aggregation



감사합니다

Any Question?