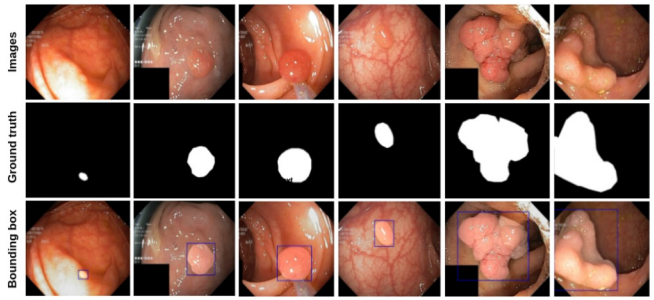
3099704: AI for Digital Health

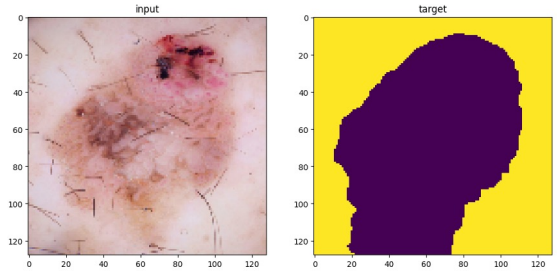
Semantic Segmentation

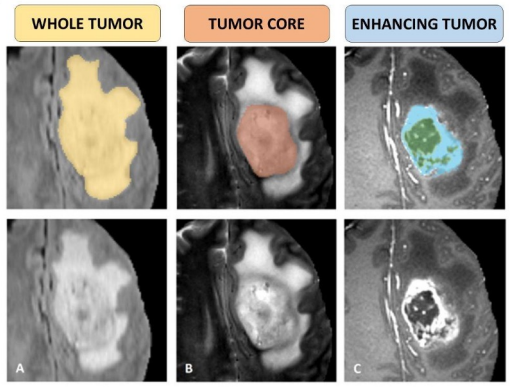
Prof. Peerapon Vateekul, Ph.D.
Peerapon.v@chula.ac.th

Outline

- Semantic Segmentation
  - UNet
  - Common Pretrained Datasets
  - Evaluation Metrics
- Case Studies
- Modern Trends: Beyond CNNs







Semantic Segmentation

Imaging Tasks

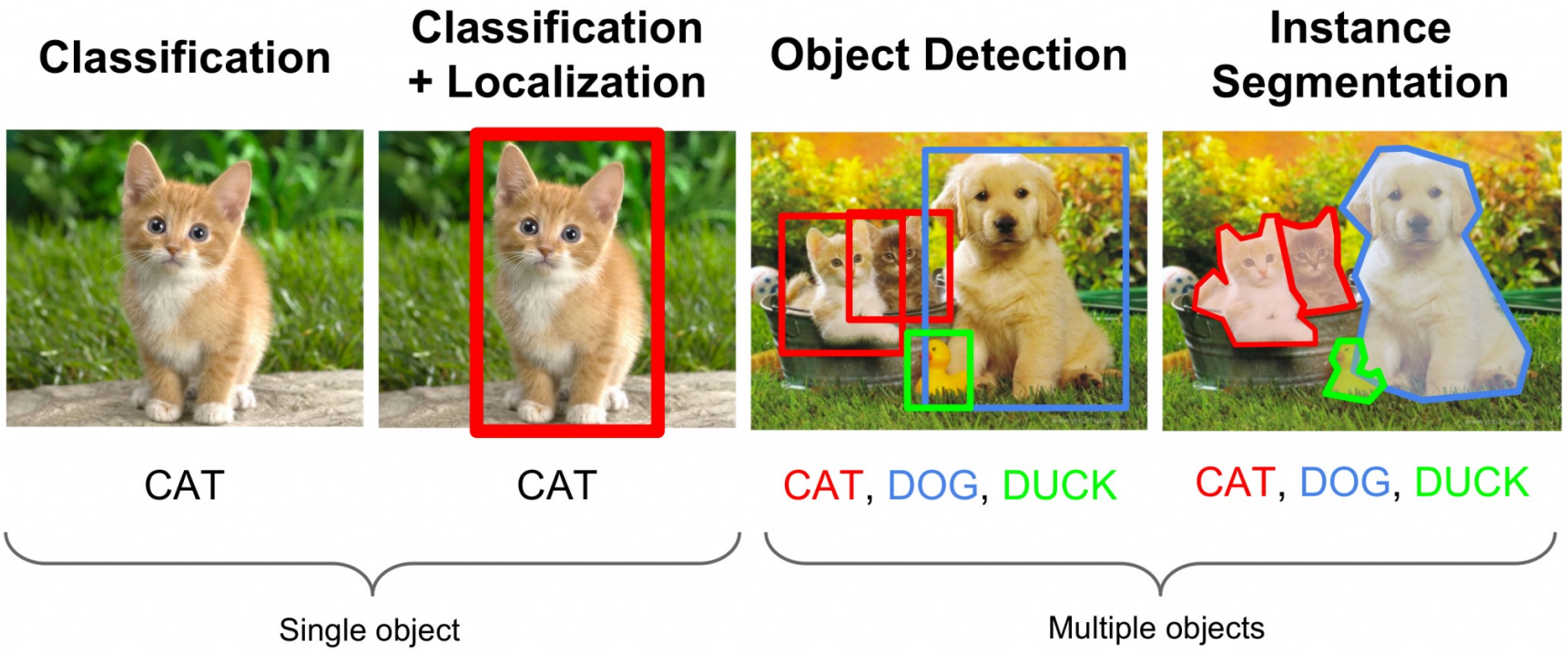
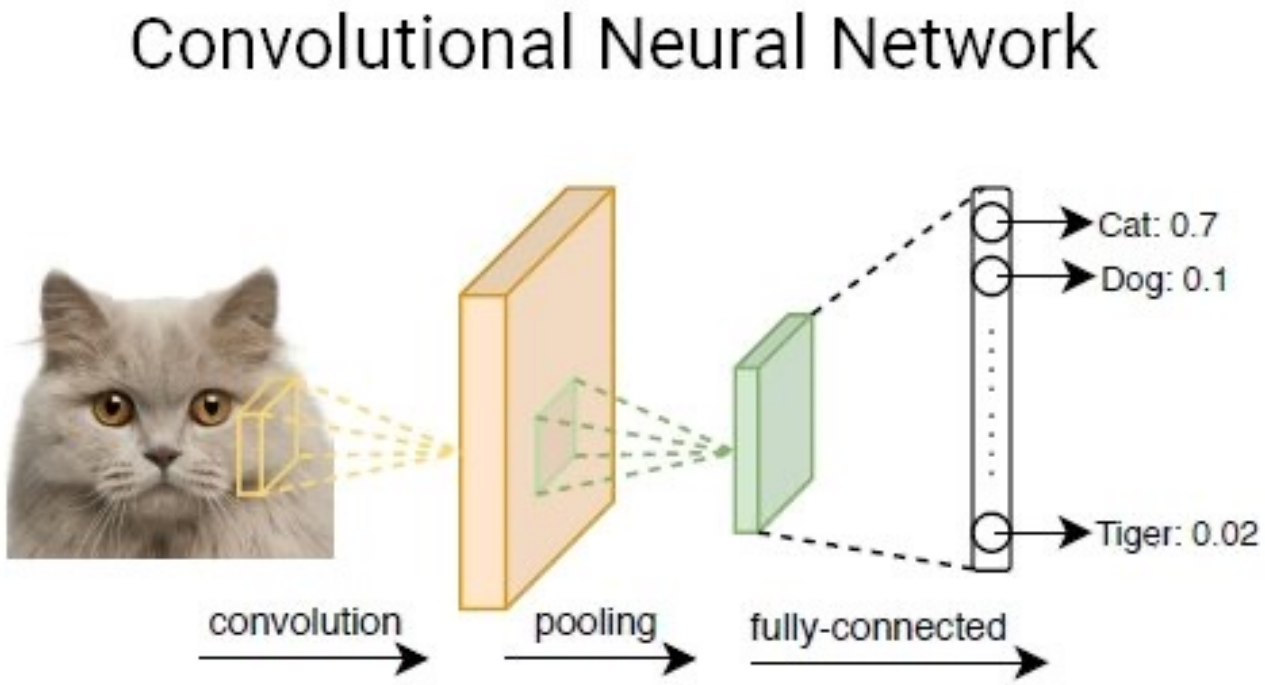
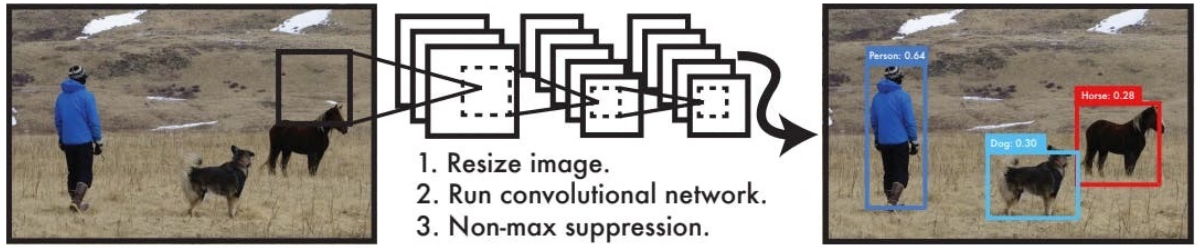


Image Classification (recap)



Object Detection (recap): YOLO



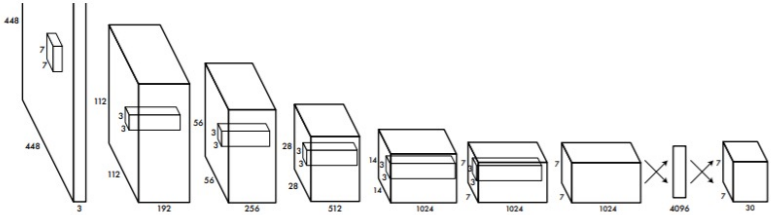
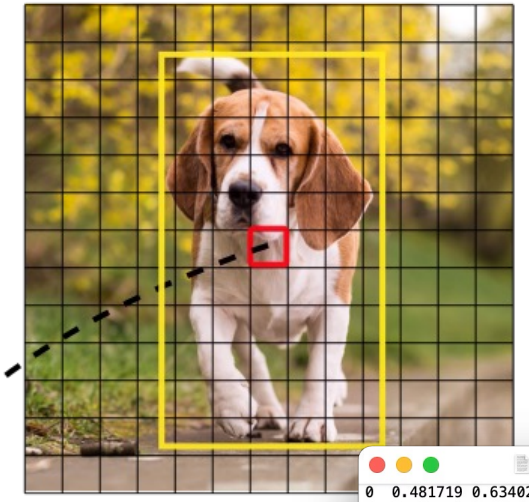


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

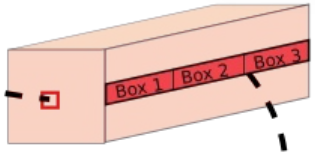
Image Grid. The Red Grid is responsible for detecting the dog



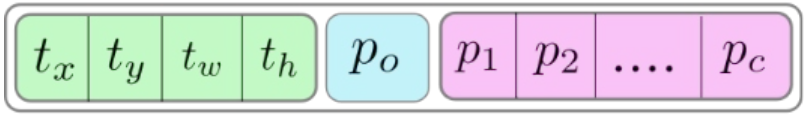
zidane.txt

```
0  0.481719 0.634028 0.690625 0.713278
0  0.741094 0.524306 0.314750 0.933389
27 0.364844 0.795833 0.078125 0.400000
```

Prediction Feature Map



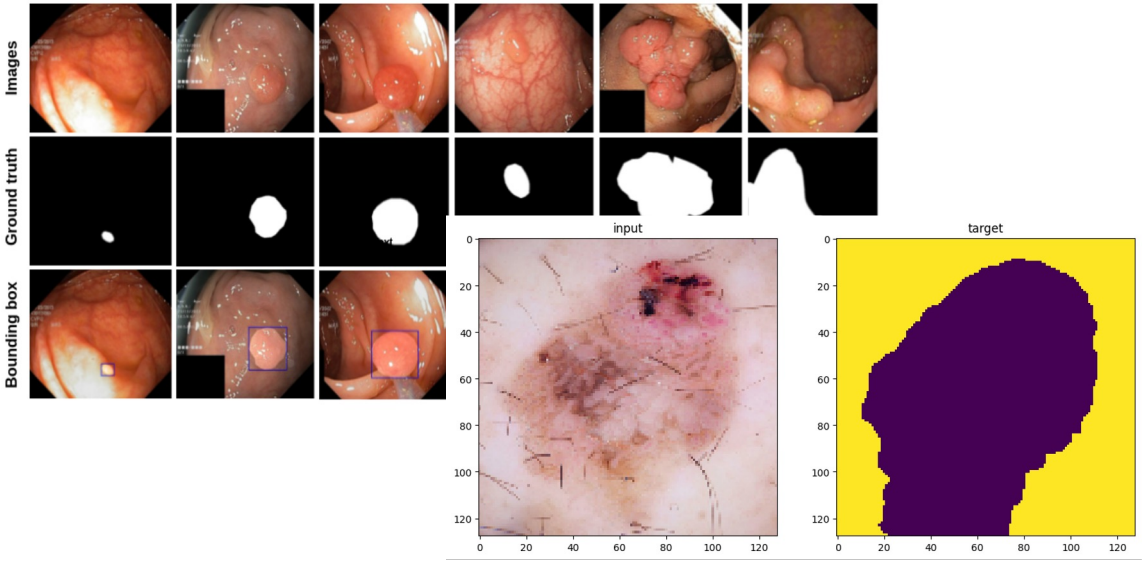
Attributes of a bounding box

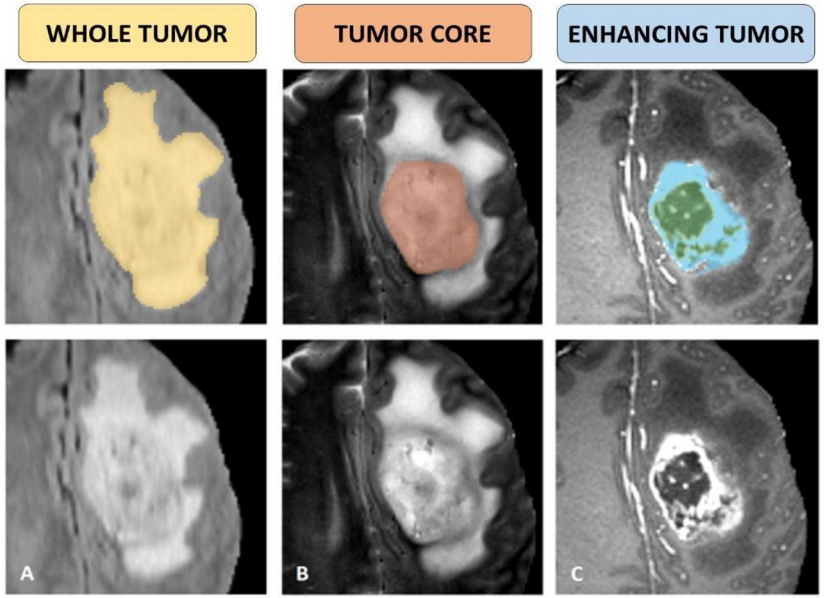


https://blog.paperspace.com/how-to-implement-a-yolo-object-detector-in-pytorch/

Semantic Segmentation

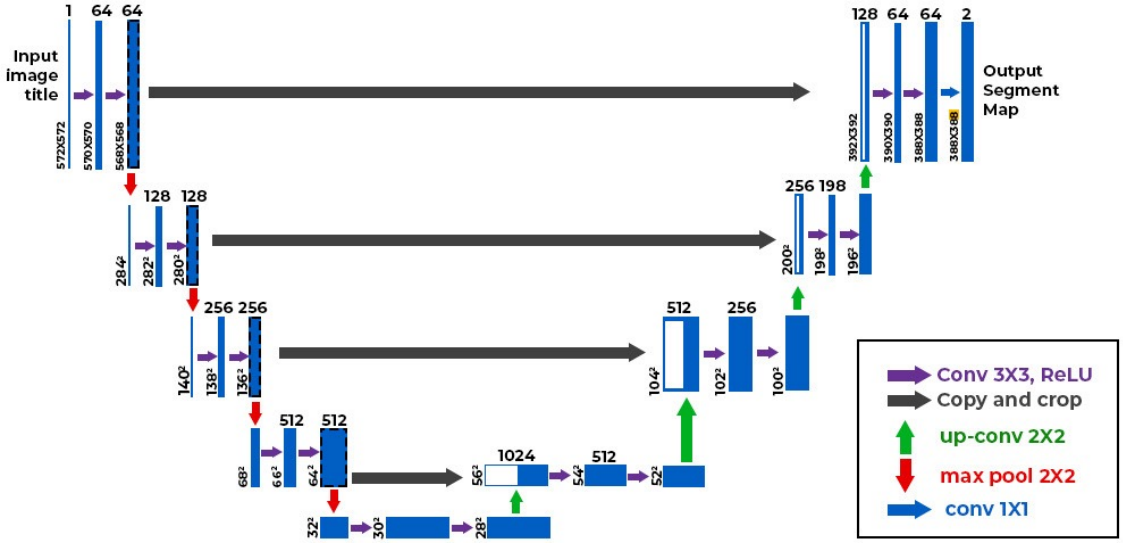
- Semantic segmentation = pixel-wise classification
- Each pixel gets a class label (road, car, polyp, tumor, sky, background).

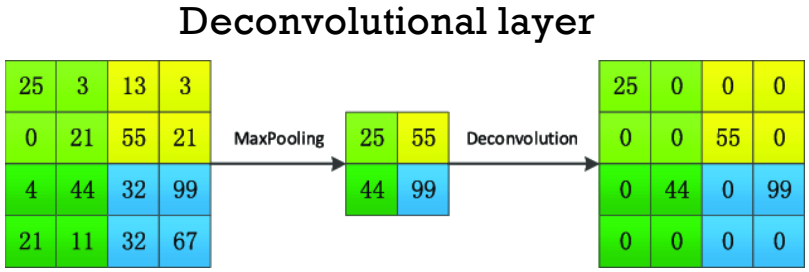




UNet: Encoder-Decoder Network: Encoder, Decoder, Skip Connections

- **U-Net**





$\mathcal{L} = \lambda_1 \cdot \mathrm{CE} + \lambda_2 \cdot \mathrm{Dice}$; CE (pixel loss) & Dice (region loss; semantic)

Architecture of the U-net for a given input image. The blue boxes correspond to feature maps blocks with their denoted shapes. The white boxes correspond to the copied and cropped feature maps.
Source: O. Ronneberger et al. (2015)

Pretrained Datasets

Model family	Typical pretraining
YOLOv8-Seg	COCO
DeepLab	COCO → Cityscapes
U-Net (medical)	MSD / domain-specific
ViT-Seg	ADE20K
SAM	Massive private + public mix

Bundle name	Dataset	Task
spleen_ct_segmentation	MSD	Spleen
liver_ct_segmentation	MSD	Liver
pancreas_ct_segmentation	MSD	Pancreas
multi_organ_ct_segmentation	BTCV	Abdominal organs

MONAI

Pretrained Datasets: MSD

- Medical Segmentation Decathlon is a large-scale benchmark designed to evaluate general-purpose medical image segmentation models across multiple organs, modalities, and tasks.

- Key Characteristics
  - Modalities: CT, MRI
  - Dimensionality: 3D volumes
  - Tasks: 10 diverse segmentation challenges

- Example Tasks
  - Brain tumor (MRI)
  - Liver, spleen, pancreas (CT)
  - Lung tumor (CT)
  - Colon cancer (CT)
  - Heart & prostate (MRI)

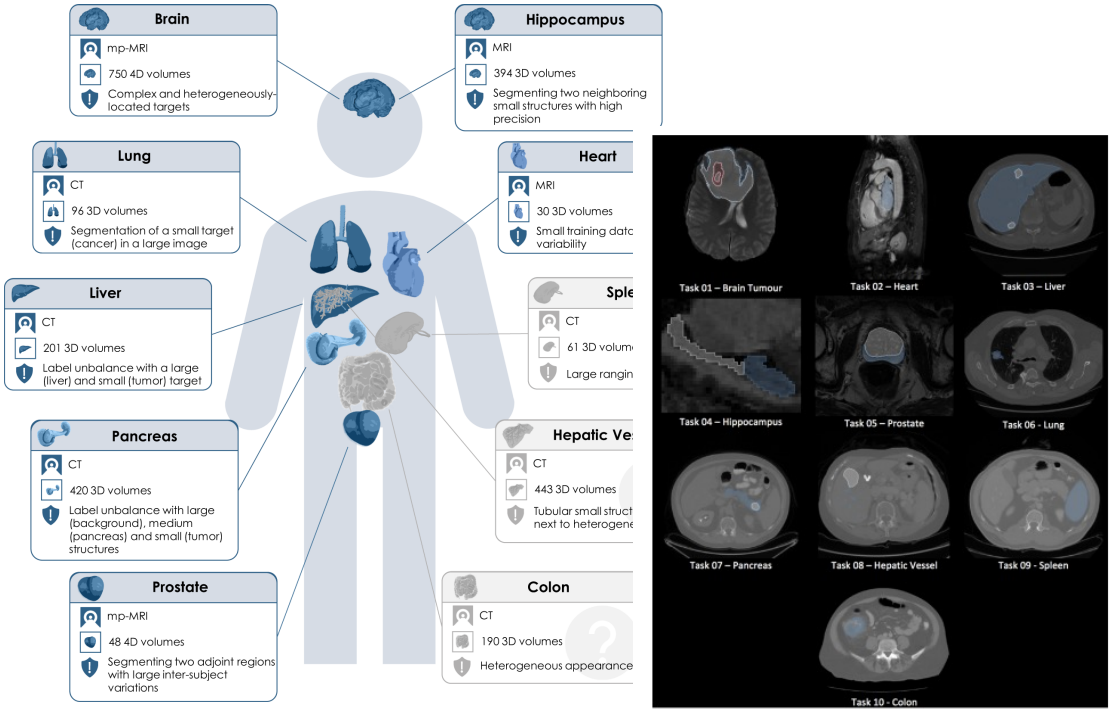
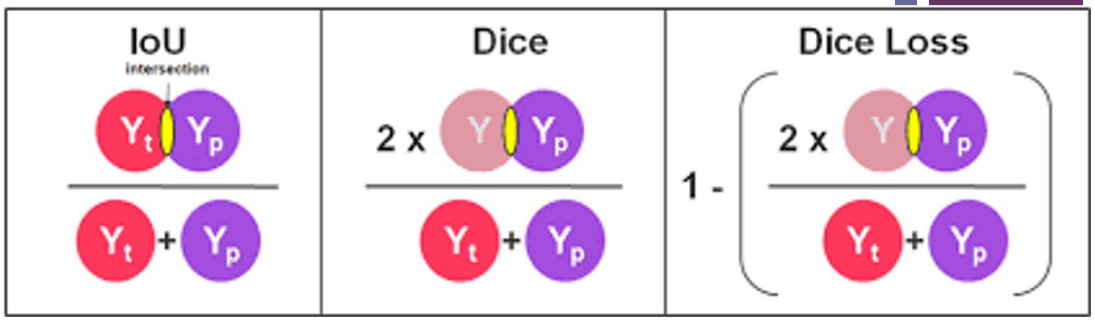


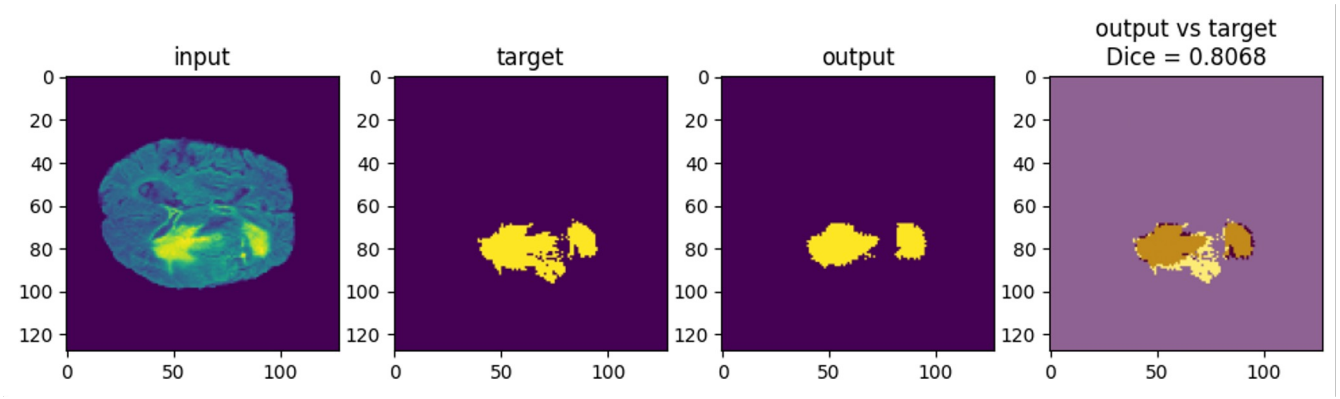
Figure 1: Exemplar images and labels for each dataset. Blue, white, and red correspond to labels 1, 2, and 3, respectively, of each dataset. Not all tasks have 3 labels.

Evaluation Metrics

Metric	Meaning
Pixel Accuracy	Easy but misleading
Mean IoU (mIoU)	Standard benchmark
Dice / F1	Popular in medical



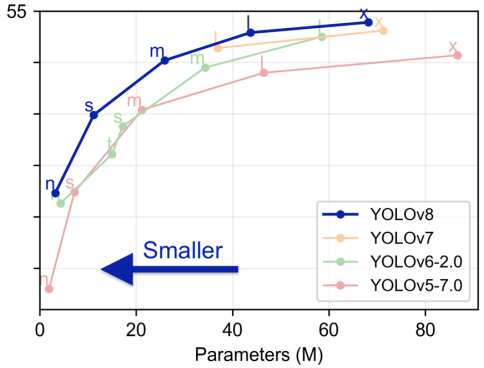
$$\text{IoU} = \frac{TP}{TP + FP + FN} \qquad \text{Dice} = \frac{2TP}{2TP + FP + FN} \qquad \text{Dice} = \frac{2 \cdot \text{IoU}}{1 + \text{IoU}}$$

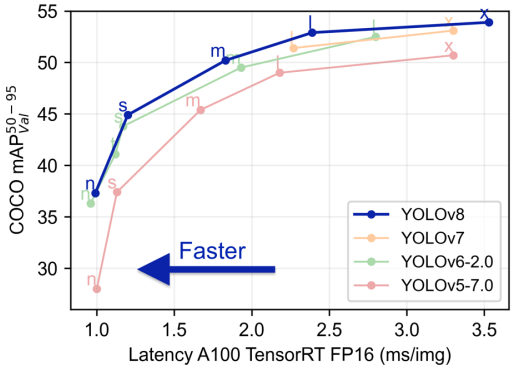


Mean Dice: 0.7252

Ultralytic's YOLO (extra segmentation model)

YOLOv5 (2020) → YOLOv8 (2023) → Ultralytics YOLO 11 (2024)





English | 简体中文

Ultralytics CI passing | codecov 87% | DOI 10.5281/zenodo.7347926 | docker pulls 22k

Run on Gradient | Open in Colab | Open in Kaggle

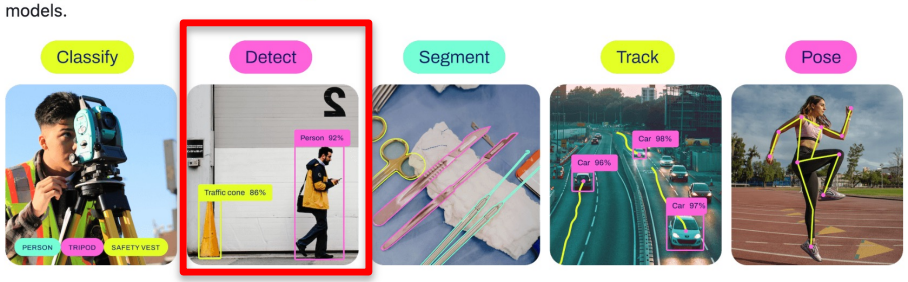
Ultralytics YOLOv8 is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility. YOLOv8 is designed to be fast, accurate, and easy to use, making it an excellent choice for a wide range of object detection and tracking, instance segmentation, image classification and pose estimation tasks.

We hope that the resources here will help you get the most out of YOLOv8. Please browse the YOLOv8 Docs for details, raise an issue on GitHub for support, and join our Discord community for questions and discussions!

To request an Enterprise License please complete the form at Ultralytics Licensing.

Models

YOLOv8 Detect, Segment and Pose models pretrained on the COCO dataset are available here, as well as YOLOv8 Classify models pretrained on the ImageNet dataset. Track mode is available for all Detect, Segment and Pose models.



All Models download automatically from the latest Ultralytics release on first use.

YOLOv8-Seg: Object Detection → Segment

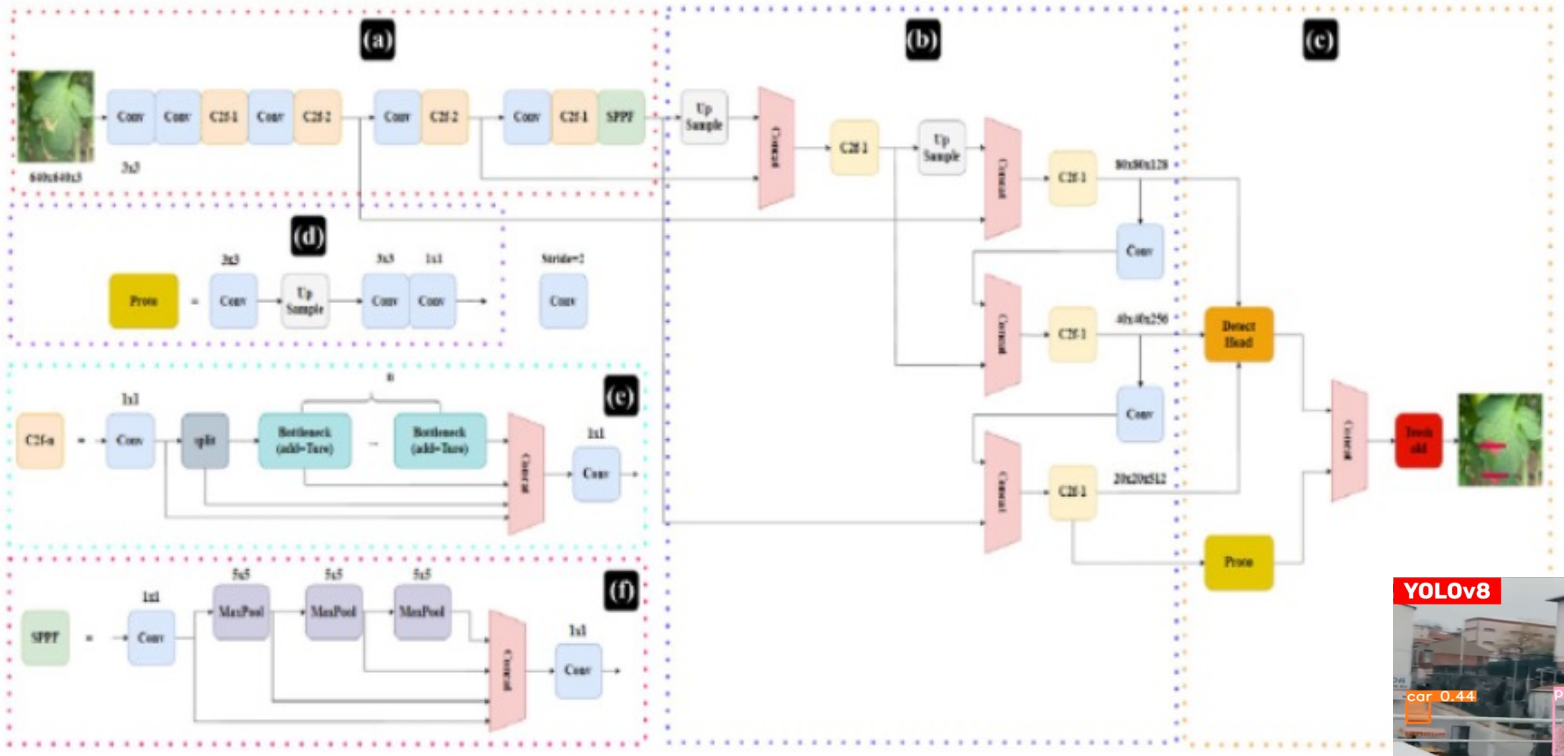
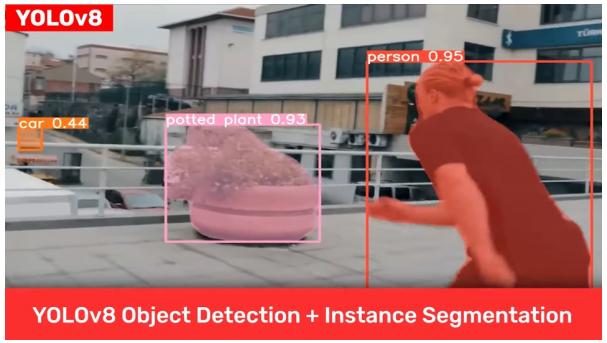


Fig. 3 YOLOv8-Seg architecture; **a** backbone, **b** neck, **c** head, **d** Proto module, **e** C2f module and **f** SPPF module



Link

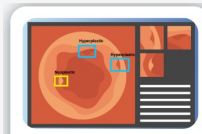
Case Studies

DeepGI: GIM Segmentation

Meticuly: Automatic Skull and Mandible Reconstruction

DeepGI: Real-Time Colonic Polyp Detection

An AI-assisted solution that aims to improve colonic polyp detection in real-time. It is compatible with all standard colonoscope systems.



Feature 1

It can support both detection and characterization. It is based on the latest deep-tech models called "Deep Learning" to achieve the best detection rate, which is higher than 95%. The most challenging is that DeepGI can analyze the endoscopy video in real-time. Thus, the endoscopists can seamlessly improve their diagnosis performance.

Feature 2

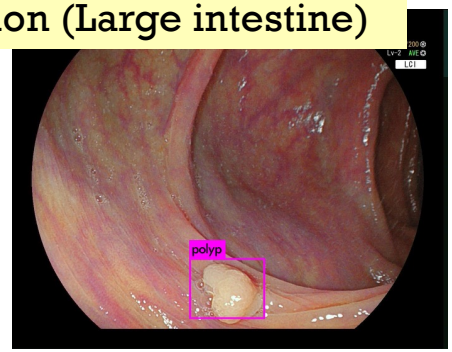
In the operation room, there can be various kinds of endoscopy cameras from different vendors among with different versions. DeepGI can support all those endoscopy cameras. Thus, DeepGI can be easily integrated into the existing system in the operation room.

Feature 3

Apart from the polyp detection in colonoscopy, DeepGI can also support other endoscopy operations in real-time: gastroscopy and cholangioscopy. For the gastroscopy, DeepGI can segment regions Gastrointestinal Metaplasia (GIM) in stomach. For the cholangioscopy, it can classify the benign area in real-time. Thus, DeepGI is a complete solution for the whole GI tracts.

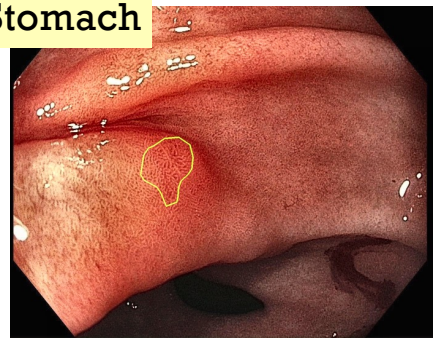
AI assisted solution in many parts of GI tracts

Colon (Large intestine)



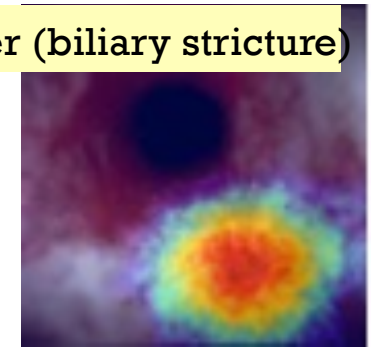
Real-time Polyp Detection on Colonoscopy Videos

Stomach



Real-Time Gastrointestinal Metaplasia (GIM) Segmentation on Gastroscopy Videos

Liver (biliary stricture)

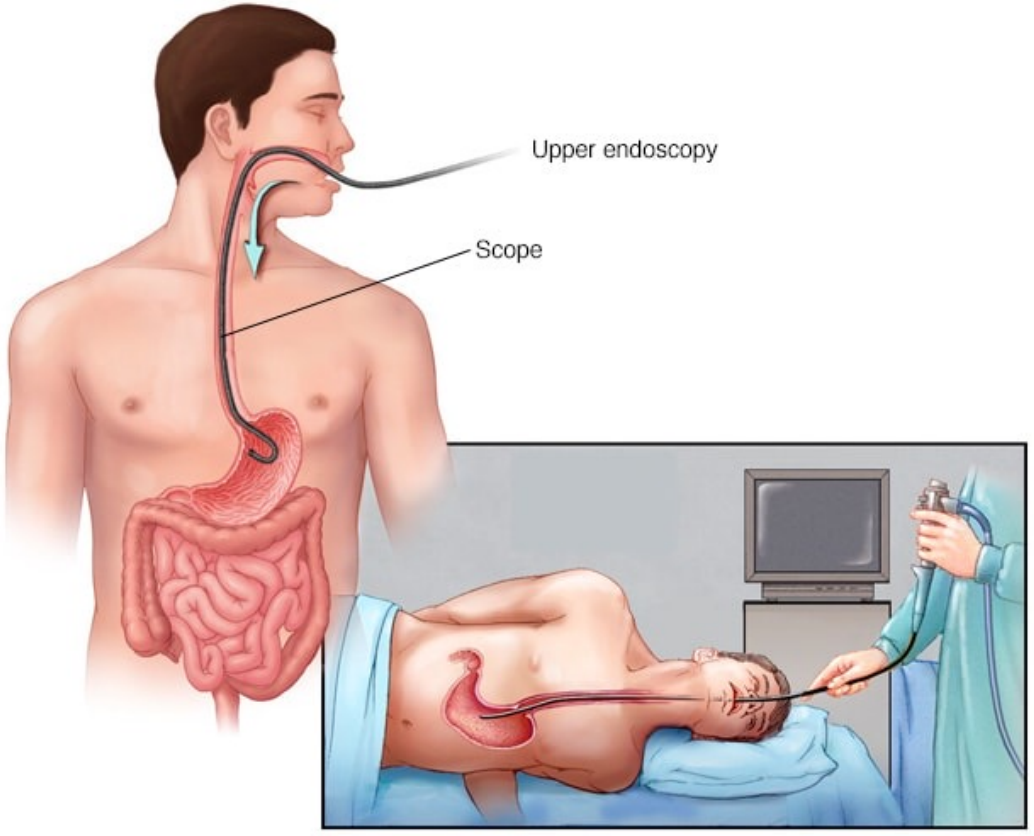


Real-Time Image Classification for Malignant Biliary Strictures on Cholangioscopy Images

Aims & objectives

Gastroscopy

Gastrointestinal metaplasia (GIM)



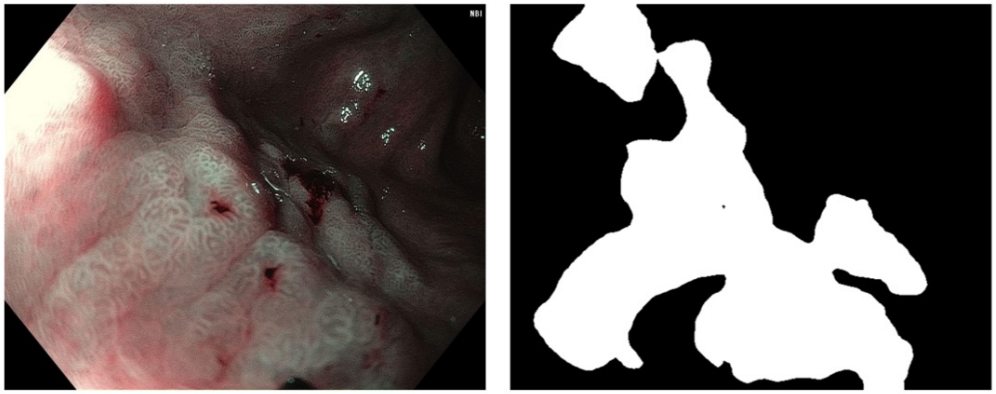


Figure: *The semantic segmentation of GIM. (A) Original image. (B) Prediction: white is GIM, and black is healthy.*

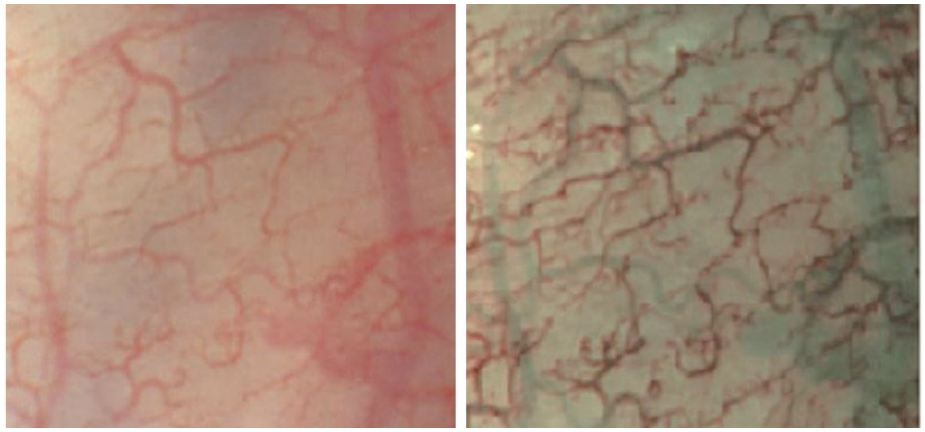


Figure: *The membrane of the human tongue. (A) White light image. (B) Narrow* band imaging.


Gastroscopy
(new model)

Automatic Skull and Mandible Reconstruction

METICULY®

https://www.meticuly.com/

Aj.Titipat Aj.Peerapon

Every year, more than **10k cases** in Thailand suffer from **skull damage** and **require implant**.

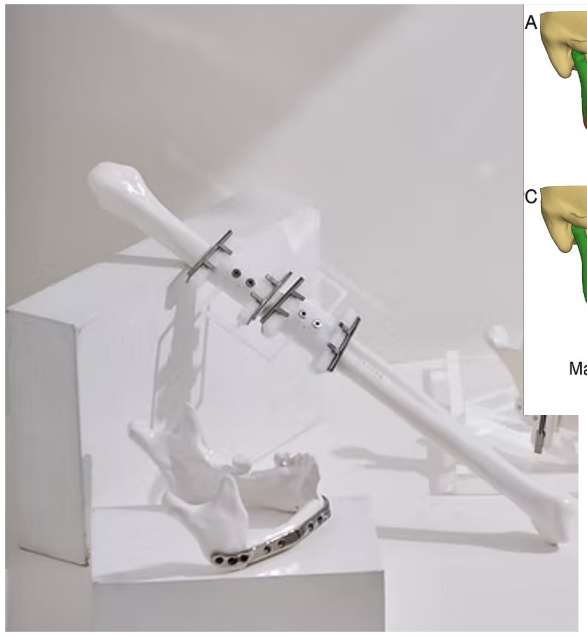
- **Meticuly** has been doing **patient-specific 3D printing**.
- Provided → 1000 personalized cranial implant (Cranioplasty)
- We apply Autoimplant and Mandible reconstruction in the pipeline

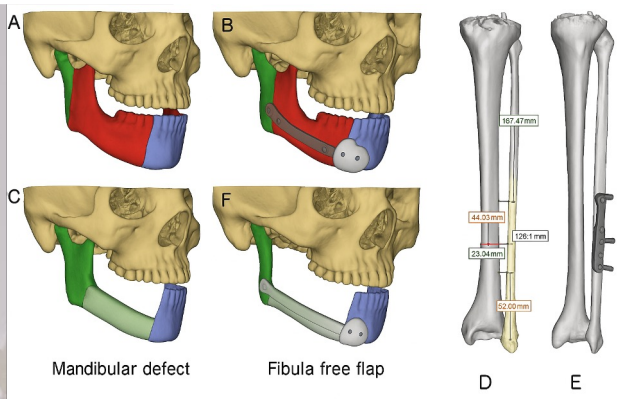
1 case → Takes ~ 1 week → AI-assisted design

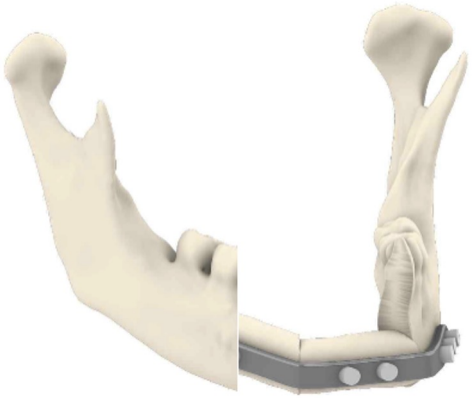
fasten the process

Credit: Aj.Titipat's slide

Background and Motivation







Similarly, thousands of patients suffer from **mandibular defects** due to cancer, trauma, or infection, which severely impacts **chewing, speech, swallowing, and facial appearance**.

Support **pre-operative planning** and **custom plate design** tailored to each patient's anatomy.

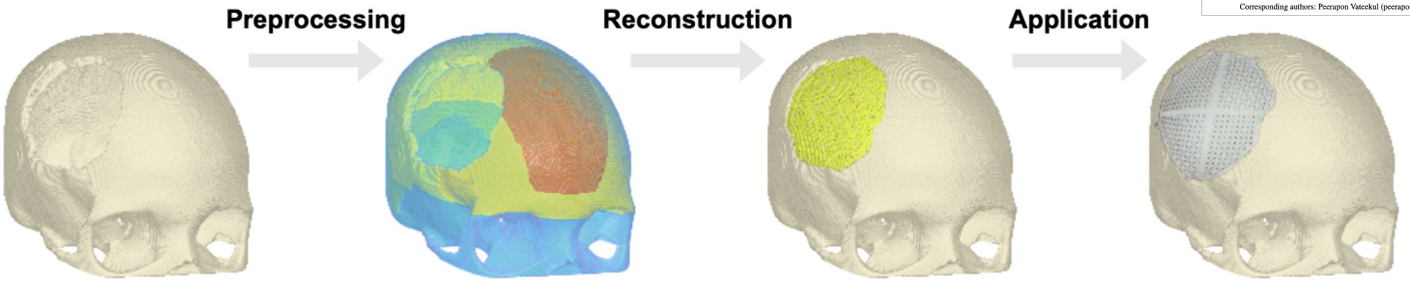
https://www.meticuly.com/

Our Aim

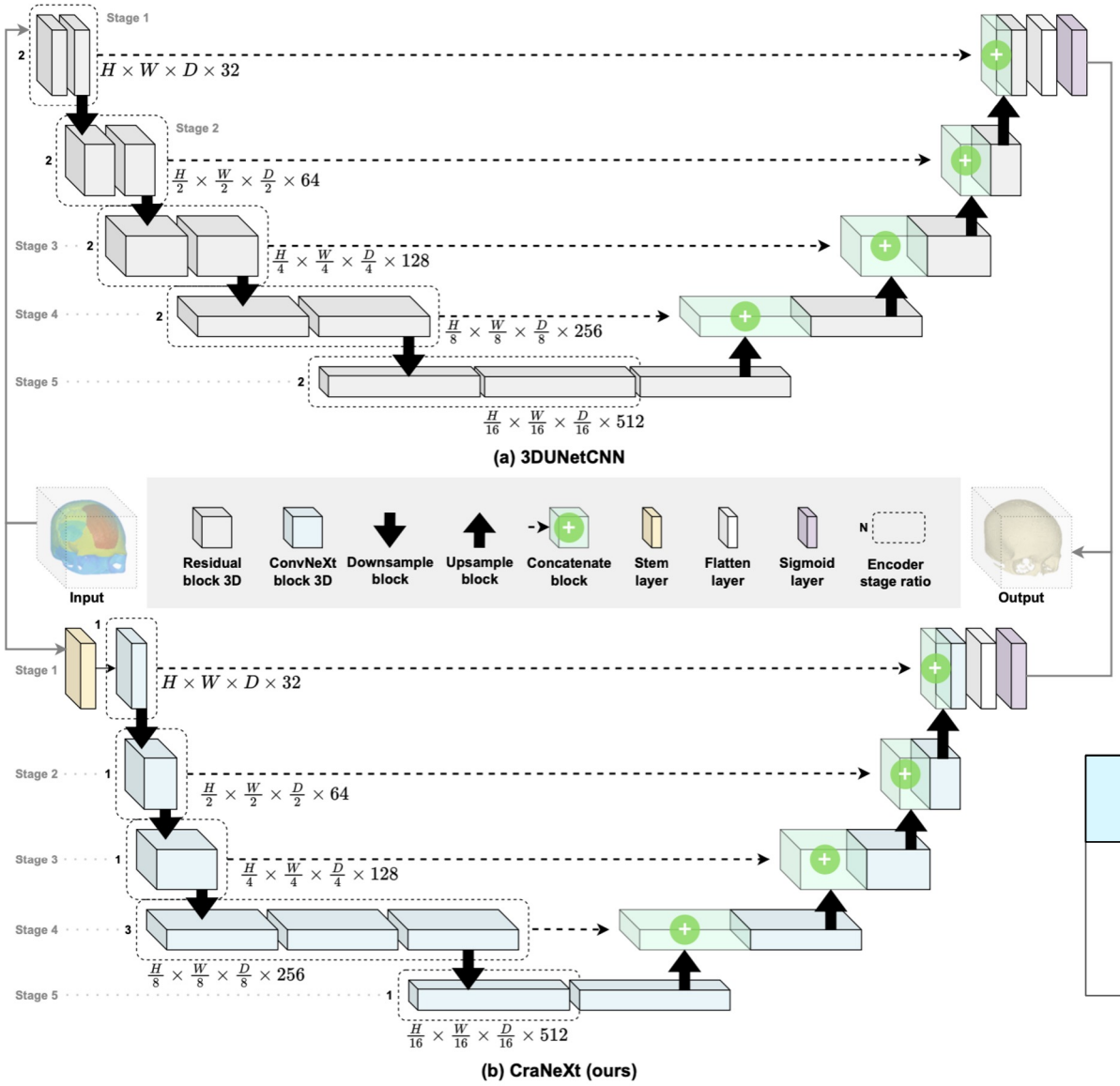
To enhance the **accuracy**, **efficiency**, and **robustness** of skull reconstruction on clinical defective skulls

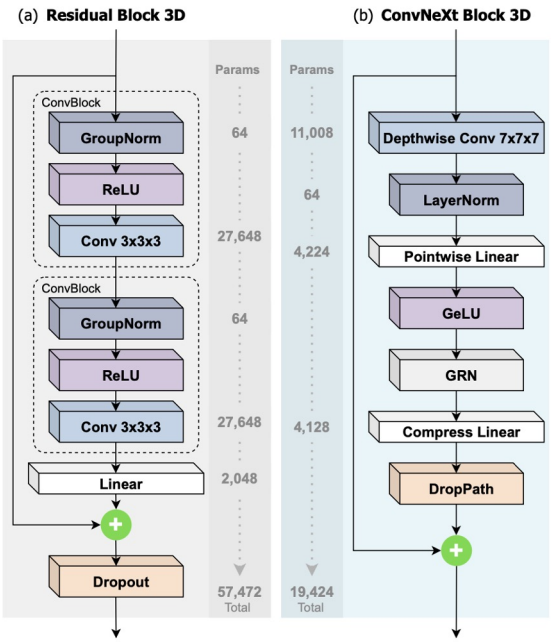
We proposed:

- Improved deep learning backbone for 3DUNet
- Analysis on both **synthetic** and **real** clinical datasets



Model

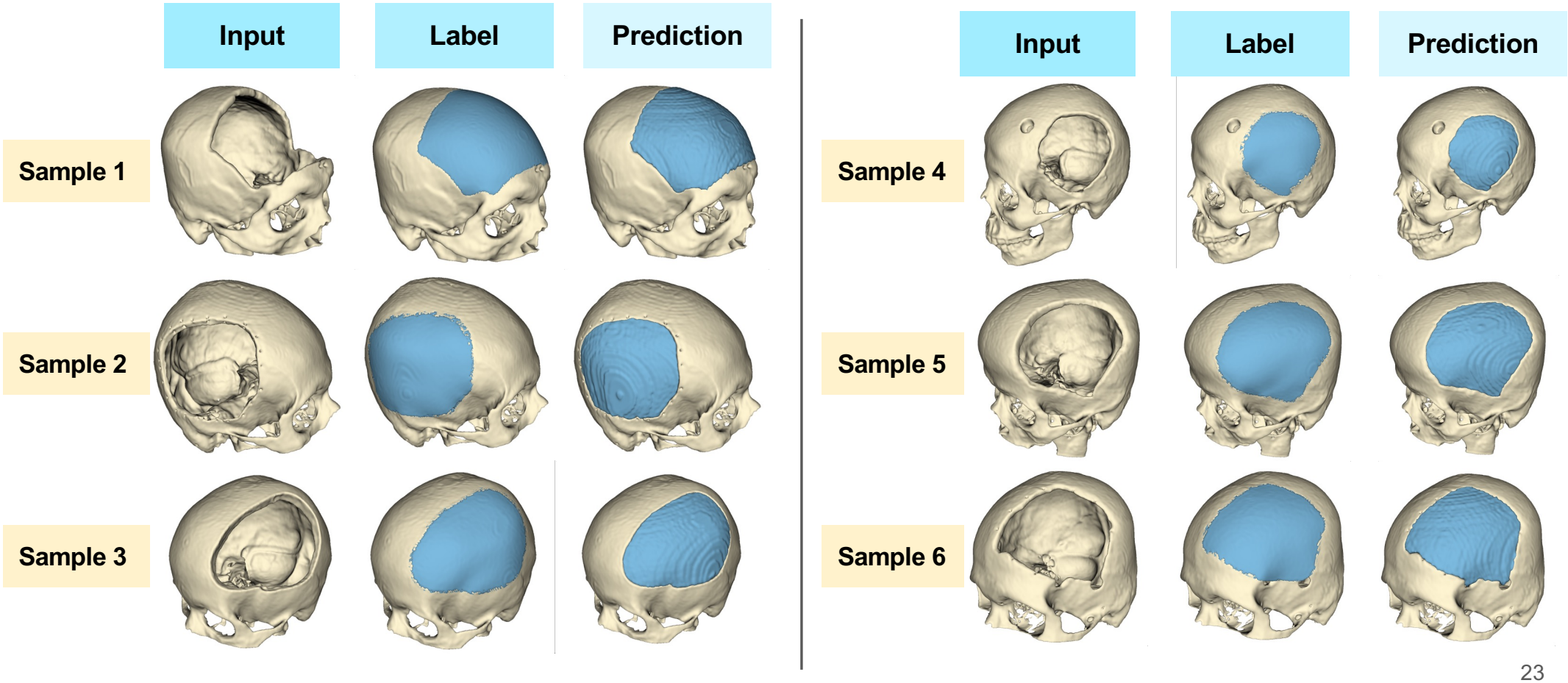




Main differences

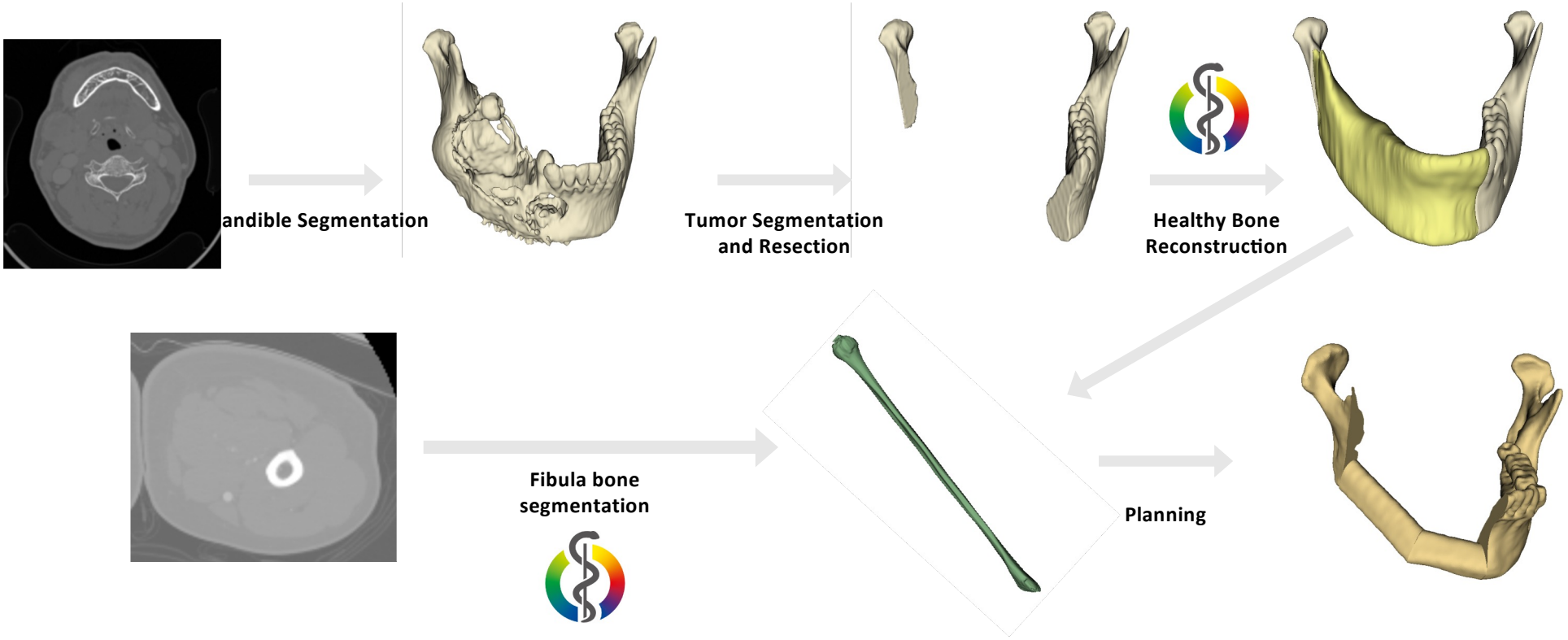
- Stage ratio
- Convolution block

Model Performance Evaluation

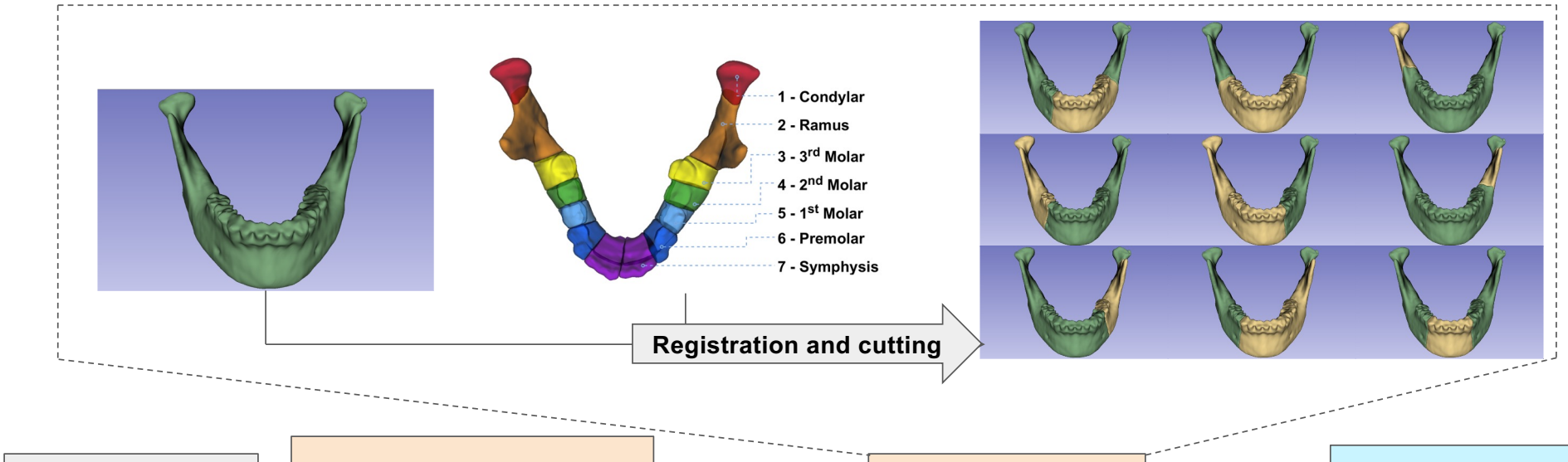


Mandible Reconstruction

- We can use a similar models for mandible reconstruction.



Dataset

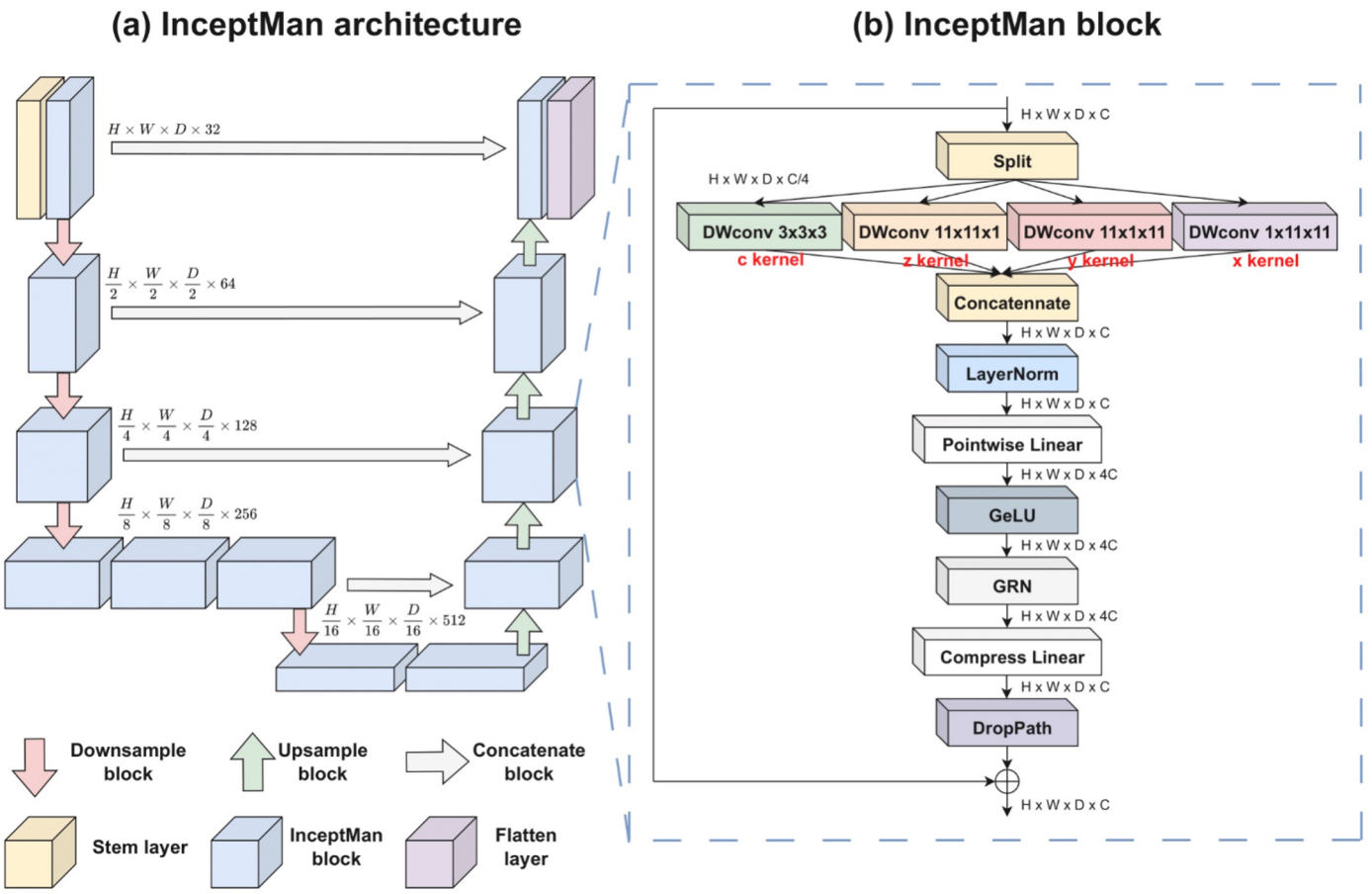


In House Dataset → **Resample, Resize, Foreground cropping** → **Augmentation** → **Model training**

90 unique mandible
Dim: 0.5 x 0.5 x 0.625 mm
Voxel: [500 -512] x [512 -530] x [200 -759]

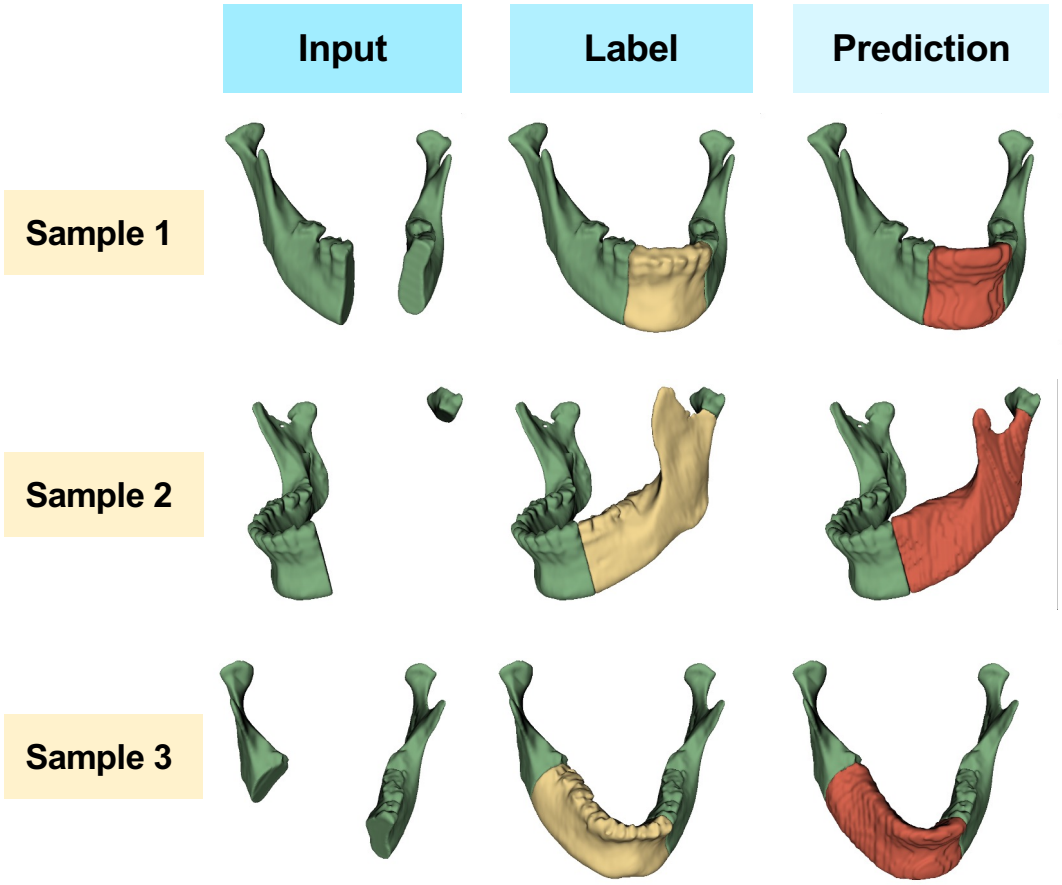
+ Defect Categorization from template
Implant: Synthetically generated

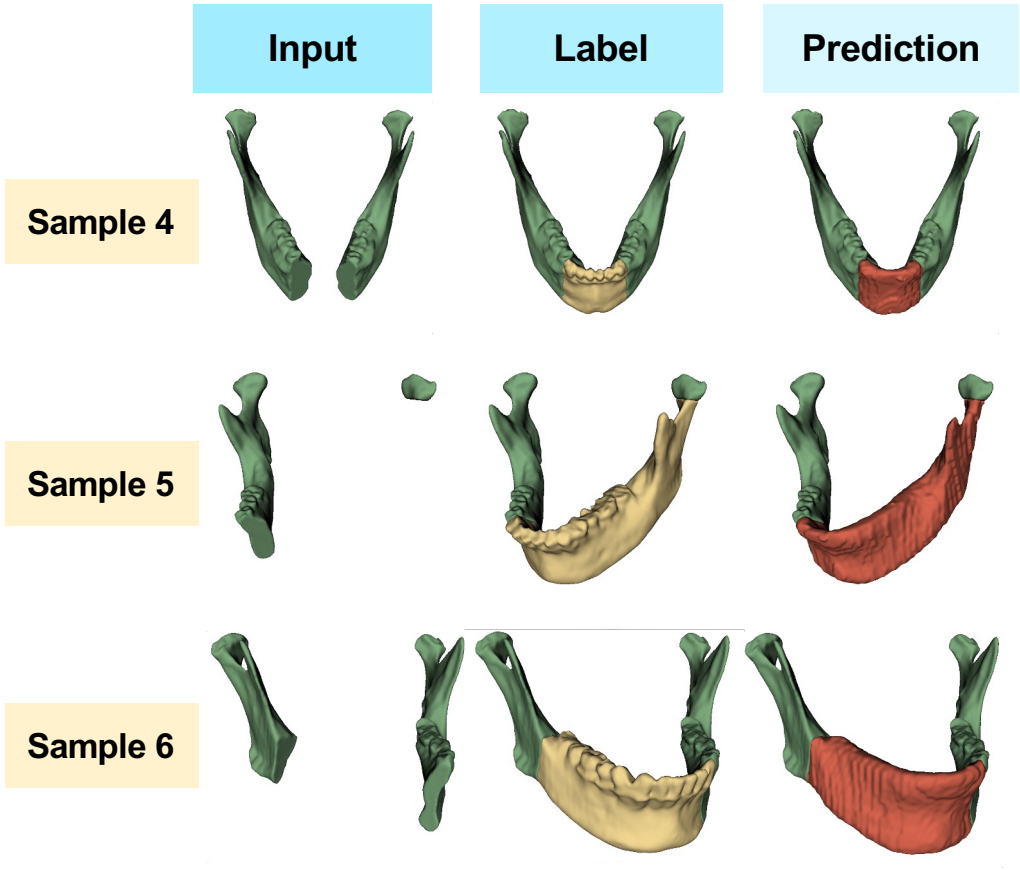
Model



Main differences
InceptMan block modified 3D ConvNeXt block with Inception block. It uses stage ratio of 1:1:1:3:1 (similar to CraNeXt).

Model Performance Evaluation





CraNext and InceptMan are an Open Source

CraNeXt

CraNeXt is a deep learning model for skull reconstruction tasks. The input is a binary voxel is a defective skull, and the output is a binary voxel representing a complete skull.

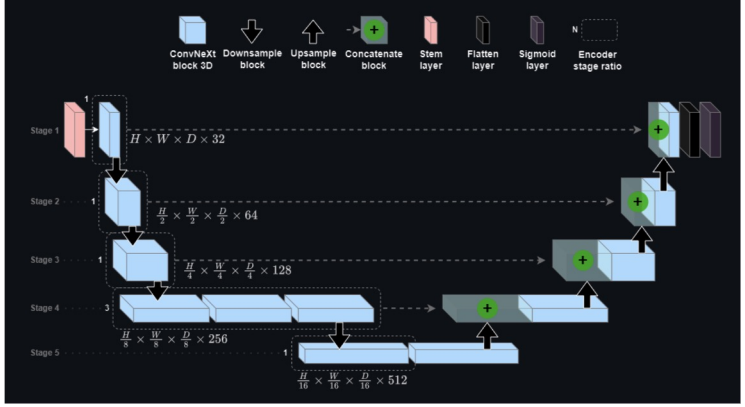
Usage Example

Open in Colab

For a detailed demonstration of how to use CraNeXt, please refer to our `example.ipynb` notebook. This notebook provides step-by-step instructions and code examples for:

- Downloading sample SkullBreak dataset and preparing input data
- Using the CraNeXt model with pretrained weights on SkullBreak
- Running skull reconstruction and performing evaluation
- Visualizing reconstruction result

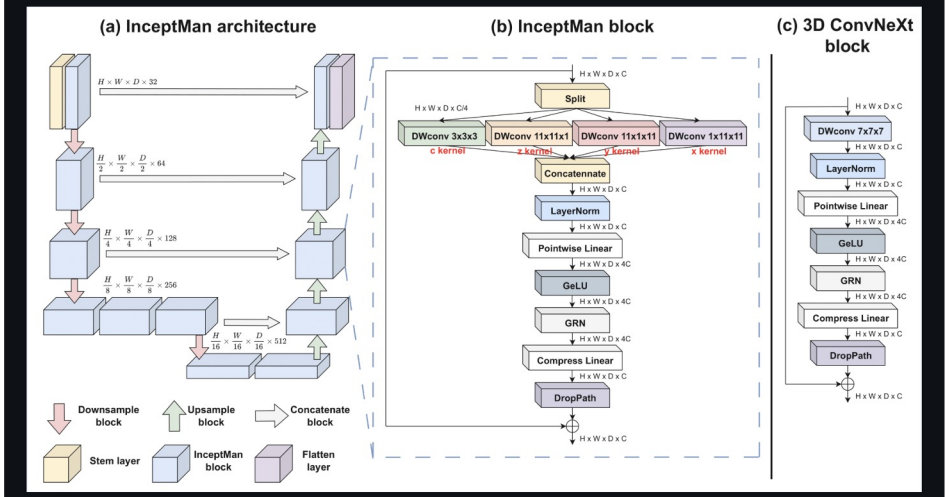
You can run the notebook directly in Google Colab by clicking the Colab badge above or locally.



Available at: https://github.com/guitared/CraNeXt

InceptMan

InceptMan is a deep learning model designed for mandible reconstruction, as proposed in our paper, "InceptMan: An InceptionNeXt-Based Architecture for End-to-End Mandible Reconstruction" The input is a binary volumetric incomplete mandible, and the output is a binary volumetric complete mandible.

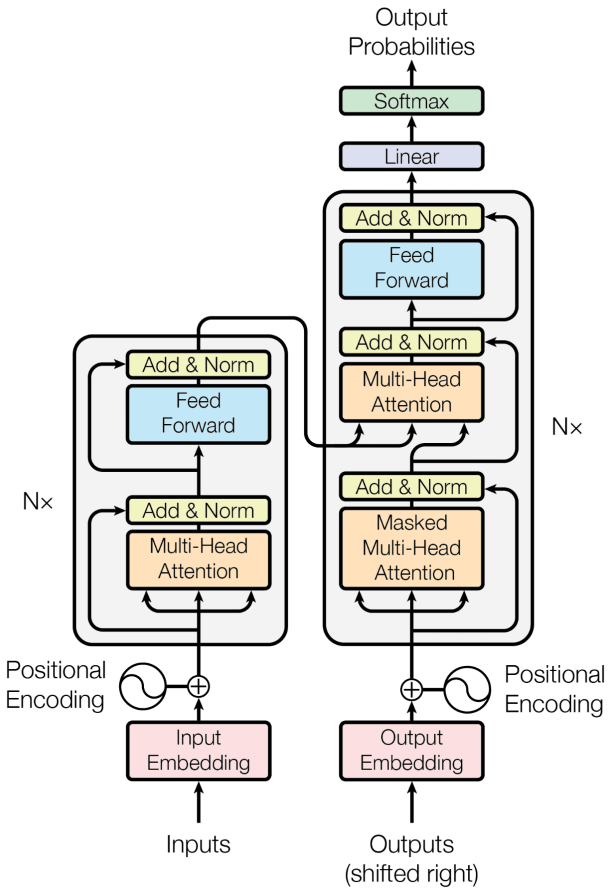


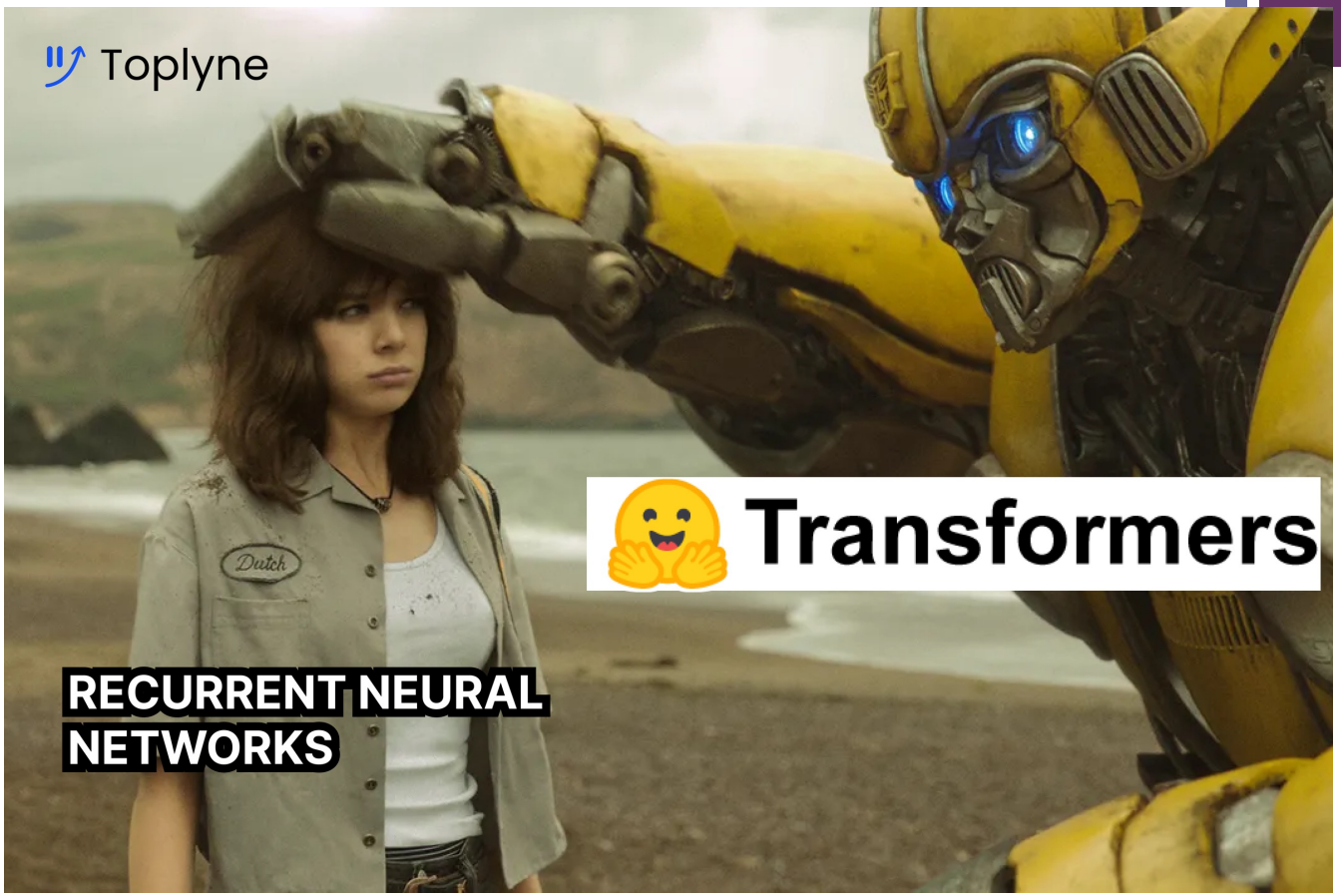
Available at: https://github.com/oxygen-ii/InceptMan/

+

Modern Trends: Beyond CNNs

Rise of Transformer (2017)

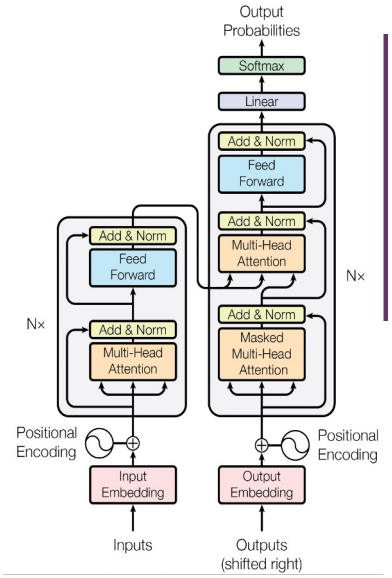




Transformer-Based Models

All Transformer based

1. Encoder-based model: BERT (2018)
2. Decoder-based model: GPT (2018)
3. Encoder and Decoder: BART (2019)



BERT [Devlin, et al, 2018]:
Bidirectional Encoder Representation from Transformers

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**
Google AI Language
{jacobdevlin,mingweichang,kentonl,kristout}@google.com

Abstract

We introduce a new language representation model called **BERT**, which stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. Unlike recent language representation models (Peters et al., 2018a; Rad-

There are two existing strategies for applying pre-trained language representations to downstream tasks: *feature-based* and *fine-tuning*. The feature-based approach, such as ELMo (Peters et al., 2018a), uses task-specific architectures that include the pre-trained representations as addi-

2019

OpenAI GPT (Generative Pre-Training) [Radford, 2018]

Improving Language Understanding by Generative Pre-Training

Alec Radford OpenAI alec@openai.com
Karthik Narasimhan OpenAI karthikn@openai.com
Tim Salimans OpenAI tim@openai.com
Ilya Sutskever OpenAI ilyasu@openai.com

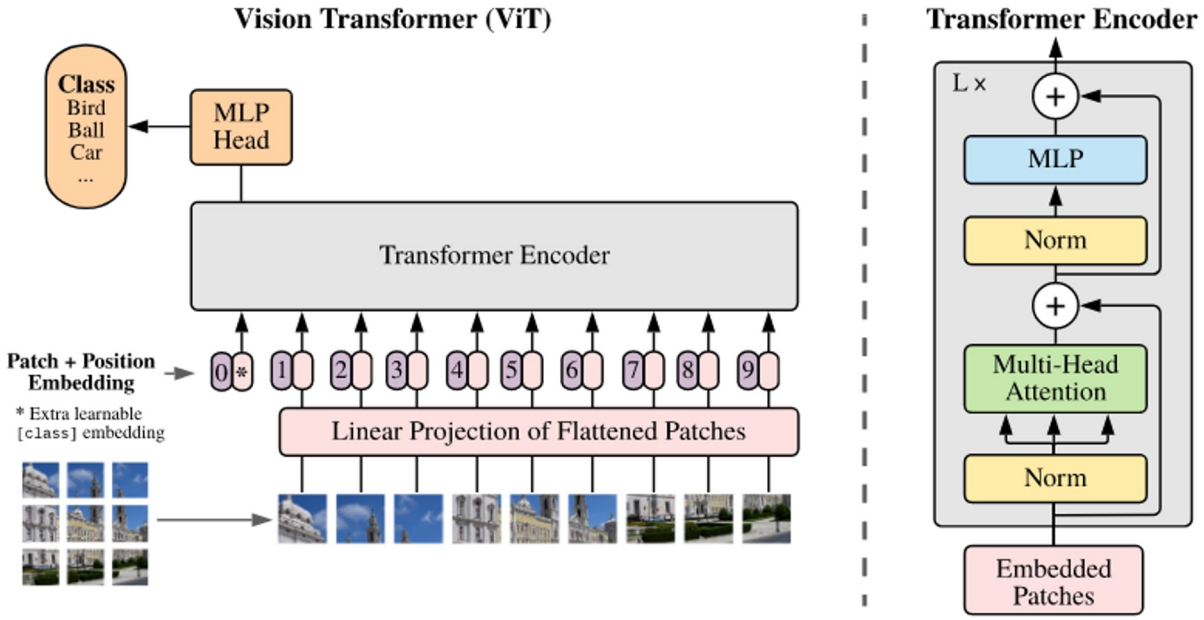
Abstract

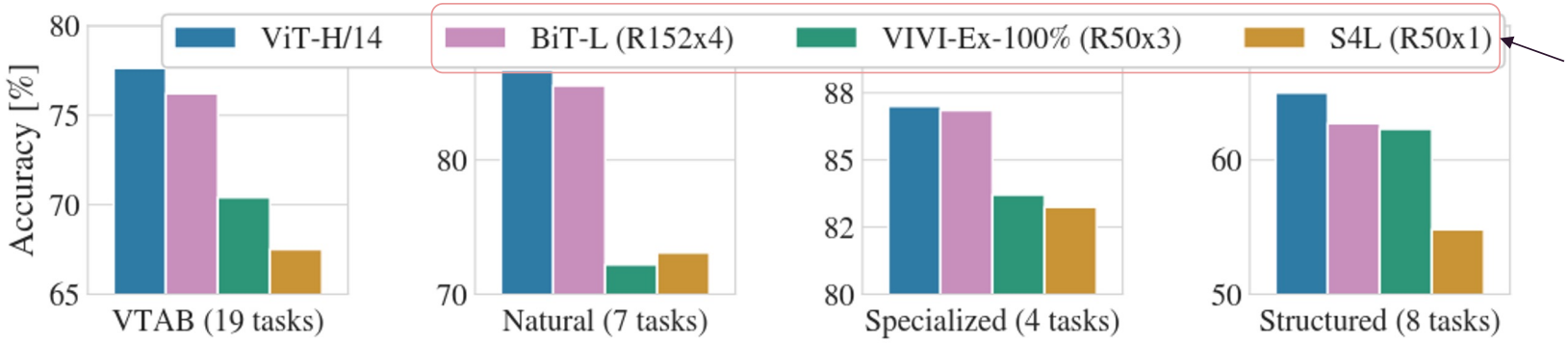
Natural language understanding comprises a wide range of diverse tasks such as textual entailment, question answering, semantic similarity assessment, and document classification. Although large unlabeled text corpora are abundant, labeled data for learning these specific tasks is scarce, making it challenging for discriminatively trained models to perform adequately. We demonstrate that large gains on these tasks can be realized by *generative pre-training* of a language model

Beyond NLP

Computer Vision

ViT uses only encoder.





Baselines are ResNet

AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

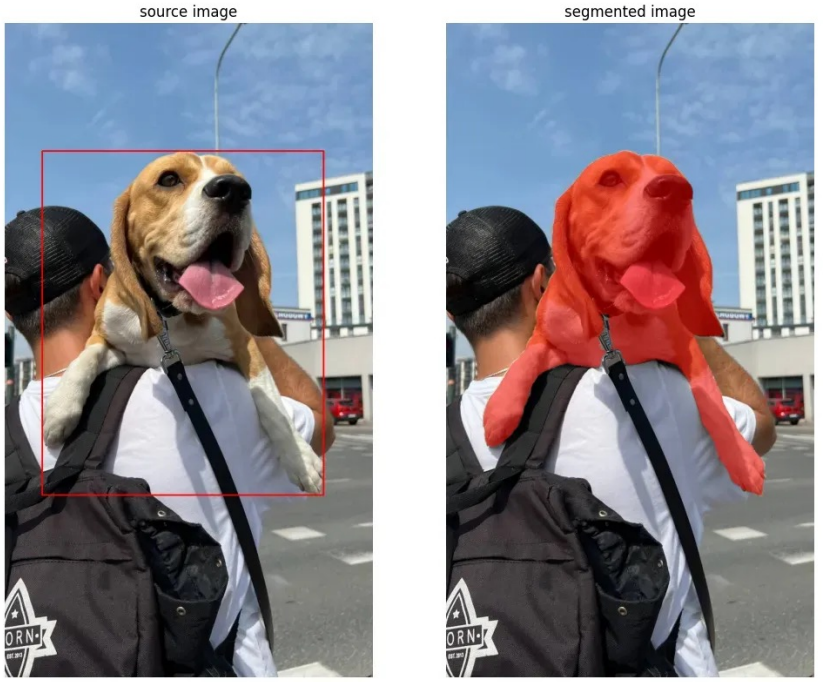
Segment Anything Model (SAM)

- SAM is a foundation model for promptable segmentation, developed by Meta, that can segment any object in an image without task-specific training.

- Supported Prompts
  - 📍 Point (foreground / background)
  - ⬛ Bounding box
  - 🖌️ Partial mask
  - ✍️ (Optional) Text (via extensions)

- SAM variations:
  - SAM-1 segments anything in a single image.
  - SAM-2 extends this to videos with memory.
  - "SAM-3" is not an official model
  - only a placeholder for ongoing research beyond SAM

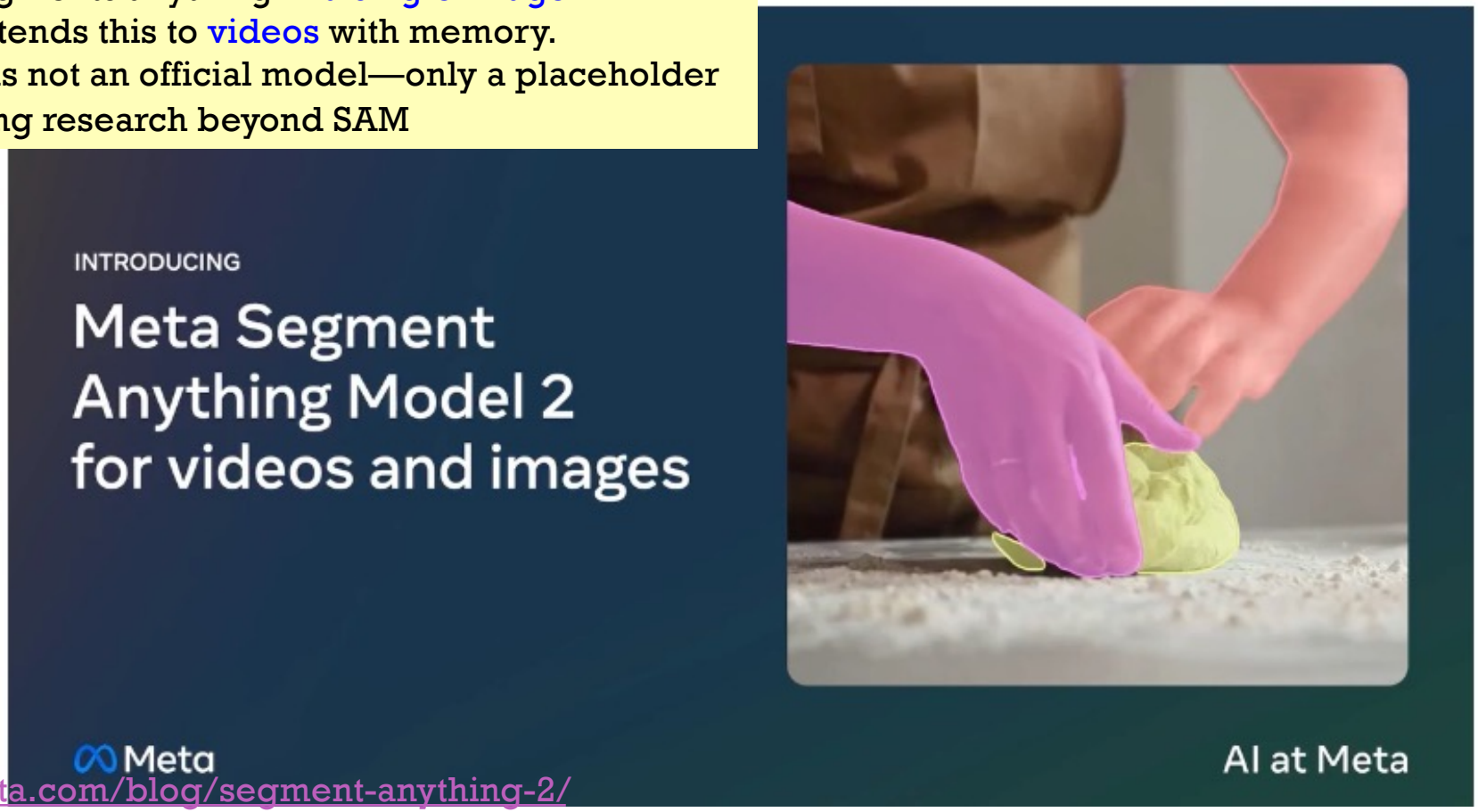
- Architecture (High Level)
  - Image Encoder (ViT): extracts visual features
  - Prompt Encoder: encodes user intent
  - Mask Decoder: generates object masks



Introducing SAM 2: The next generation of Meta Segment Anything Model for videos and images

July 29, 2024 · 15 minute read

SAM-1 segments anything in a single image.
SAM-2 extends this to videos with memory.
"SAM-3" is not an official model—only a placeholder for ongoing research beyond SAM



https://ai.meta.com/blog/segment-anything-2/

+ Thank you
& any questions