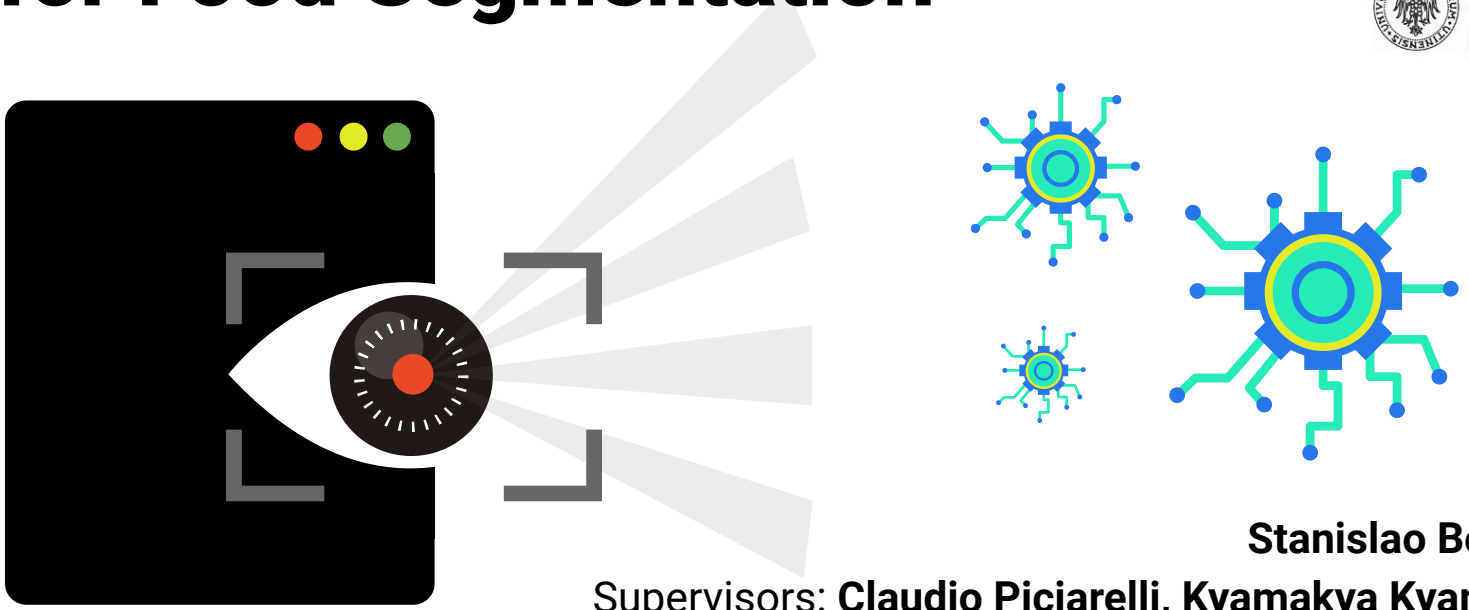
# Computer Vision Techniques for Food Segmentation



**Stanislao Bellantoni**

Supervisors: **Claudio Piciarelli, Kyamakya Kyandoghere**

*Università di Udine & Alpen-Adria-Universität Klagenfurt*

Academic years: 2023/2024

# Thesis Objectives

01

## Comparing

Compare various deep learning segmentation models.

02

## Dataset

Utilize UEC FoodPixComplete, with 10.000 images across 103 food classes.

03

## Performances

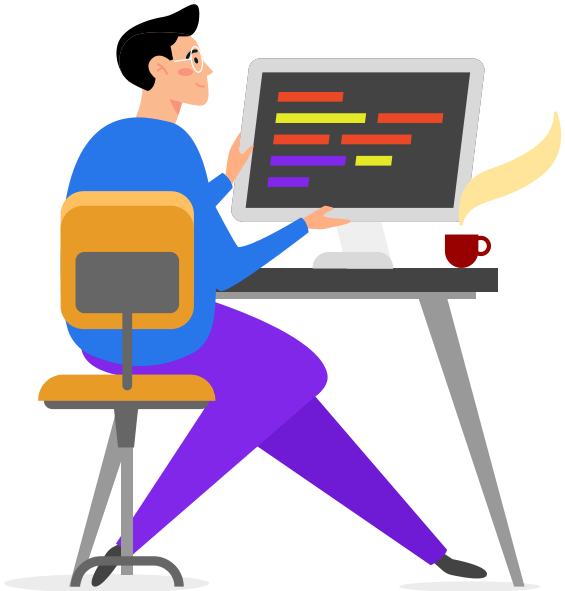
Assess model performance using a range of metrics.

04

## Analyze Results

Analyze and provide insights into the effectiveness of deep neural networks.

# Motivation



01

## Proliferation of Food Images

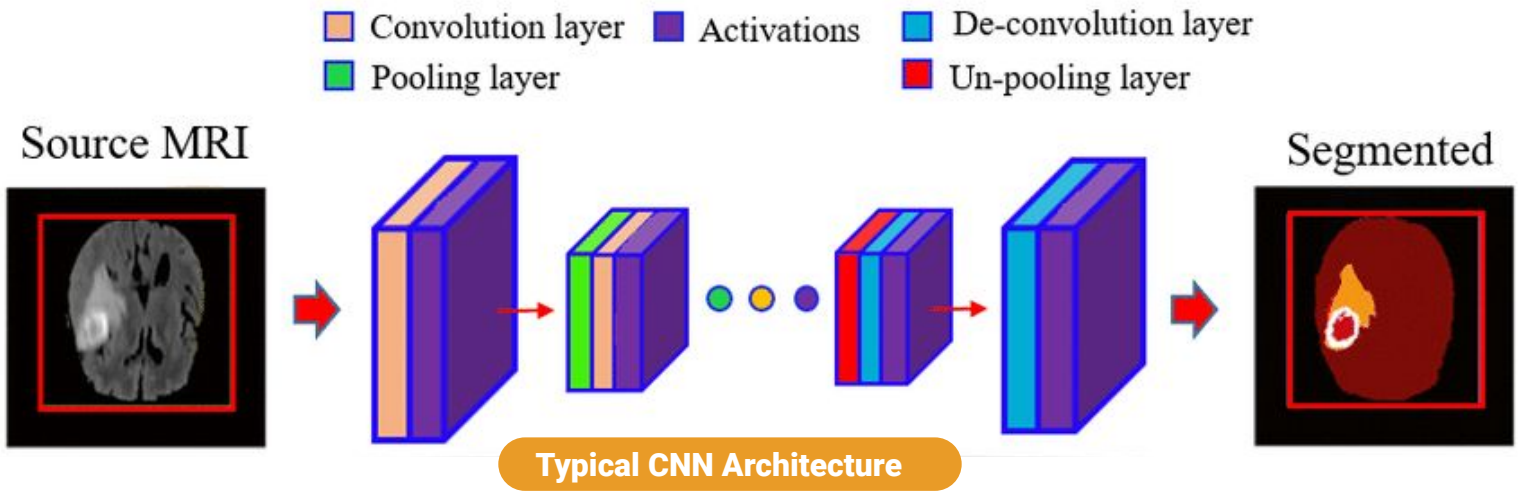
In today's digital age, social media platforms and recipe websites are **flooded** with appetizing **food photos**. This trend has sparked a demand for sophisticated systems **capable of analyzing culinary content**.

02

## Leveraging Technological Advances

Recent progress in deep learning has improved image recognition and segmentation. This research aims to tackle the **unique challenges** of food image segmentation, such as **texture variation** and **complex plating**, enhancing the **accuracy and reliability** of these models for practical applications.

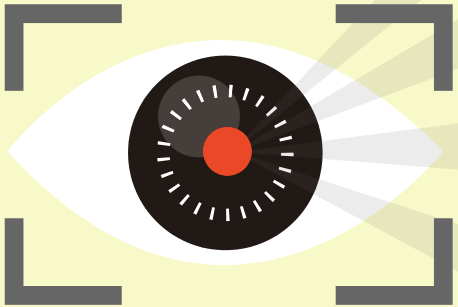
# Semantic Images Segmentation




The goal is to **assign a class label to each pixel in the image**, effectively partitioning the image into meaningful **regions or objects.** This approach is particularly useful in applications such as scene understanding, autonomous driving, and medical image analysis, where it is **crucial to accurately identify and delineate** different objects or structures within the image.

# Methodology

## Methodological Steps



**Literature Review**

Comprehensive analysis of existing segmentation techniques and datasets.

01

**Implementation**

Utilized Python 3.8 and PyTorch to implement six segmentation algorithms.

02

**Training & Validation**

Models were trained and validated on the UECFoodPix dataset.

03

**Evaluation**

Evaluated and compared the models' performances using standard segmentation metrics.

04

# Literature Review

**2016** 01

## UNIMIB2016

*Ciocca et al.*

Introduced a **dataset** for **canteen tray** image segmentation and classification. Evaluated classification strategies based on **global and local features.**

**2019** 02

## Enhanced Food50Chen

*Alan et al.*

Created a comprehensive **dataset** with **augmented images** to evaluate food localization and semantic segmentation algorithms.

**2021** 03

## FoodSeg103

*Wu et al.*

Developed a **large-scale dataset** with fine-grained ingredient labels, and **proposed a multi-modality** pre-training approach **incorporating recipe information.**

**2023** 04

## Automated Nutritional Estimation

*Sultana et al.*

**Reviewed methods** for food image classification, **volume/weight estimation** and **nutrition** estimation, emphasizing the transition to deep learning.

# Dataset UEC-FoodPixComplete

**70%**

## Training Set

Used to train the model.

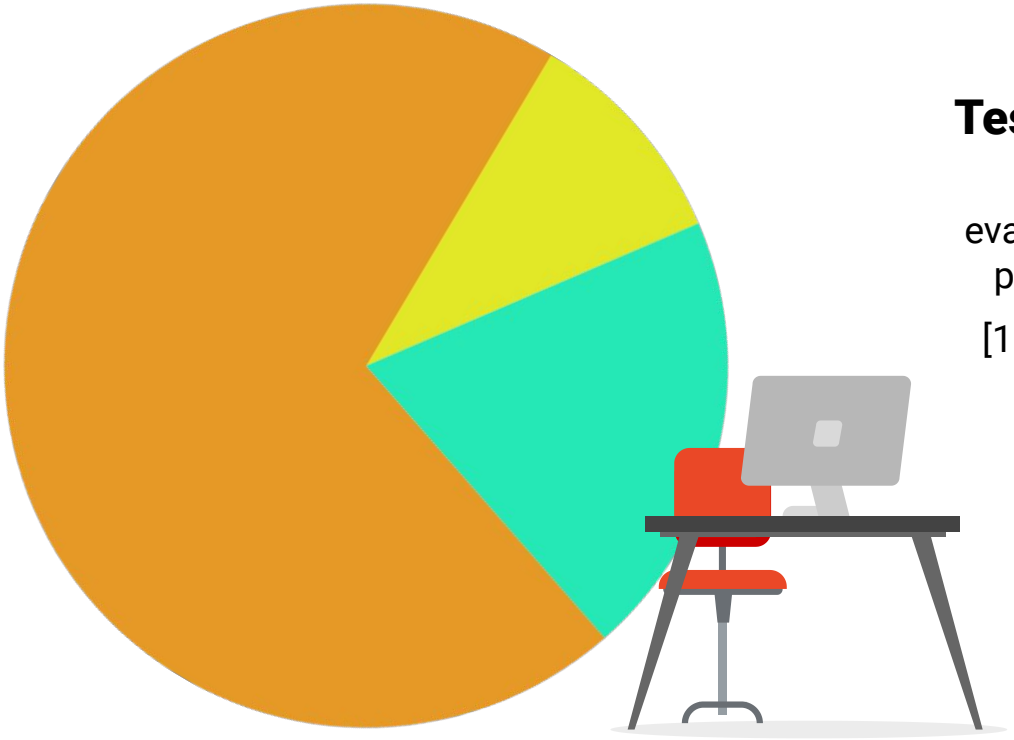
[7200 images]

**20%**

## Validation Set

Used to tune model parameters and for early stopping.

[1800 images]



**10%**

## Testing Set

Used to evaluate model performance.

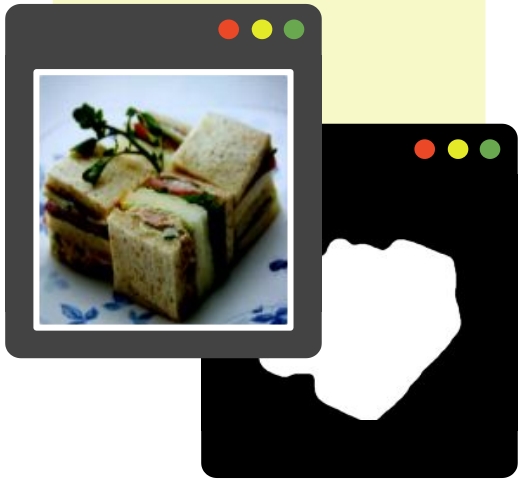
[1000 images]

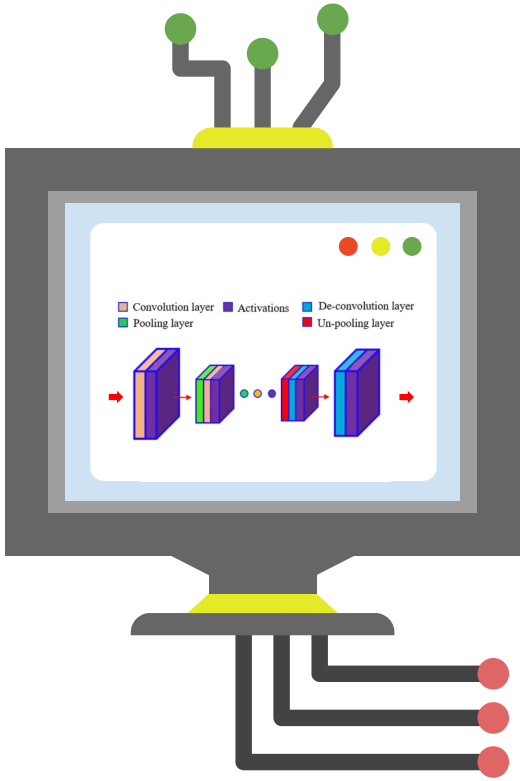
# Model Input & Output

## Input

Dataset food images and manually refined segmented GT Masks.

**Image +**
**Ground Truth Mask**

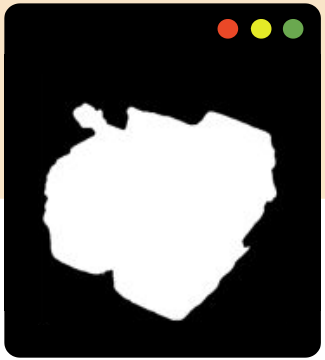





## Output

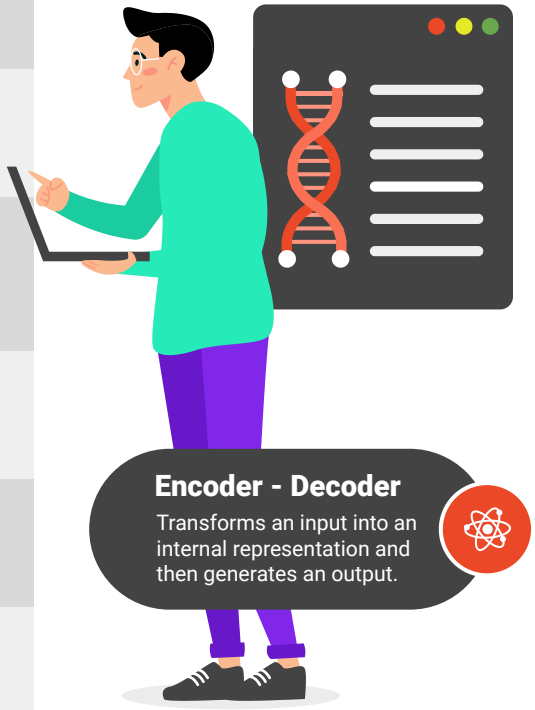
Pixel-wise classification of the images, distinguishing food items from the background.

**Predicted Mask**



# Segmentation Models

| | |
|---|---|
| **UNet** | Encoder-decoder for precise localization |
| **UNet++** | Enhanced U-Net with dense skip connections |
| **Feature Pyramid Network** | Top-down, multi-scale feature extraction |
| **PSPNet** | Pyramid pooling for global and local context |
| **DeepLabV3** | Atrous convolutions for multi-scale context |
| **DeepLabV3+** | Atrous convolutions with encoder-decoder |



# Evaluation Metrics

Single Metrics

## Accuracy

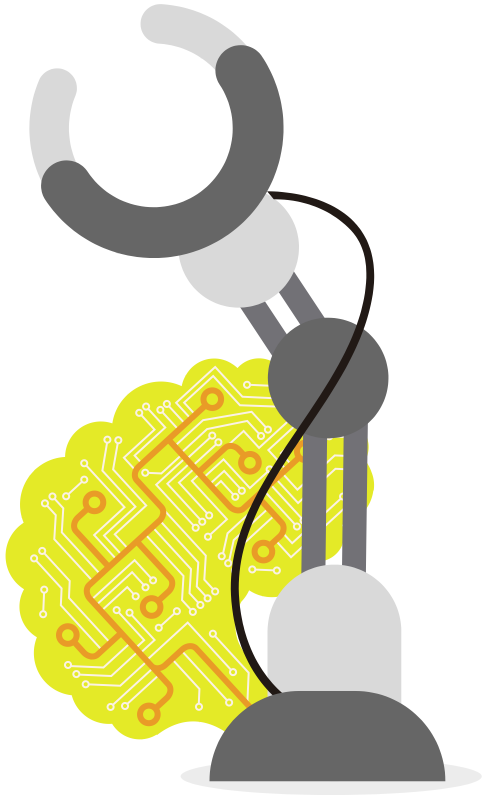
Correctness of segmentation predictions.

## Recall

True positives out of actual positives.

## Precision

True positives out of predicted positives.



Combined Metrics

## Jaccard Score

Overlap between predicted and true masks.

## F1 Score

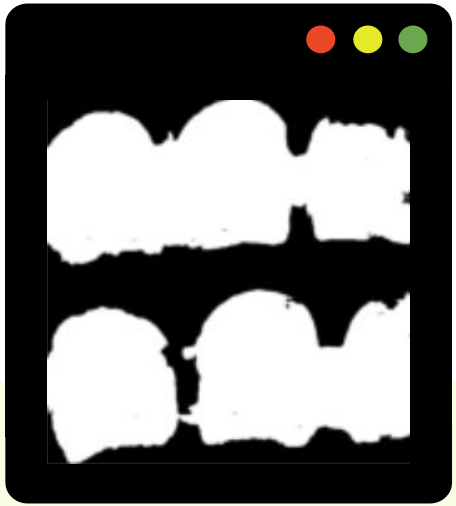
Harmonic mean of precision and recall.

# Results Comparison

| * | Model | Accuracy | Precision | Recall | F1 Score | Jaccard Score |
|---|---|---|---|---|---|---|
| 05 | UNet | 0.904 | 0.809 | 0.963 | 0.880 | 0.785 |
| 06 | UNet++ | 0.895 | 0.800 | 0.951 | 0.869 | 0.769 |
| 02 | FPN | 0.928 | 0.860 | 0.959 | 0.907 | 0.829 |
| 03 | PSPNet | 0.923 | 0.853 | 0.952 | 0.900 | 0.818 |
| 04 | DeepLabV3 | 0.917 | 0.831 | 0.97 | 0.895 | 0.810 |
| 01 | DeepLabV3+ | 0.929 | 0.853 | 0.974 | 0.909 | 0.834 |

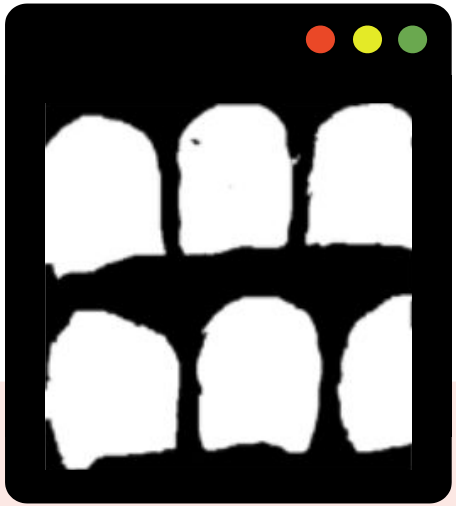
* ranking of the best and worst models according to evaluation metrics.

# Worst Model: UNet



Jaccard Score

0.785



**Image**

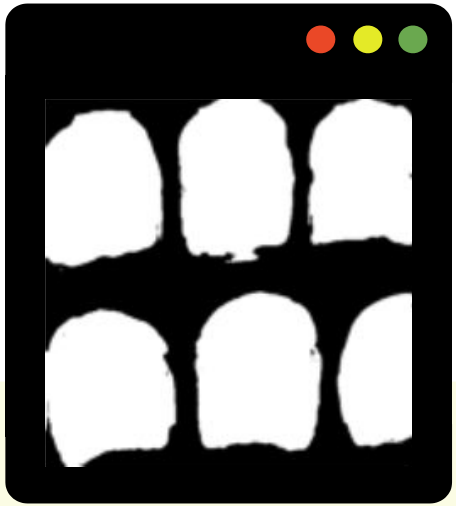
**Predicted Mask**

Lower precision, struggles with complex textures, inefficiency in handling scale variations.

**Ground Truth Mask**

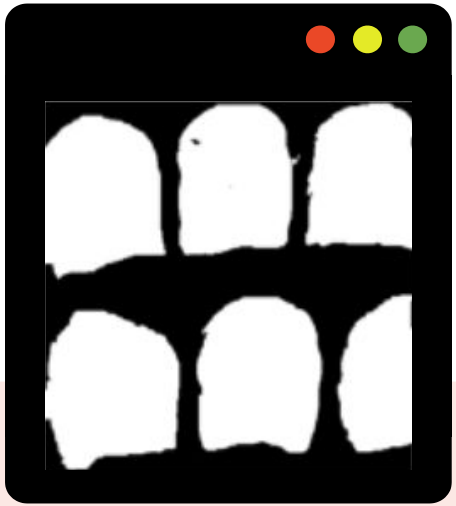
Reference segmentation used for evaluating.

# Best Model: DeepLabV3+



Jaccard Score

0.834



**Image**

**Predicted Mask**

High precision and recall, excellent in capturing local and global contexts, robust against scale and texture variations.

**Ground Truth Mask**

Reference segmentation used for evaluating.

# Summary of Findings

## Insights

Provided valuable insights into model performance and suggest practical applications in health and nutrition.

## Effectiveness

Demonstrated the effectiveness of deep learning models for the semantic segmentation of food images.

## Datasets

The use of large and diverse dataset like UEC FoodPixComplete improve model robustness and generalizability.

# Practical Applications

01

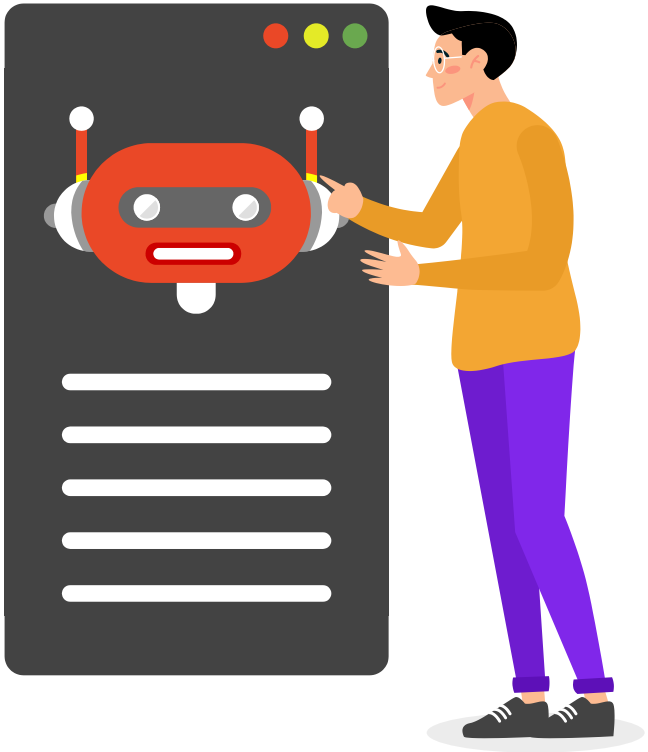
## Dietary Planning

Facilitates precise calculation of nutritional intake and meal planning.

02

## Food Logging

Supports automated food logging for restaurant and tracking apps.



## Food Quality Assessment

03

Enhances the ability to assess and monitor food quality and presentation.

# Future Work

01

## Exploration of New Models

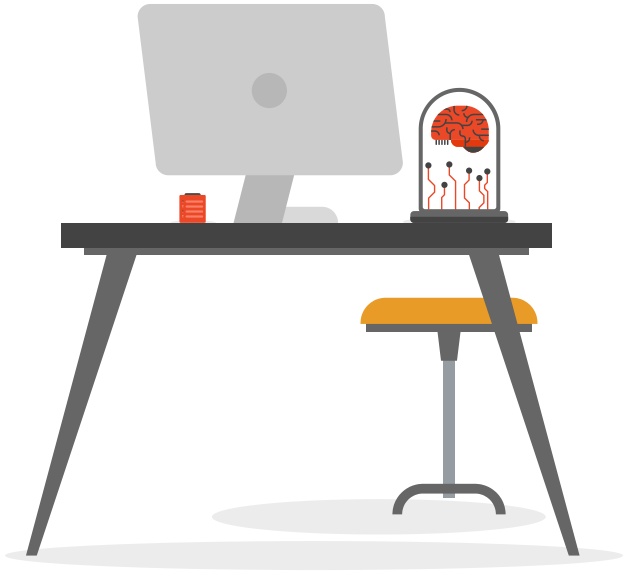
Continue exploring new and improved segmentation models.

02

## Larger Datasets

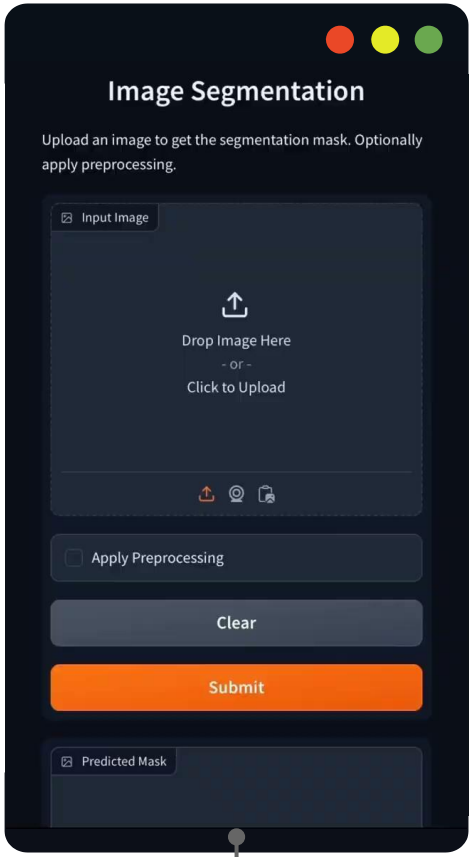
Utilize larger and more diverse datasets to improve generalizability and robustness.

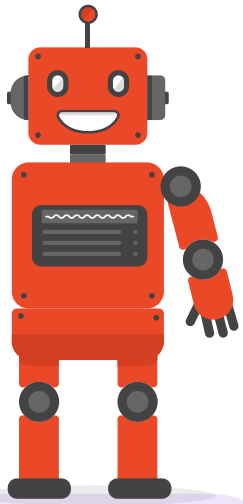
# THANK YOU
## For the Attention!



## On Device Test

Using the Gradio interface and weights of our best trained model, DeepLabV3+



## Insights

With ResNet34 preprocessing, the model recognizes more food and is more accurate.