
UEC-FoodPIX Complete: A Large-scale Food Image Segmentation Dataset

Kaimu Okamoto and Keiji Yanai

The University of Electro-Communications, Tokyo



Introduction

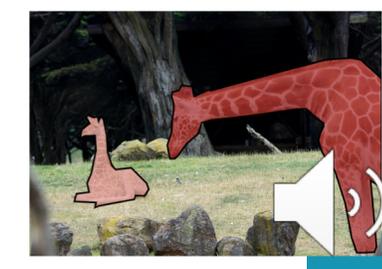
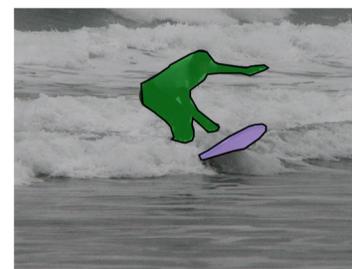
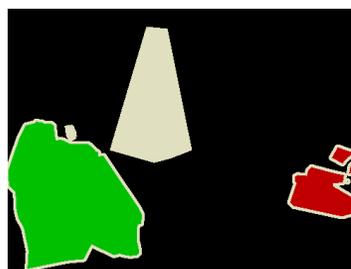
Widely used as large-scale segmentation datasets [3][4]



22categories
About 10k
images



80categories
About 330k
images



Introduction

Widely used as large-scale segmentation datasets [3][4]

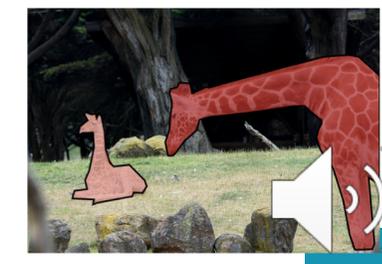
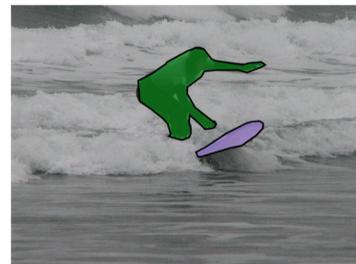


22categories
About 10k
images



80categories
About 330k
images

These includes **only a limited number of food categories.**
(Banana, Pizza, etc.)



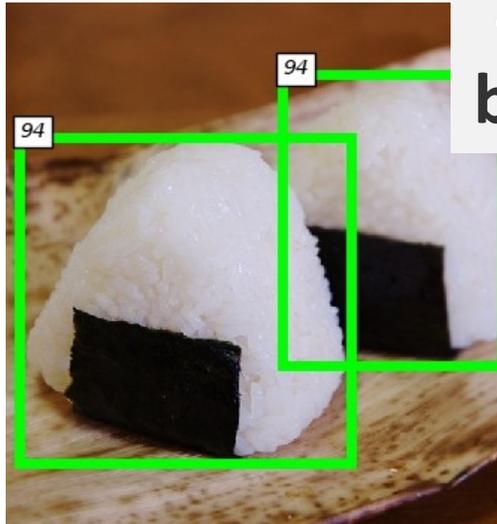
Introduction

Food segmentation dataset

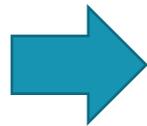
- UEC-Food Pix by Ege et al.[2]

⇒Semi-automatically annotated by GrabCut [16] based on the bounding boxes annotated in the UECFood-100 dataset

BBox



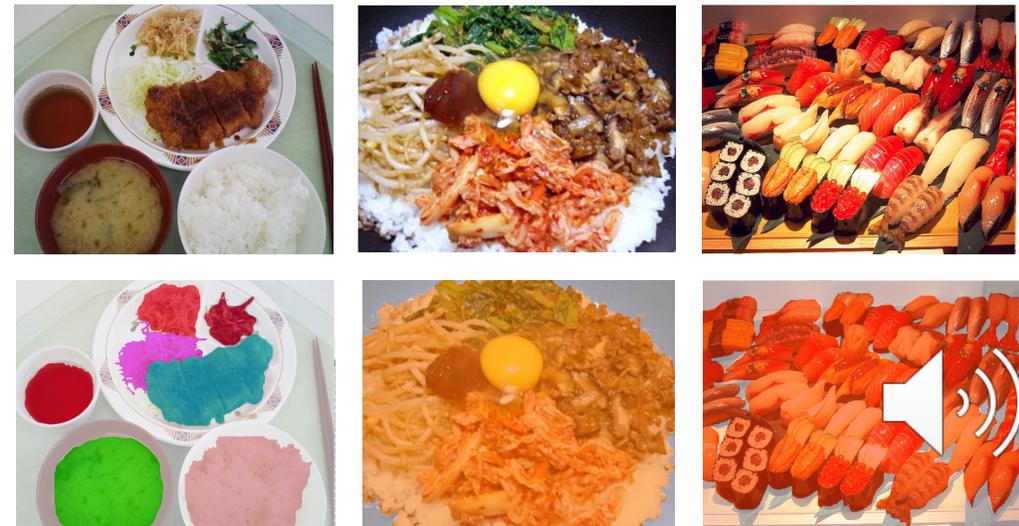
annotated
by Grab Cut



Mask



UEC-Food Pix dataset



Introduction

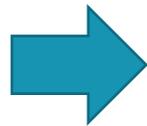
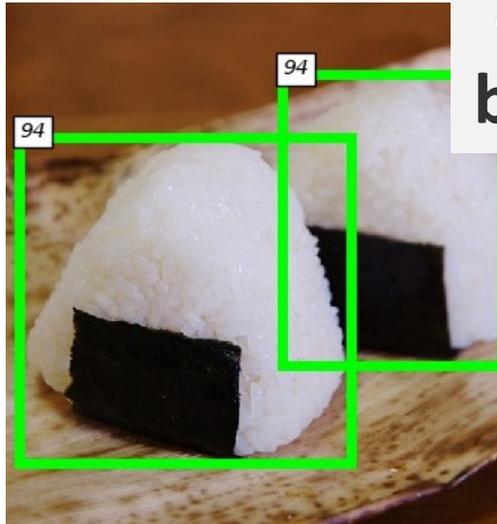
Food segmentation dataset

- UEC-Food Pix by Ege et al.[2]

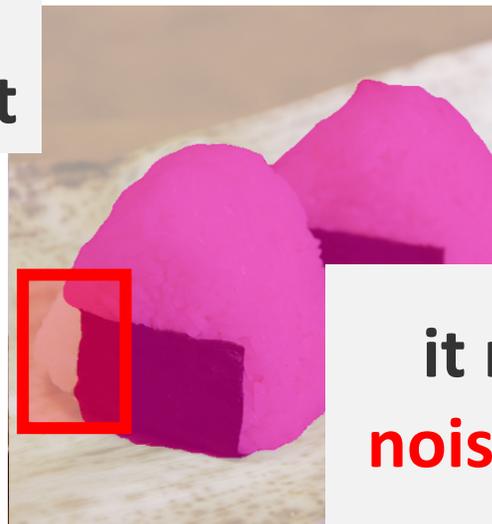
⇒Semi-automatically annotated by GrabCut [16] based on the bounding boxes annotated in the UECFood-100 dataset

BBox

annotated
by Grab Cut

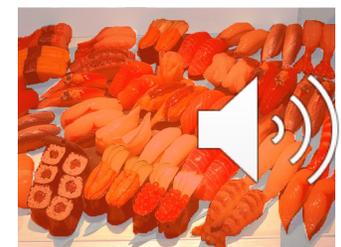


Mask



it may contain
noisy annotations

UEC-Food Pix dataset

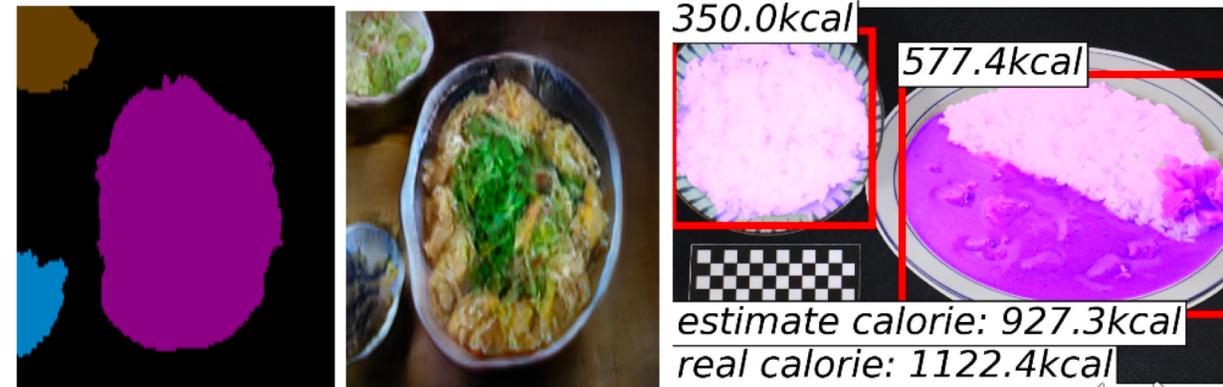
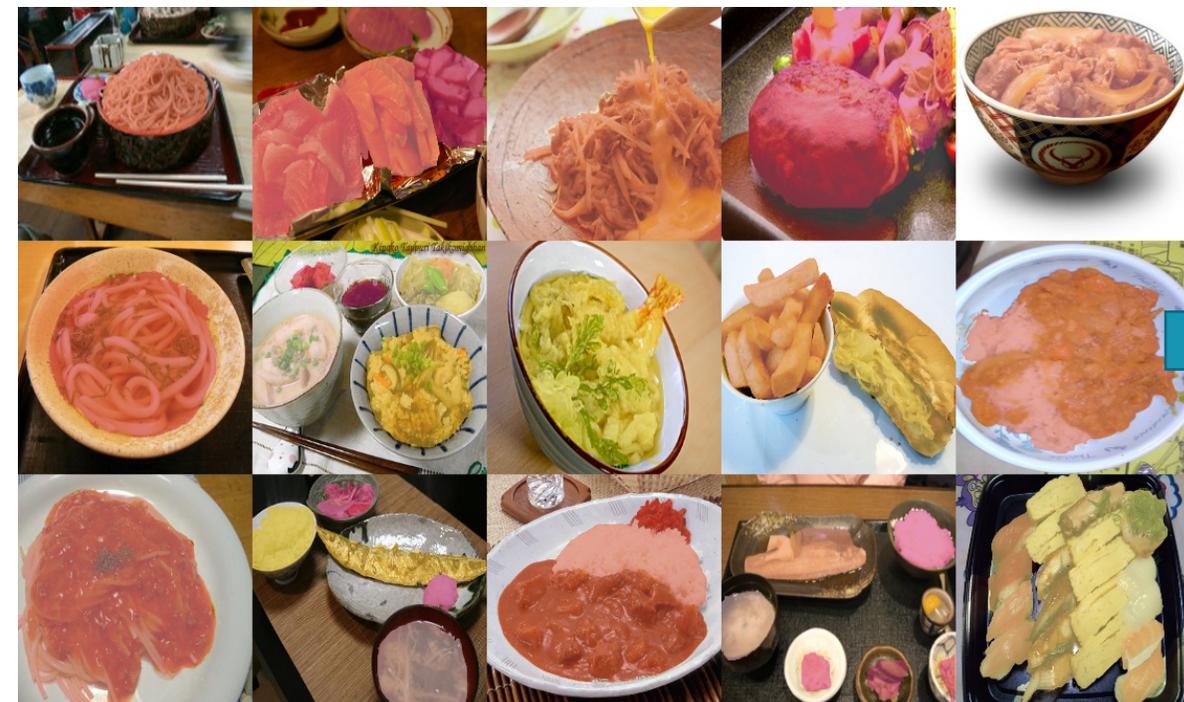


Objective

- Updating UEC-Food Pix to make new food segmentation datasets
- Introducing application examples of this dataset

UEC-Food Pix Complete

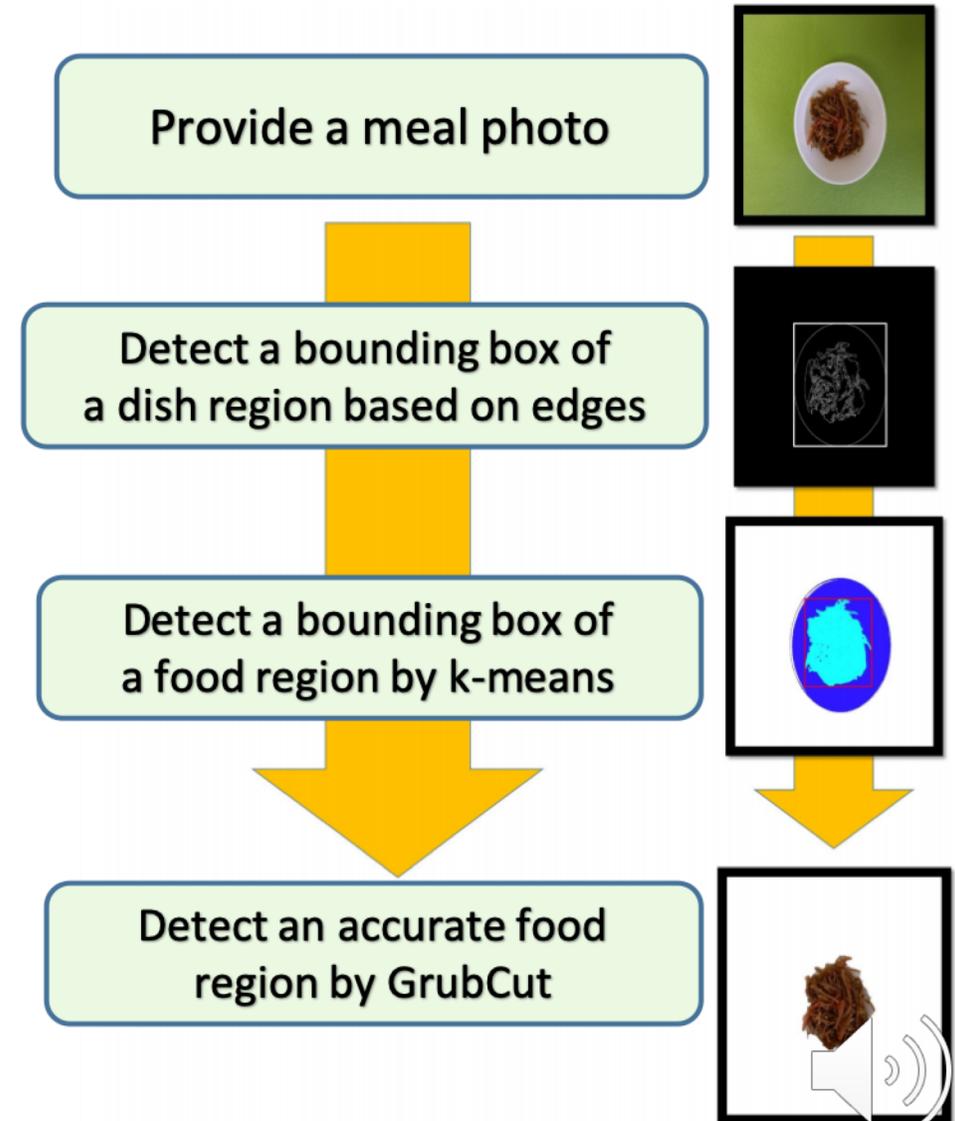
- Food images synthesis
- Calorie estimation



Related Work

A region based food calorie estimation

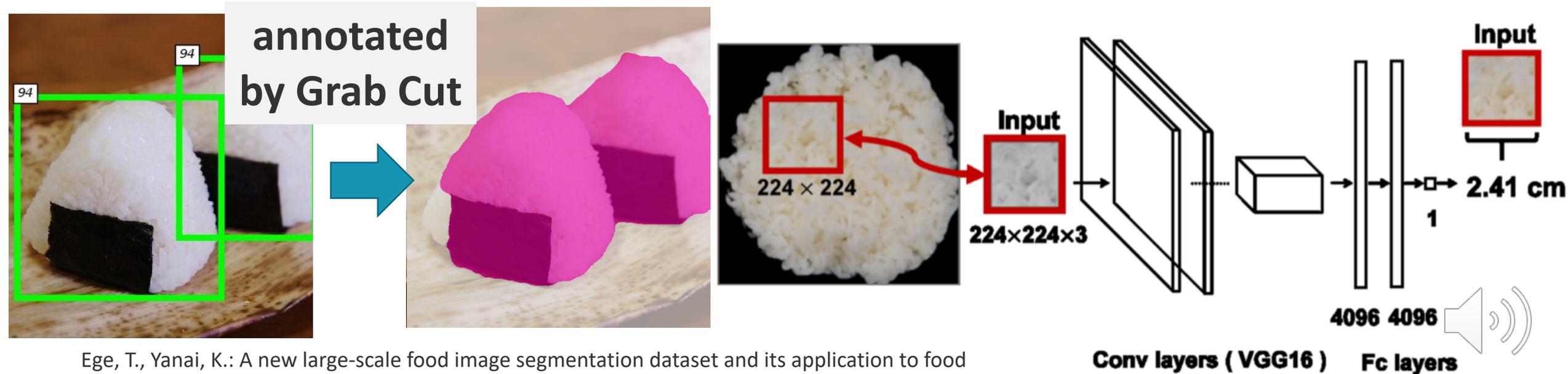
- Calorie Cam [13]
 - ⇒ running on a mobile phone
 - ⇒ Using GrabCut[16] to detect food region



Related Work

Estimating actual size of foods without a reference card

- Rice image actual size estimation system[6]
 - ⇒ Estimating actual size of foods by using the size of rice grains
 - ⇒ Creating the dataset for food segmentation applying calorie estimating



Ege, T., Yanai, K.: A new large-scale food image segmentation dataset and its application to food calorie estimation based on grains of rice. MADiMa, (2019)

Overview of the datasets

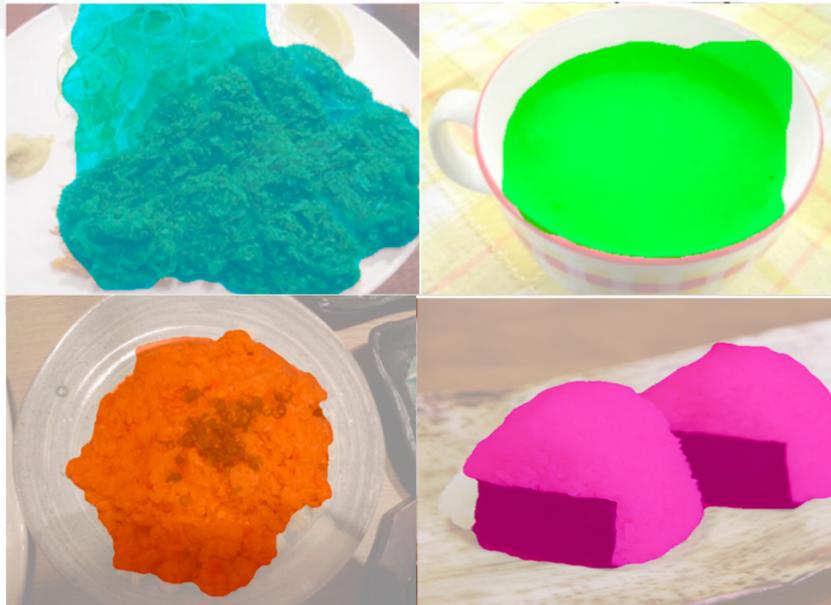
Introducing the datasets

UEC-Food Pix

102 food categories

9,000 train images(Automatically)

1,000 test images(hand annotation)

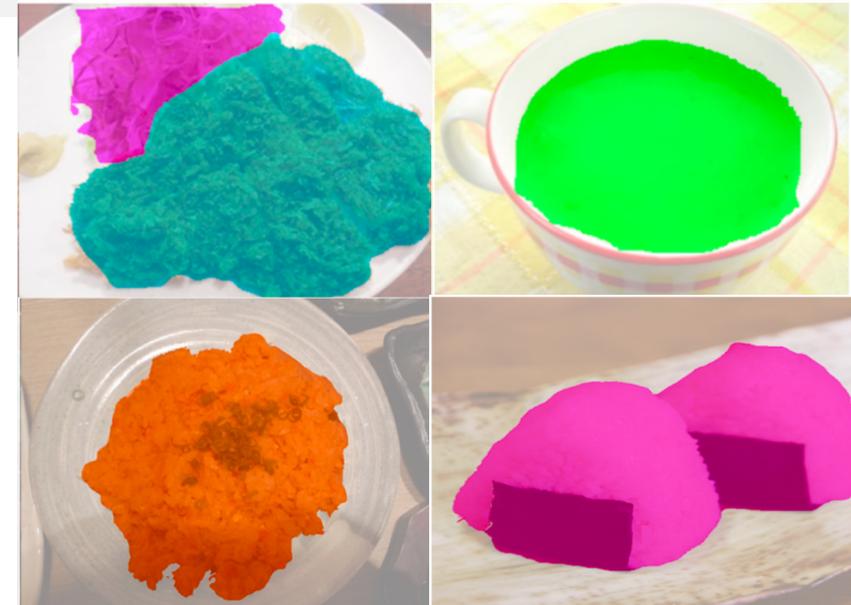


UEC-Food Pix Complete

102 food categories

9,000 train images(**hand annotation**)

1,000 test images(hand annotation)

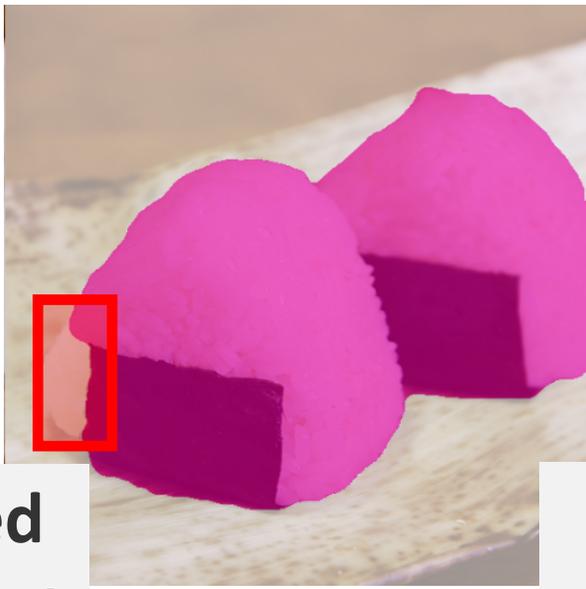
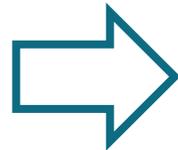
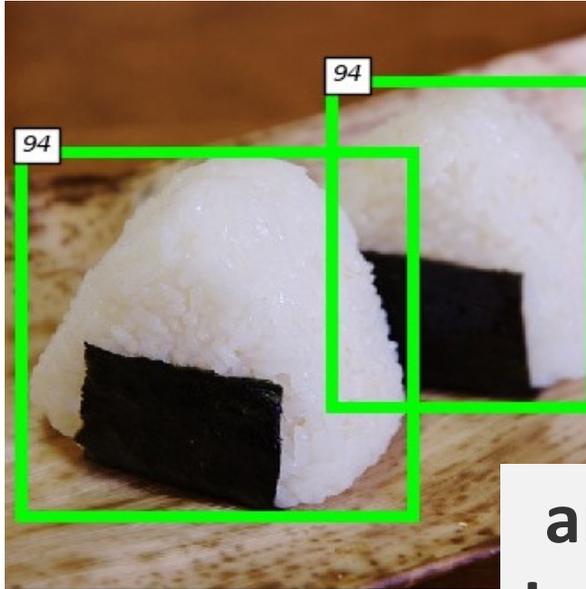


Dataset Construction

We created **UEC-FoodPix Complete** as a higher quality food image segmentation dataset by updating UEC-FoodPix manually.

UEC-FoodPIX

UECFoodPIX Complete



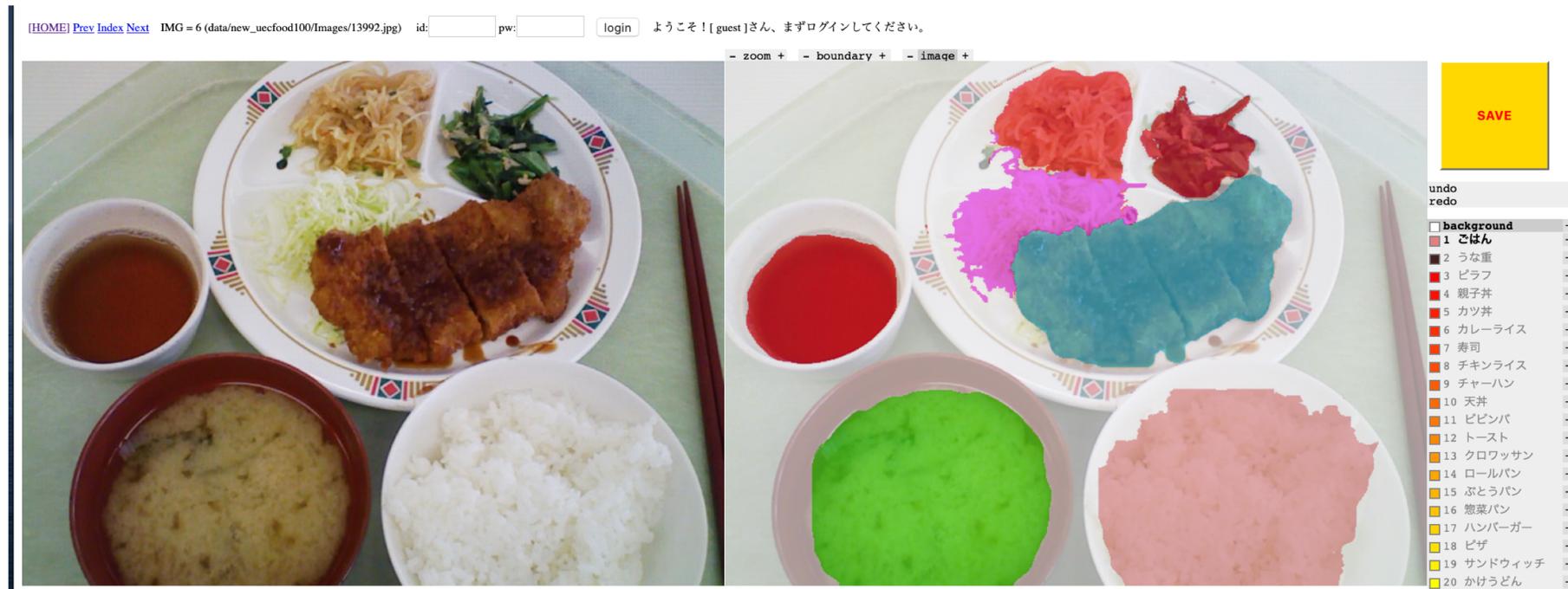
annotated
by Grab Cut

Manually
modifying



Using Web-based pixel-wise annotation tool for create this dataset

- Proposed web tool by Pongsate et al. [18].
⇒ allows easy synthesis and separation of food regions with super-pixels



Tangseng, P., Wu, Z., Yamaguchi, K.: Looking at outfit to parse clothing. arXiv:1703.01386 (2017)



Dataset Construction

When done manually, the mask will differ depending on the person who works

⇒ set annotation some rules on how to create food region masks

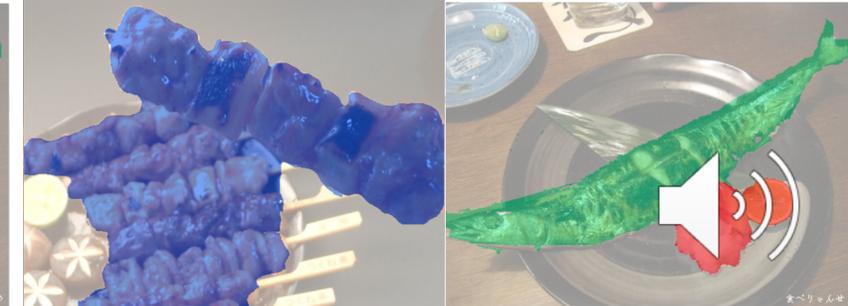
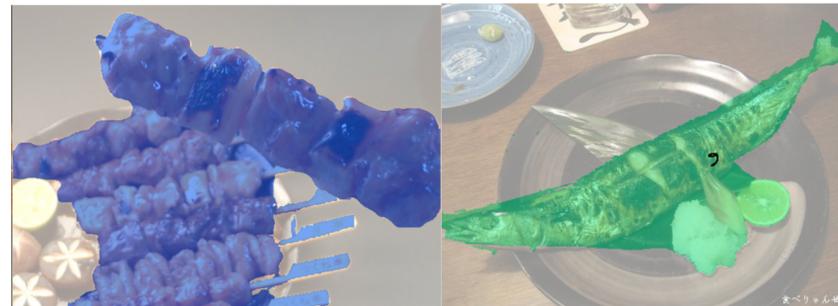
ex.) ‘Yakitori’ skewers do not include, ‘Grated radish’ is set in the other food category

⇒ 10,000 images, 4 months creation period

Original a image

Different masks

Apply a rule to the mask



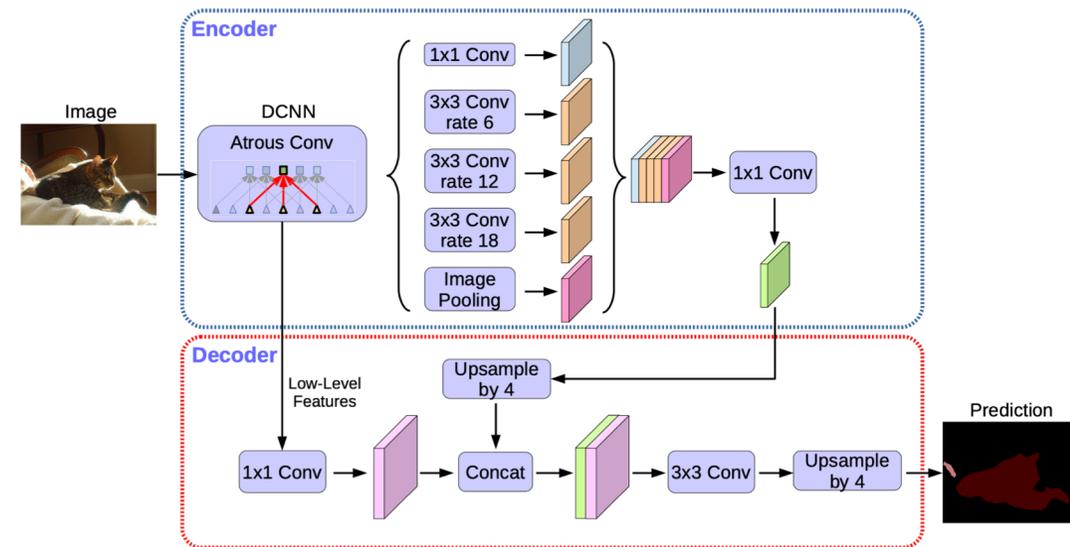
Accuracy and mIoU comparison between UEC-FoodPIX and Complete

⇒ food region estimation by Deeplab V3+[1]

Training dataset	Acc	mIoU
UEC-Food Pix	0.560	0.416
partial UEC-Food Pix Complete (2000 hand annotation)	0.597	0.436
UEC-FoodPix Complete	0.668	0.555

9,000 training images 1,000 test images

Deeplab V3+



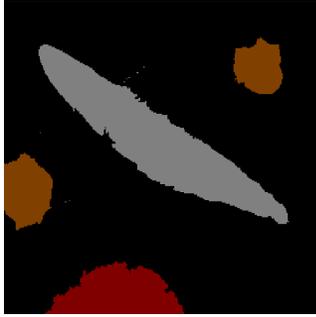
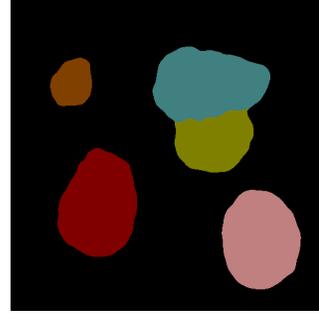
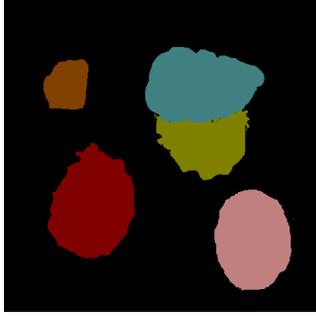
Some segmentation results by both models

Input

Ground truth

UECFODPIX

COMPLETE



The results by the **Complete model** are similar to the groundtruth



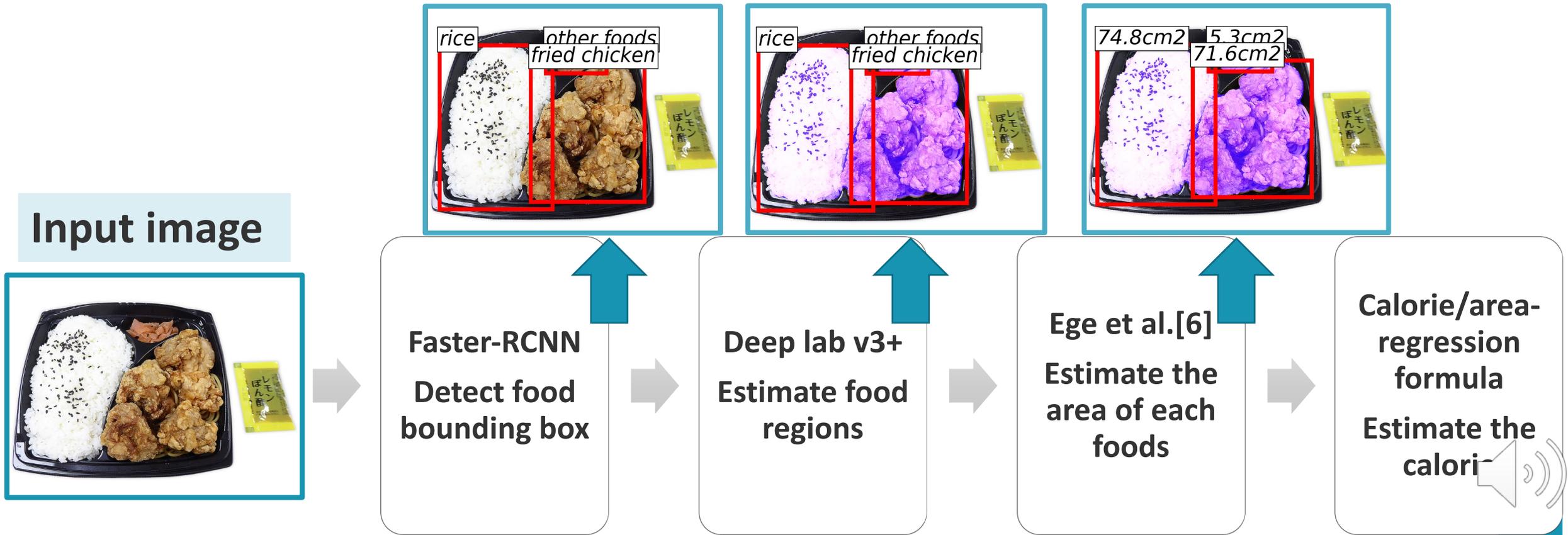
Application 1: Region-based Food Calorie Estimation



Application 1: Region-based Food Calorie Estimation

Food region and Calorie estimation system

⇒ Following Ege et al.[6]



Application 1: Region-based Food Calorie Estimation

Food region estimation

dataset	Area(rice)		Area(multiple food)		Calorie(multiple food)	
	Abs.err(cm^2)	Rel.err(%)	Abs.err(cm^2)	Rel.err(%)	Abs.err(kcal)	Rel.err(%)
UECFoodPix[6]	7.21	8.73	30.0	14.2	240.8	29.8
Complete	3.03	3.67	44.7	20.7	268.5	33.4

- Only rice area estimation accuracy was improved
- Multiple dishes, both errors of the original UEC-Food Pix were lower.



Application 1: Region-based Food Calorie Estimation

Estimation of Food region size

Ground truth

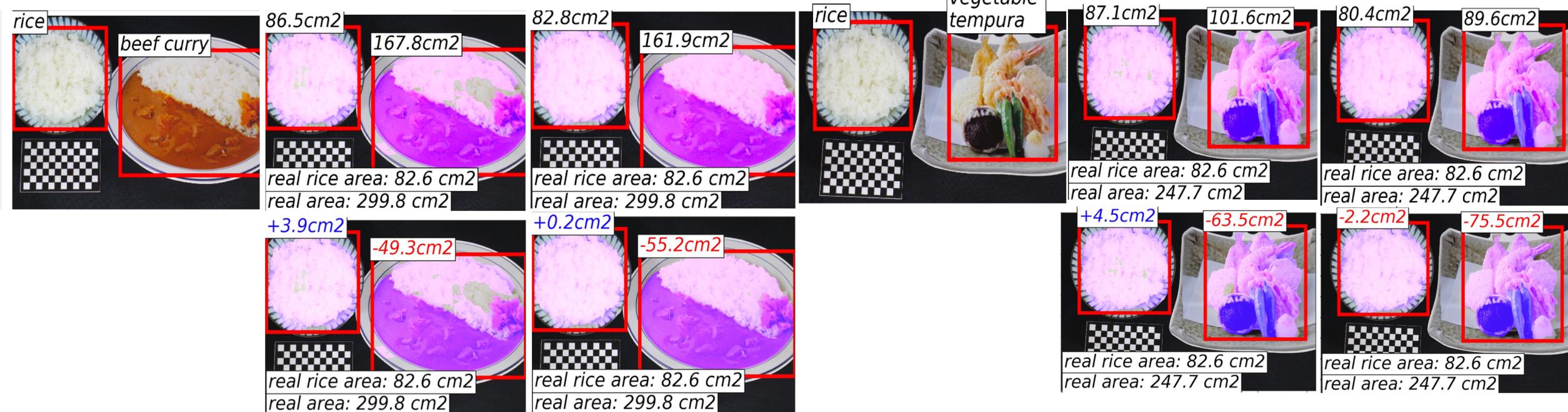
UECFoodPix

Complete

Ground truth

UECFoodPix

Complete



- Higher-quality area estimation is possible using the model trained with “Complete”.



Application 1: Region-based Food Calorie Estimation

Estimation of Food region size

Ground truth

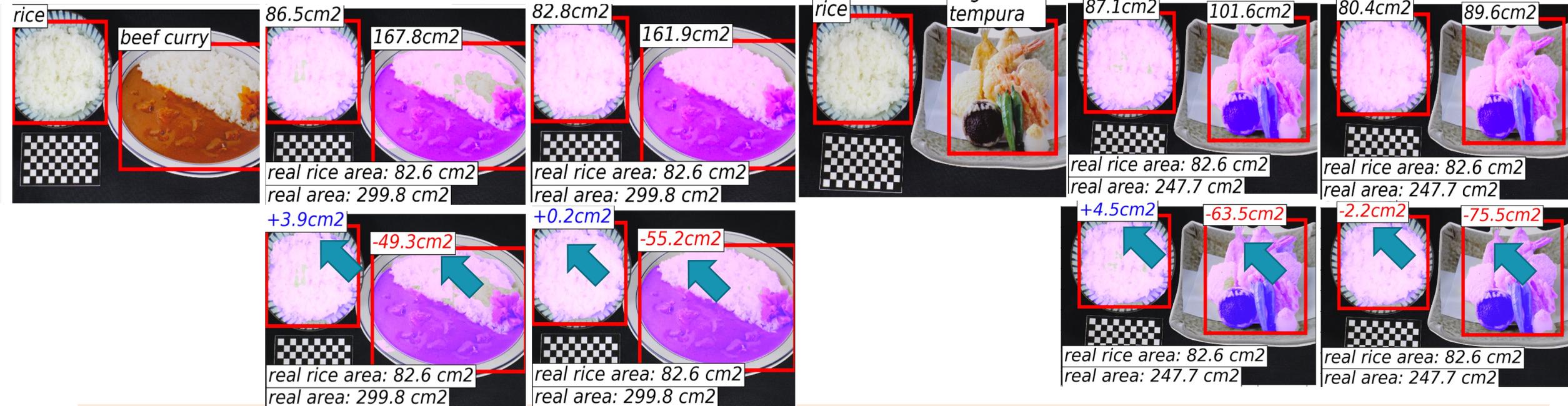
UECFoodPix

Complete

Ground truth

UECFoodPix

Complete



- When training by 'Pix', the rice part calculates the large actual size
- So even if the mask is applied a little, the value close to the total correct value is calculated.



Application 1: Region-based Food Calorie Estimation

Food Calorie estimation

Ground truth

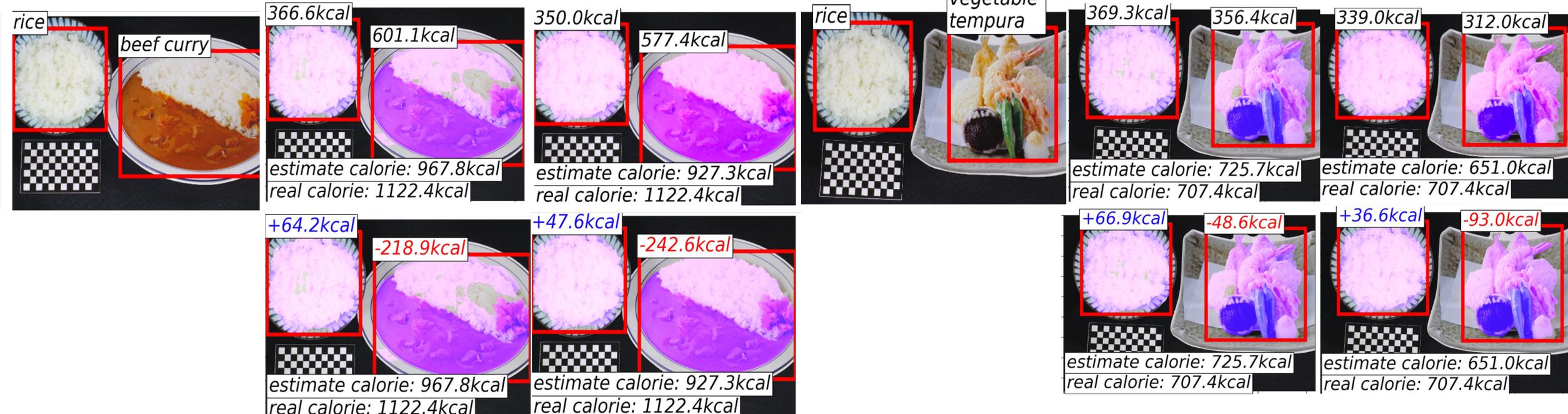
UECFoodPix

Complete

Ground truth

UECFoodPix

Complete



- The values close to the actual calories were calculated
- Since the calorie content is estimated based on the actual size, also this result is better with PIX.



Application 2: Mask-based Image Synthesis



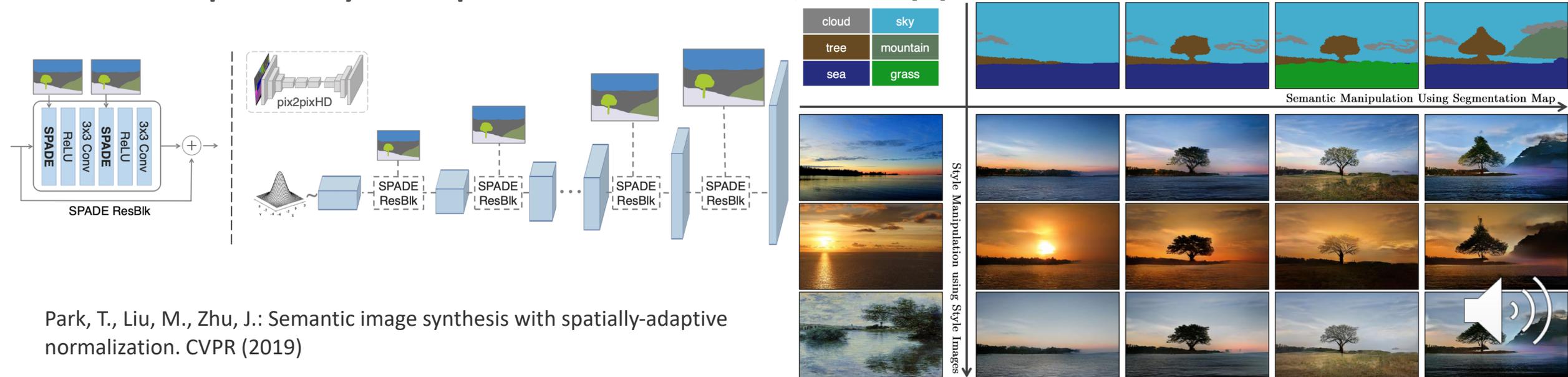
Application 2: Mask-based Image Synthesis

Mask-based Image synthesis method

- SPADE[14]

⇒the state-of-the-art mask-based image generation method

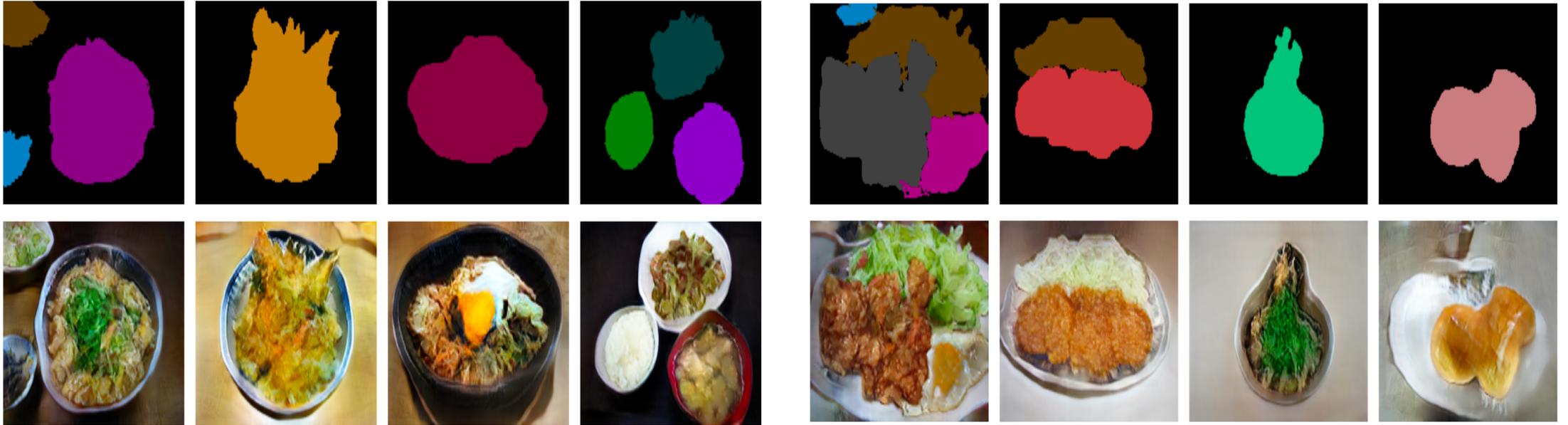
⇒spatially adaptive normalization applied



Park, T., Liu, M., Zhu, J.: Semantic image synthesis with spatially-adaptive normalization. CVPR (2019)

Application 2: Mask-based Image Synthesis

Food Images Synthesis by SPADE[14]

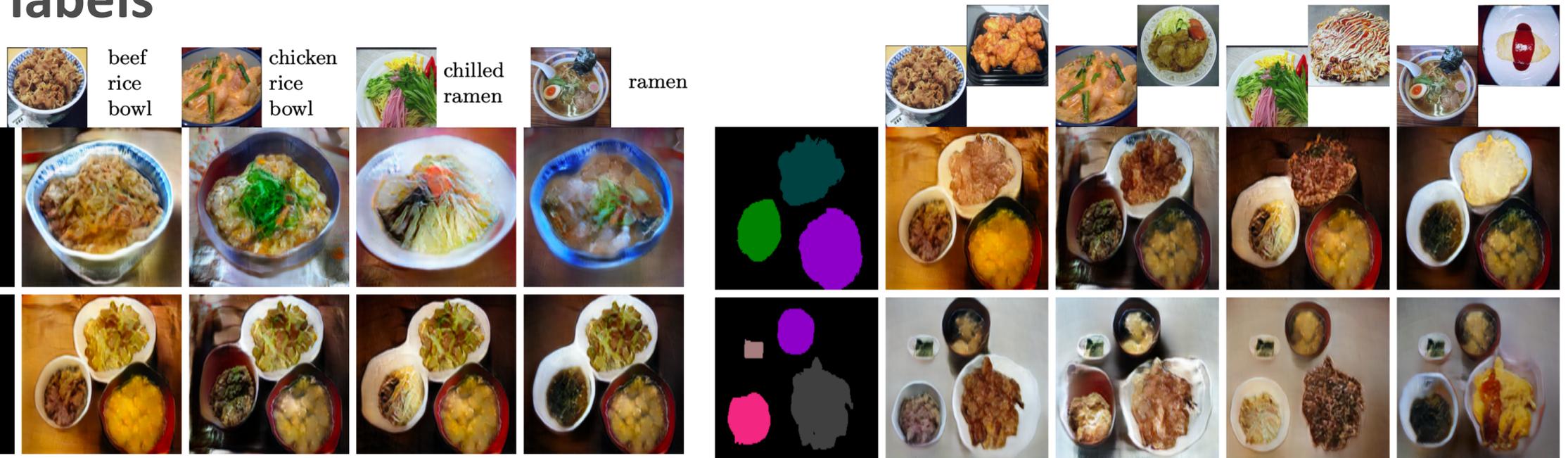


- Rice bowls and Japanese combo meals were successfully generated.
- Deformed and concatenated images are generated



Application 2: Mask-based Image Synthesis

Food Images Synthesis from the same mask images with different class labels



Multiple dish food images were synthesized well, although the shape of plates and bowls sometime were distorted and look unnatural.



Conclusions

Summary

- UEC-FoodPix Complete, by updating the existing the food image segmentation dataset UEC-FoodPix [6].
- This dataset was shown that it can be applied to calorie content estimation and image generation.

Future works

- Annotate this dataset with other information such as calorie.



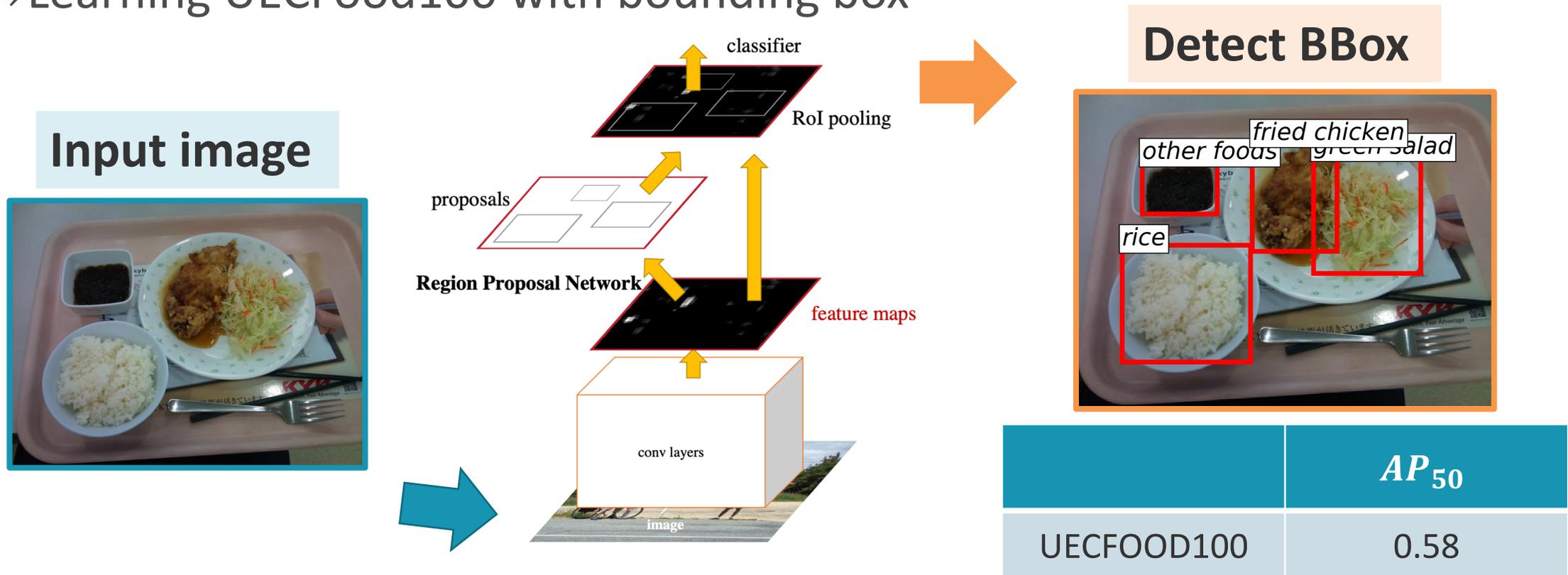
Question Slide



Application 1: Region-based Food Calorie Estimation

Detecting food bounding box using Faster R-CNN [6]

⇒ Learning UECFood100 with bounding box



Application 1: Region-based Food Calorie Estimation

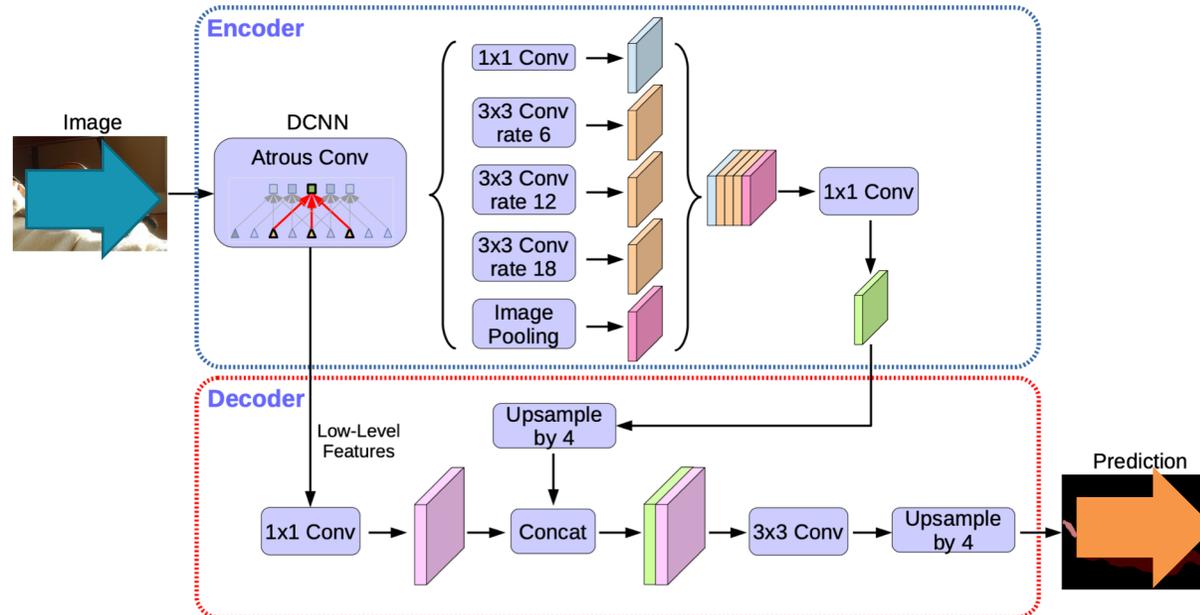
Estimate food regions with Deeplab V3+[1]

⇒ Estimate food/non food regions in areas of bounding box

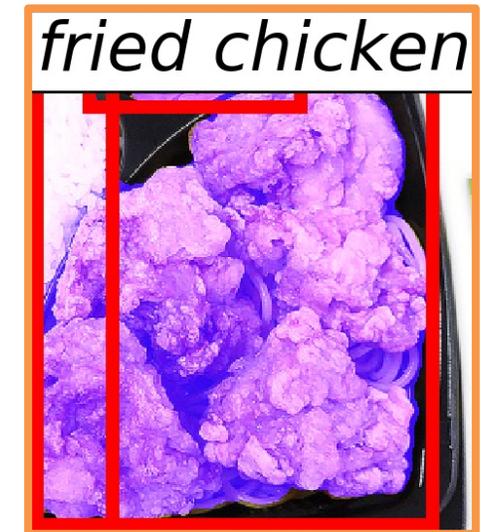
bbox areas



Estimate with Deeplab V3+



Food regions

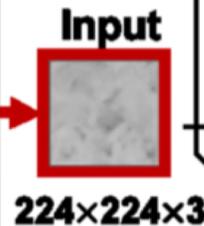
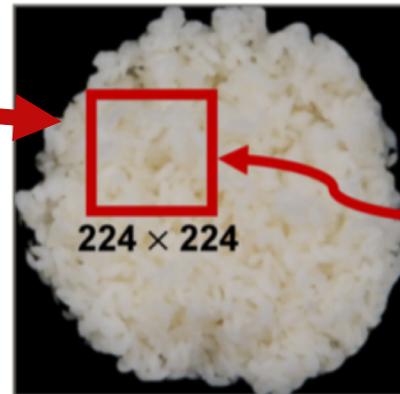


Application 1: Region-based Food Calorie Estimation

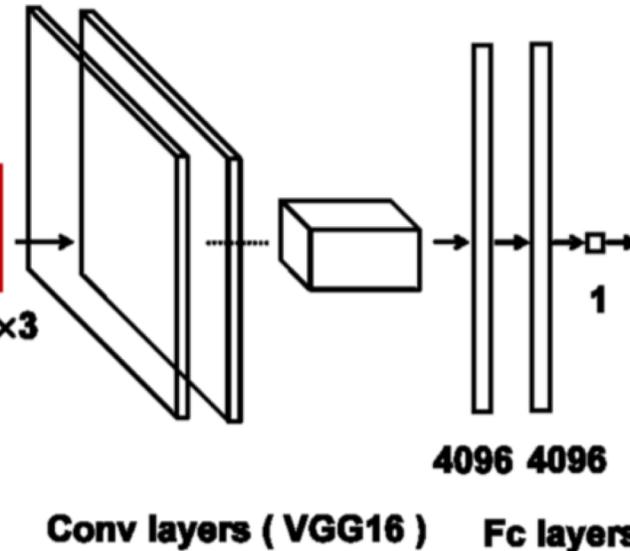
Estimating the actually area of each foods

⇒ Estimating the *cm*/pixel from rice parts in images. Using squared *cm*²/pixels

Input image



The rice system[6]



5.81 cm²

Application 1: Region-based Food Calorie Estimation

Dataset with known reference object area and calorie(63 categories)

⇒ Estimate food regions and create a regression equation for each category



Dataset

Reference card

Food region

Example

Category	Food region	Calorie
Meat spaghetti	1.0	602kcal

Calorie / Area-regression formula

