

# Segmentasi Citra berbasis Deep Learning

Nadya Aditama

# Introduction



## Nadya Aditama

S1 Informatika Universitas Telkom (2015 - 2019)

S2 Informatika Institut Teknologi Bandung (2020 - 2022)

## Research Interest

Machine Learning, Data Science, Computer Vision,  
Natural Language Processing

## Contact

[nadya.aditama@gmail.com](mailto:nadya.aditama@gmail.com)

[github.com/nadyatm](https://github.com/nadyatm)

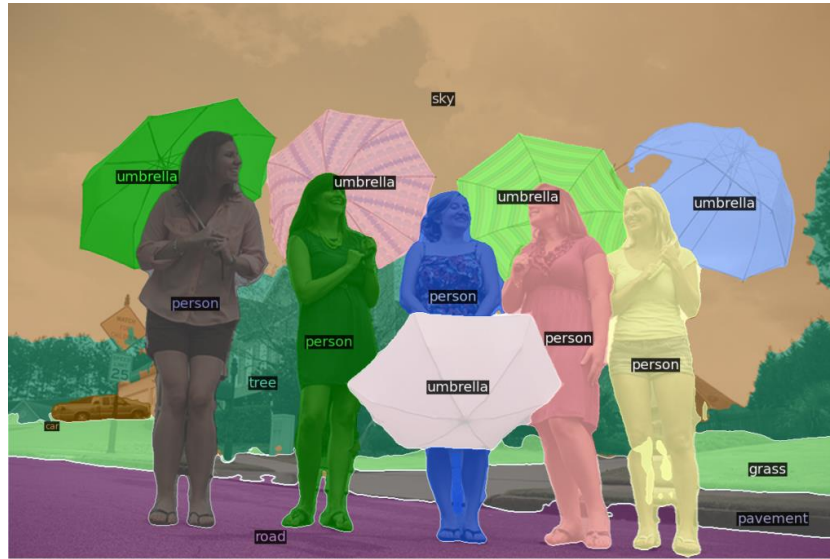
[linkedin.com/in/nadyatm17](https://www.linkedin.com/in/nadyatm17)

# Apa yang akan dipelajari hari ini?

- Segmentasi
  - Jenis Segmentasi
  - Metode Segmentasi dengan Deep Learning : Mask R-CNN
  - Metrik Evaluasi
- Pengaplikasian Segmentasi Citra
  - Master's Thesis (Indonesian Street Food Calorie Estimation using Mask R-CNN and Multiple Linear Regression)

# Perkenalan Segmentasi

# Apa itu Segmentasi?



Segmentasi merupakan sebuah proses untuk menentukan pixel mana saja dalam sebuah citra yang merupakan proyeksi dari objek yang sama dalam suatu tempat (Birchfield, 2016).

# Fakta menarik tentang segmentasi citra

Segmentasi citra mempunyai konsep yang mirip seperti object detection.

Jika objek detection hanya menandai objek dengan bounding box, maka segmentasi menandai objek per pixel.

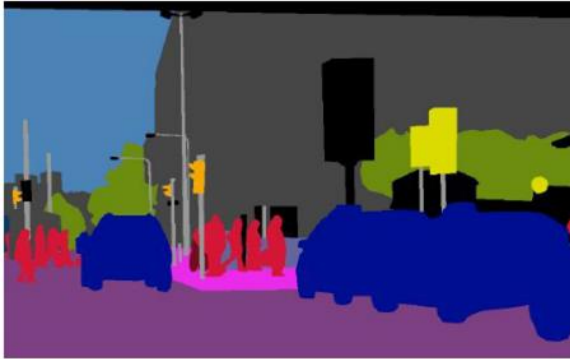
**Object  
Detection**



**Instance  
Segmentation**



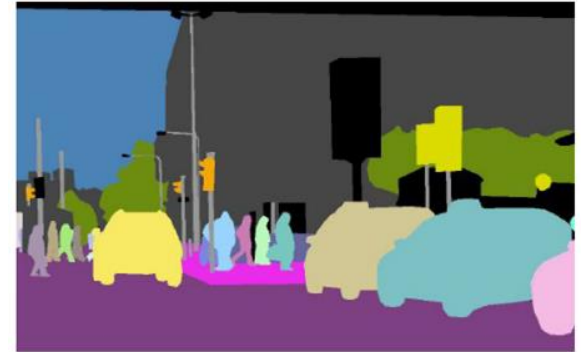
# Jenis Segmentasi Citra



Semantic  
Segmentation

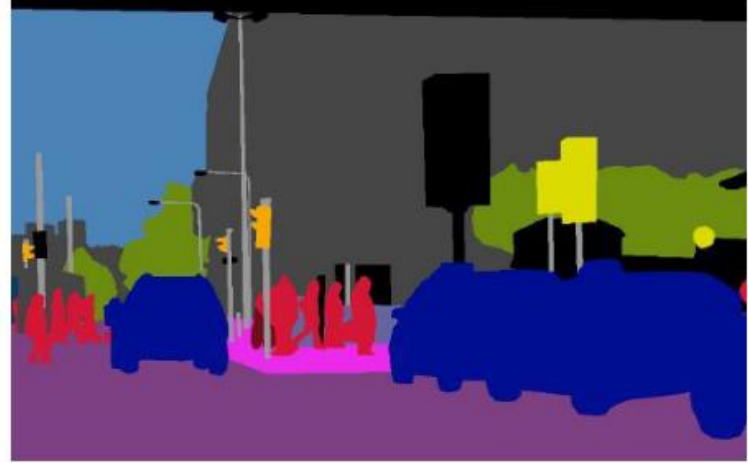


Instance  
Segmentation



Panopic  
Segmentation

# Semantic Segmentation



Sumber : Chen dkk. (2020)

Semantic Segmentation merupakan teknik segmentasi yang melakukan pelabelan tingkat pixel dengan kumpulan kategori objek dalam seluruh pixel gambar (Minaee dkk., 2021)



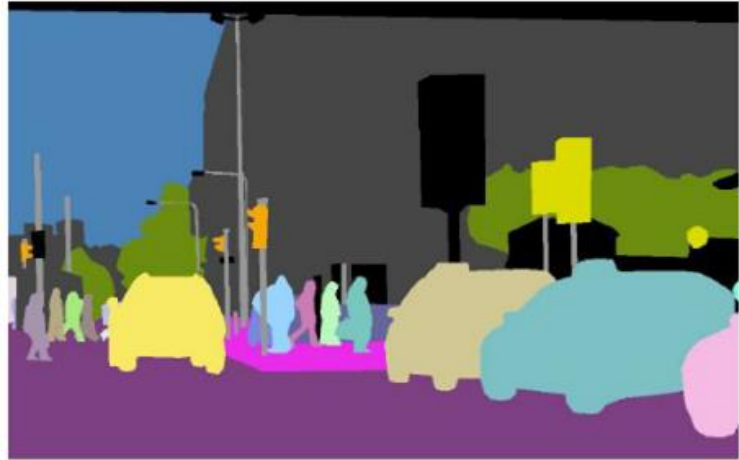
# Instance Segmentation



Sumber : Chen dkk. (2020)

Instance segmentation merupakan pengembangan dari semantic segmentation, yaitu menganggap objek yang tumpang tindih dengan kelas yang sama sebagai instance yang berbeda.

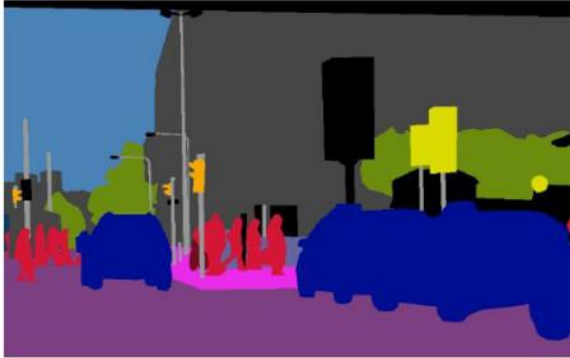
# Panoptic Segmentation



Sumber : Chen dkk. (2020)

Panoptic Segmentation merupakan gabungan antara semantic segmentation dan instance segmentation, yaitu melakukan klasifikasi per pixel, namun juga melihat objek per instance.

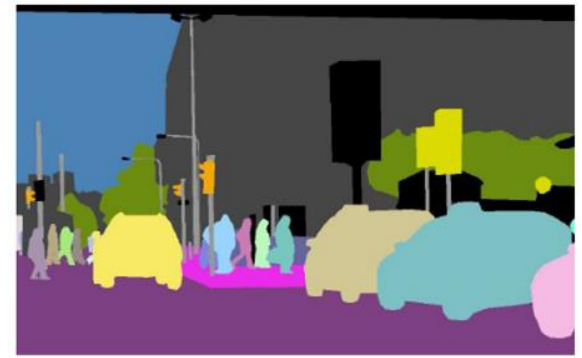
# Jenis Segmentasi Citra



Semantic  
Segmentation



Instance  
Segmentation



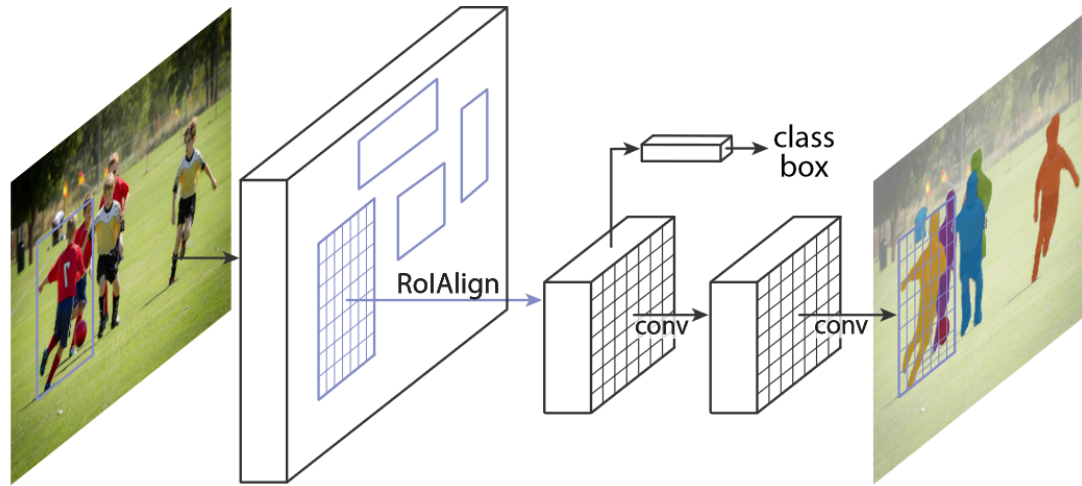
Panopic  
Segmentation

Pembahasan Metode Segmentasi  
Hari Ini

# Metode Instance Segmentation

Mask R-CNN (He dkk. 2017)

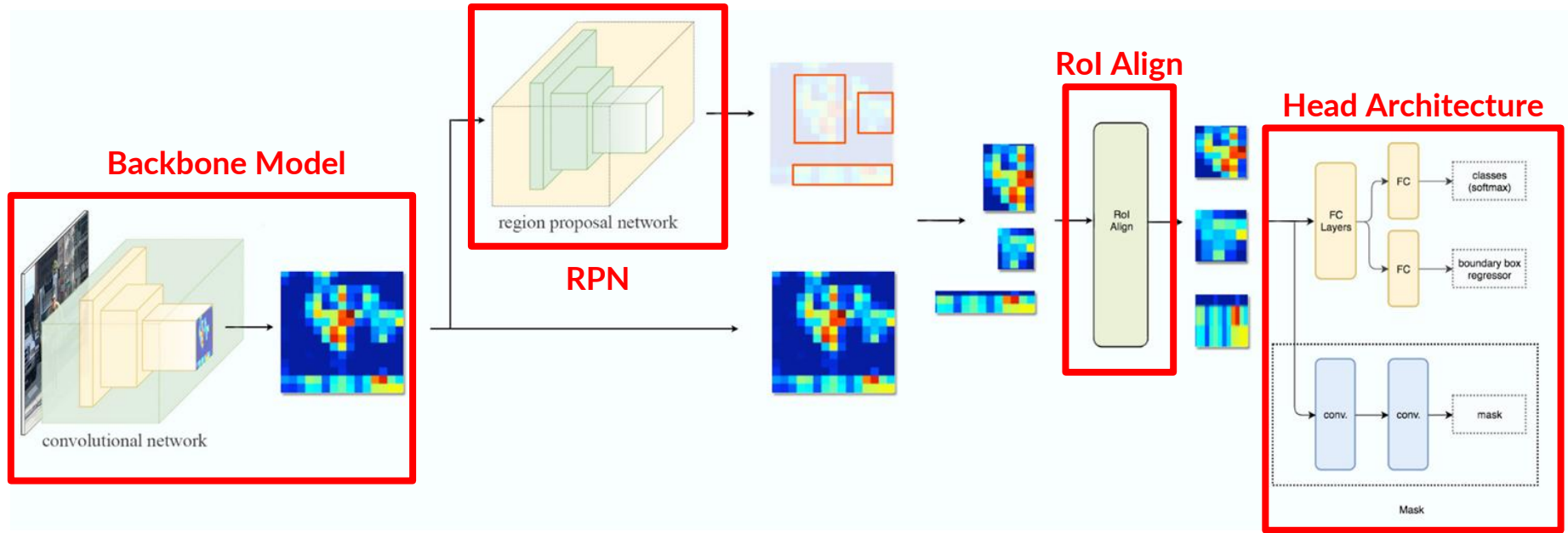
# Mask R-CNN



Sumber : He dkk. (2017)

Mask Regional Convolutional Neural Network (Mask R-CNN) adalah perluasan dari Faster R-CNN yang dikembangkan oleh penelitian (He dkk., 2017) dalam tim FAIR (Facebook AI Research).

# Bagian-bagian Mask R-CNN

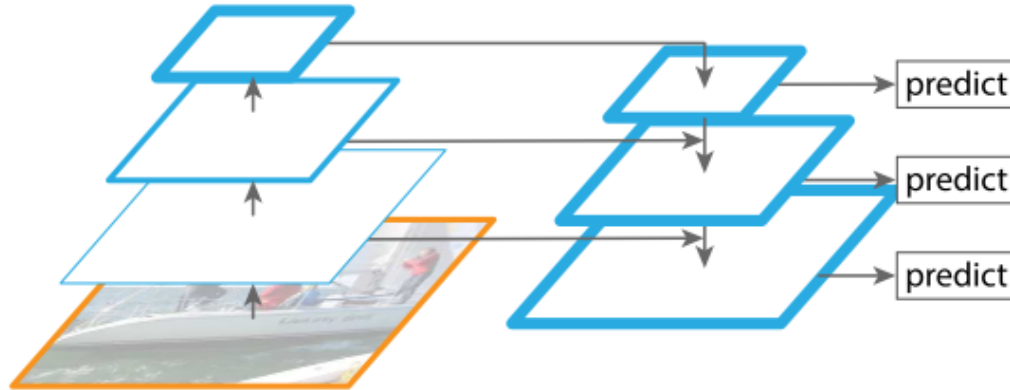


Sumber : <https://developpaper.com/mask-r-cnn/>

# Backbone Model

- Backbone merupakan arsitektur CNN utama untuk ekstraksi fitur suatu citra.
- Pada penelitian He dkk. (2017), arsitektur yang diujikan adalah ResNet dan ResNeXt dengan kedalaman 50 dan 101 layer.
- Namun, tidak menutup kemungkinan bahwa terdapat arsitektur lain yang dapat dijadikan sebagai backbone model.

# Backbone Model

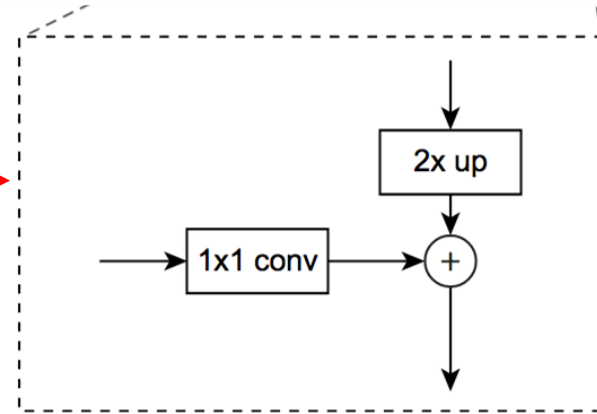
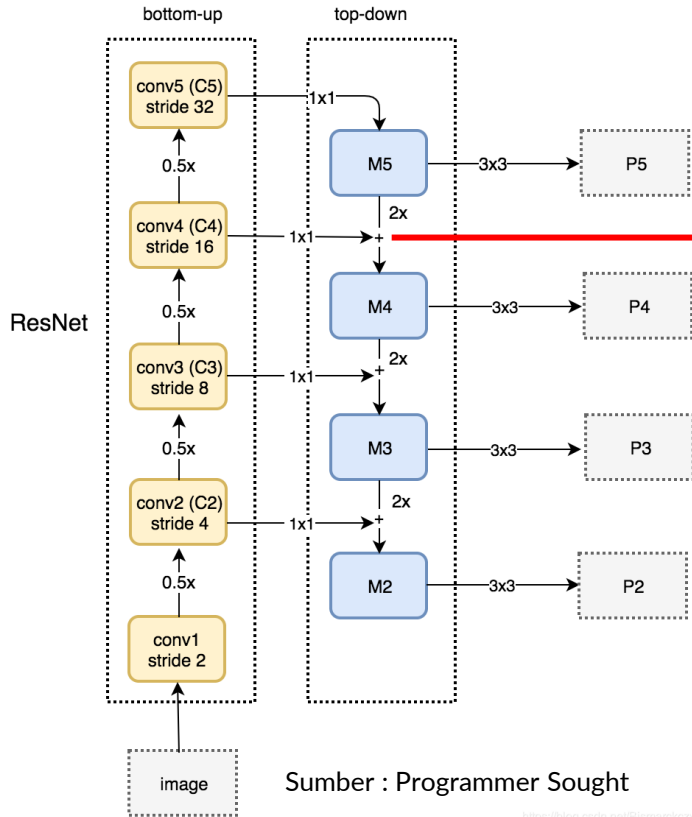


Sumber : Lin dkk. (2017)

Semakin dalam arsitektur neural network, maka feature map yang dihasilkan akan semakin kecil. Oleh karena itu, arsitektur backbone dilengkapi dengan Feature Pyramid Network (FPN).



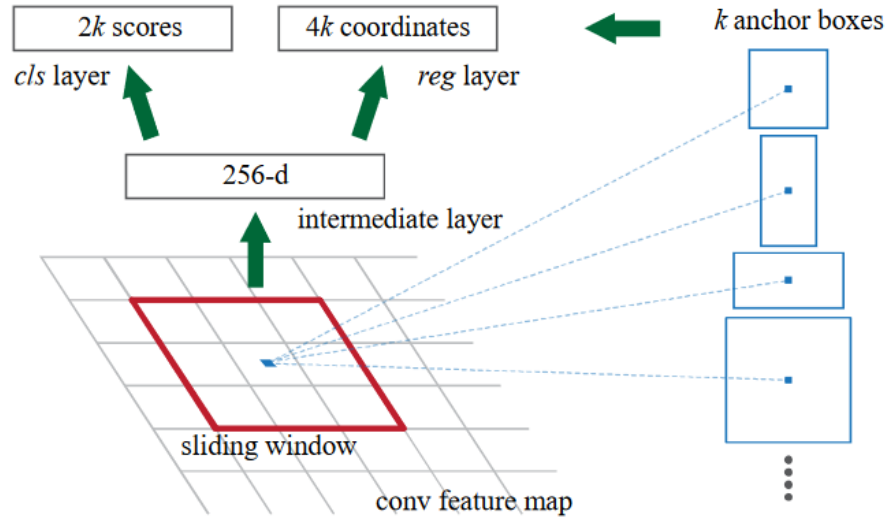
# Contoh Backbone Model ResNet + FPN



Sumber : Lin dkk. (2017)

Arsitektur terbagi menjadi top-down dan bottom up. Dalam top down, akan dilakukan penggabungan antara feature map 1 x 1 conv dengan fitur hasil upscaling.

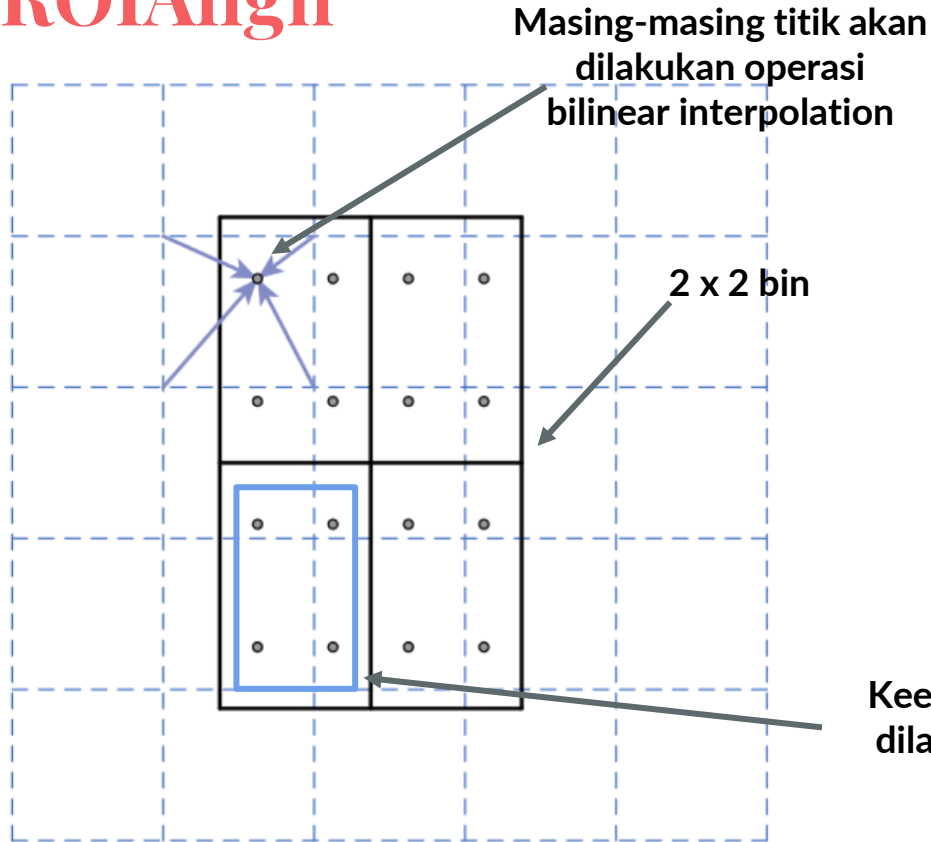
# Region Proposal Network



Sumber : Ren dkk. (2017)

Merupakan sebuah bagian yang memprediksi keberadaan objek beserta dengan bounding boxnya.  
Dilakukan sliding window dengan berbagai skala box .

# ROIAlign

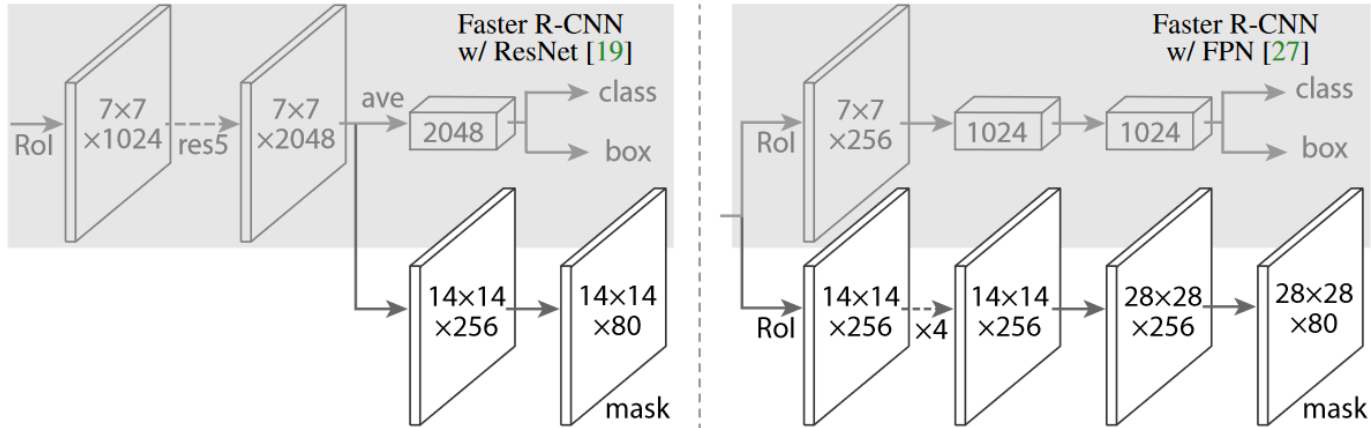


RoIAlign bertujuan untuk menyamakan dimensi dari RoI yang telah didapatkan dalam bagian RPN.

RoIAlign dilakukan dengan operasi bilinear interpolation.

Keempat poin ini dapat dilakukan operasi max atau average

# Head Architecture

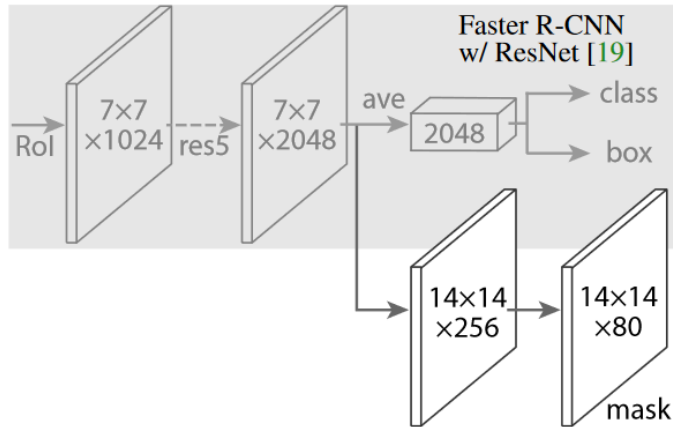


Sumber : He dkk. (2017)

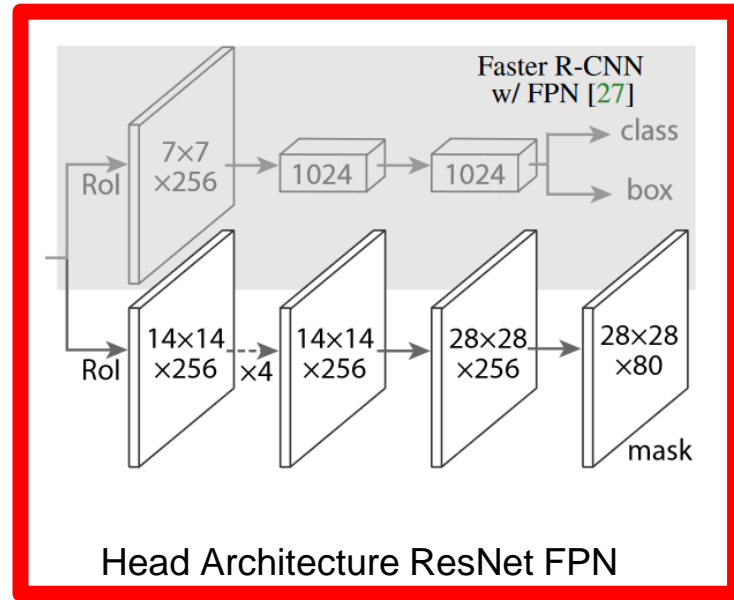
Head Architecture merupakan arsitektur terakhir pada Mask R-CNN, yang terdiri dari branch prediksi kelas dan bounding box, serta prediksi mask segmentasi

# Head Architecture

Pada paper He et al. (2017), head architecture terdiri dari dua jenis

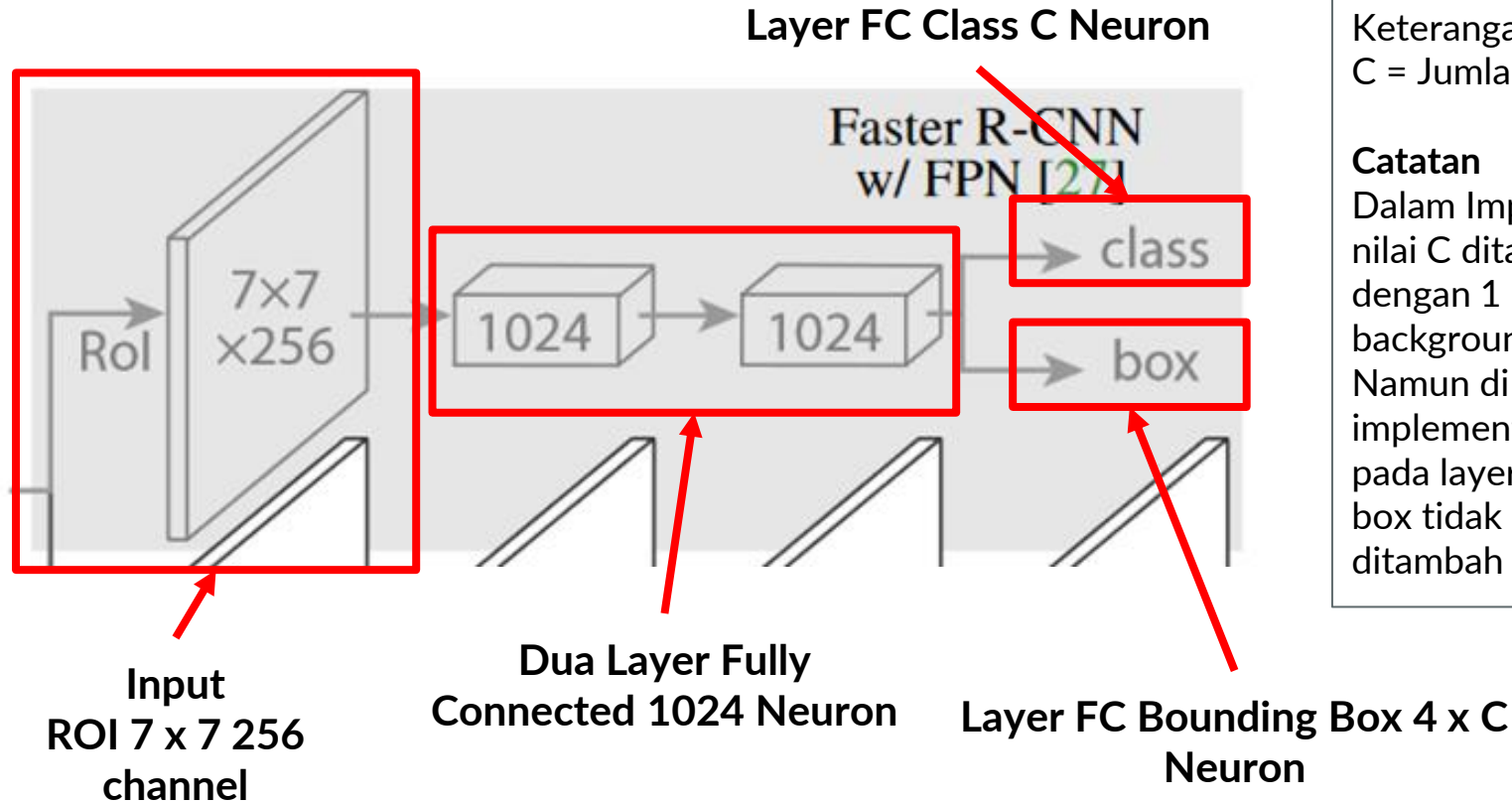


Head Architecture ResNet C4



Head Architecture ResNet FPN

# Cabang Klasifikasi dan Bounding Box

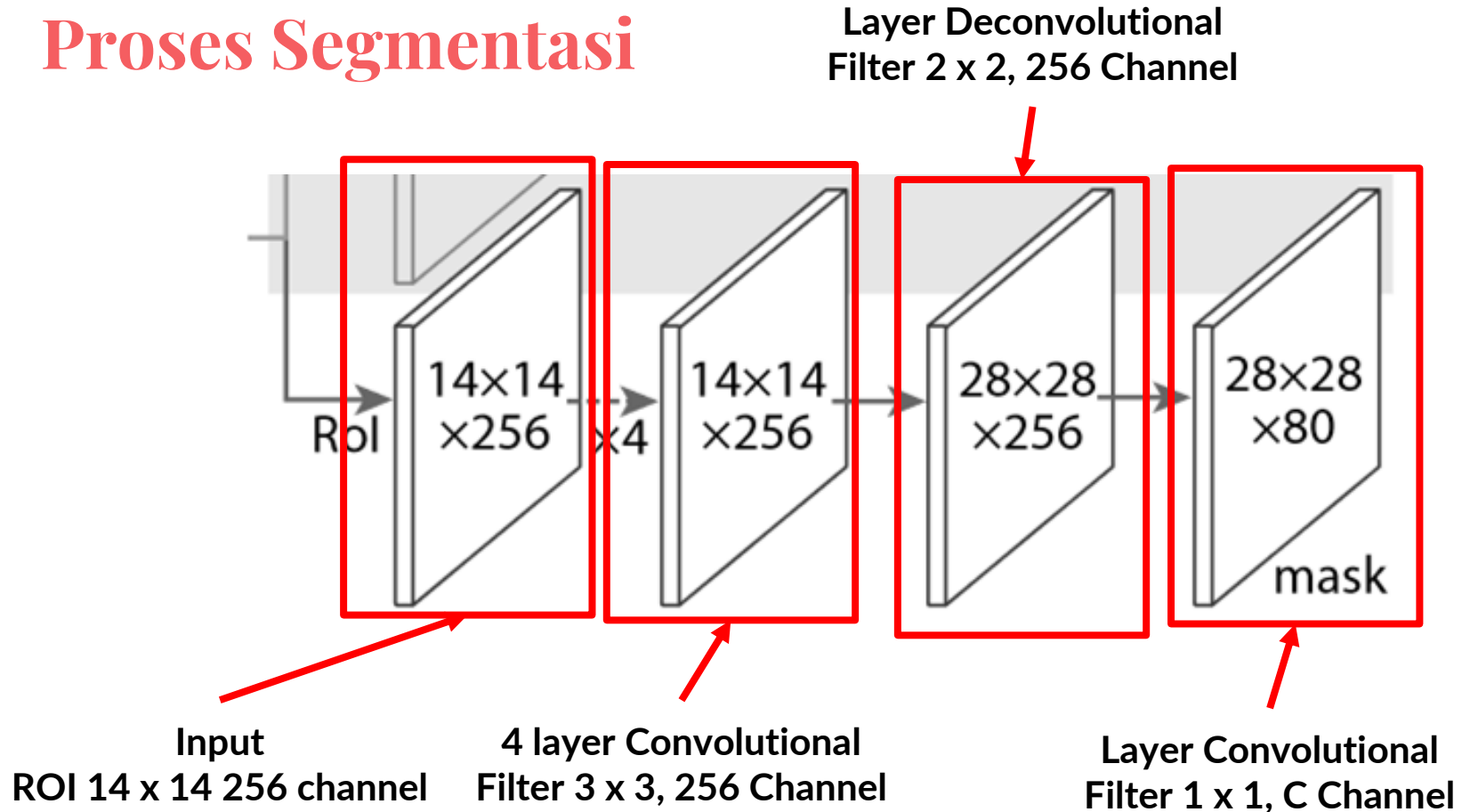


Keterangan  
C = Jumlah Kelas

## Catatan

Dalam Implementasi, nilai C ditambahkan dengan 1 karena background dihitung. Namun di beberapa implementasi nilai C pada layer bounding box tidak perlu ditambah 1.

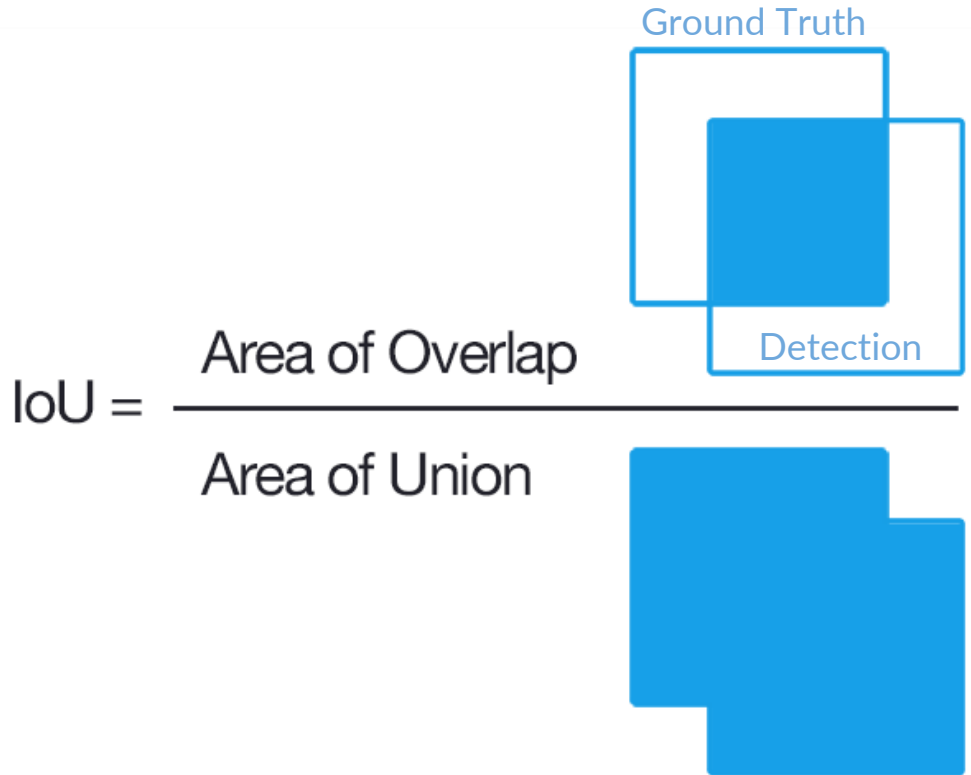
# Proses Segmentasi



# Metrik Evaluasi



# IoU (Intersection Over Union)



## Confusion Matrix

### 1. TP (True Positive)

IoU > Threshold

### 1. FP (False Positive)

IoU < Threshold / Hasil deteksi tidak berisn dengan Ground Truth

### 1. FN (False Negative)

Tidak ada hasil deteksi yang berisn dengan Ground Truth.

# Metrik Evaluasi Instance Segmentation

1. Precision, Recall, F1 Score
2. Mean Average Precision

# Precision, Recall, F1 Score

- Precision : Rasio prediksi benar dengan seluruh hasil prediksi

$$Precision = \frac{TP}{TP + FP}$$

- Recall : Rasio prediksi benar dengan keseluruhan data positif yang benar

$$Recall = \frac{TP}{TP + FN}$$

# Precision, Recall, F1 Score

F1 Score : Rata-rata harmonik precision dan recall

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall}$$

# Mean Average Precision

Merupakan salah satu metric evaluasi yang sering dipakai dalam object detection atau segmentation.

$$mAP = \sum_{i=1}^n \frac{AP_i}{n}$$

Keterangan:

AP : Average Precision dari kelas i

n : Jumlah Kelas

# Average Precision

Langkah-langkah perhitungan average precision

1. Menghitung precision dan recall dari n titik threshold IoU.

Ex : range 0.5 - 0.95 (0.5, 0.55, 0.6, 0.65, ... 0.95)

1. Membuat precision recall curve.
2. Menghitung average precision.

$$AP = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k + 1)] * Precisions(k)$$

**$Recalls(n) = 0, Precisions(n) = 1$**

**$n = \text{Number of thresholds.}$**

# Pengaplikasian Segmentasi Citra

Master's Thesis

Indonesian Street Food Calorie Estimation using  
Mask R-CNN and Multiple Linear Regression

# Permasalahan

Terdapat dua permasalahan pada estimasi kalori pada citra

1. Mengukur kalori pada objek yang saling tumpang tindih.
2. Terdapat permasalahan pada nilai R Squared yang rendah pada pengukuran berbasis regresi linear dengan satu fitur menurut penelitian Abdelhardy et al. (2019).

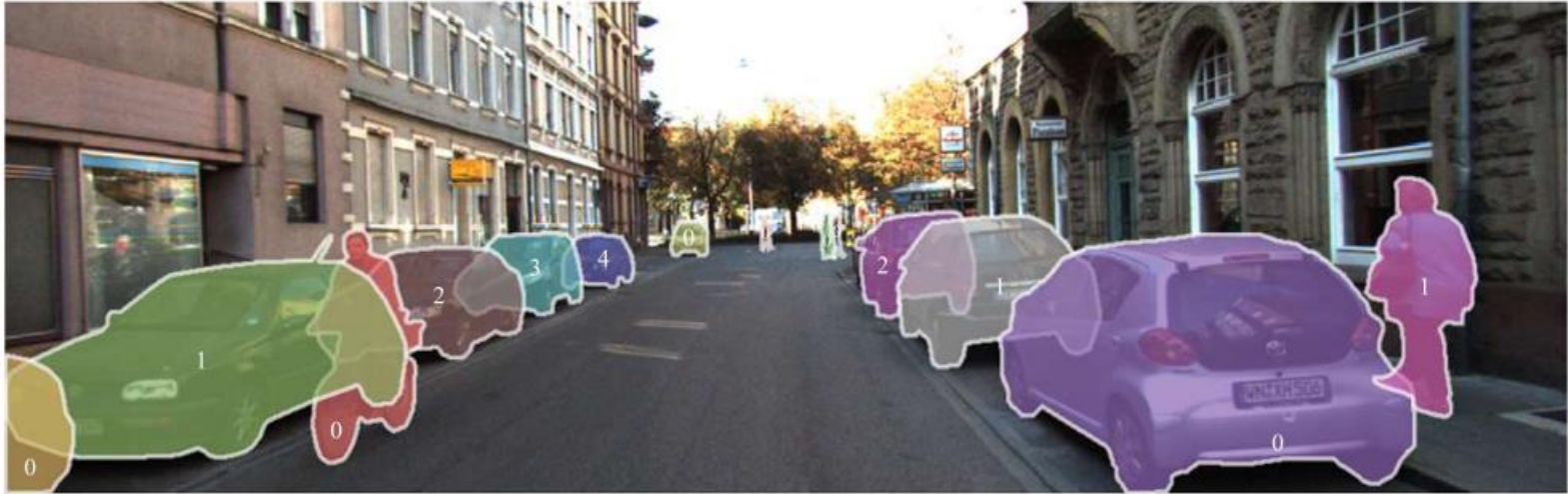


# Solusi

Terdapat dua solusi pada permasalahan tersebut

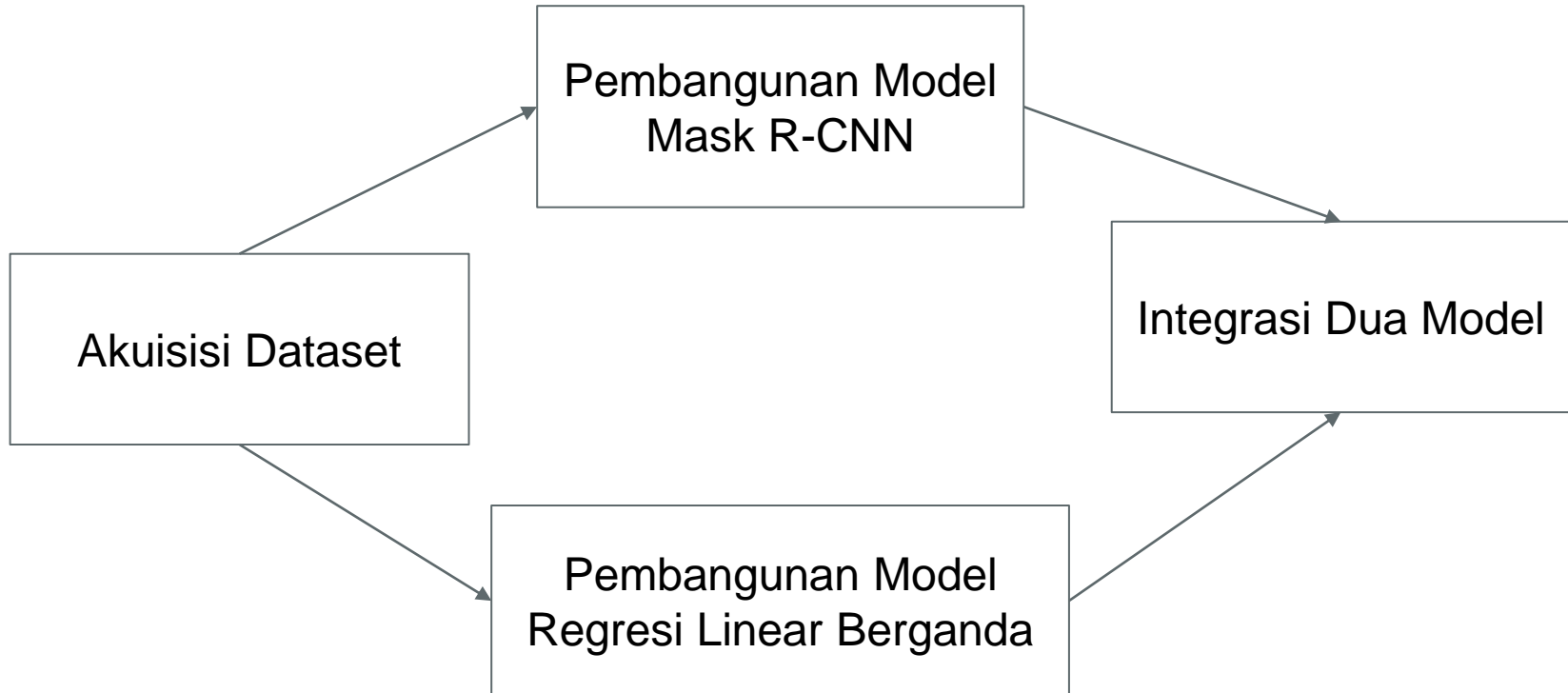
1. Pembangunan dataset jajanan pasar Indonesia yang terdiri dari 6 kelas (bakwan, bolu, cireng, serabi, tahu, tempe).
2. Melatih model Mask R-CNN untuk task Amodal Instance Segmentation.
3. Melatih model regresi linear dengan fitur yang lebih dari satu (Multiple Linear Regression)

# Amodal Instance Segmentation

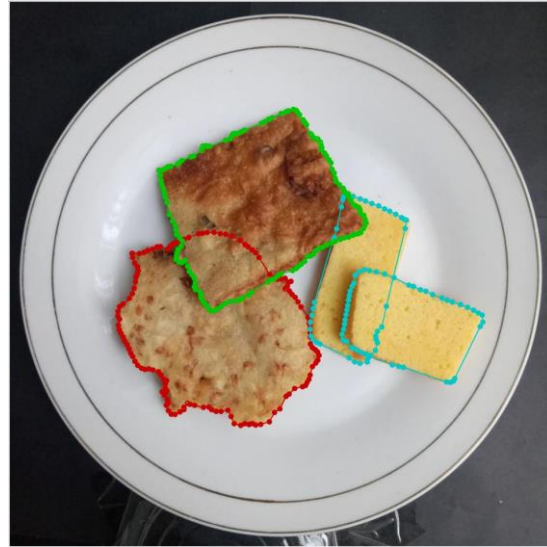


Amodal instance segmentation melakukan segmentasi pada setiap instance objek yang melibatkan bagian yang teroklusi untuk meniru kemampuan manusia dalam interpretasi objek (Qi et al., 2019).

# Tahap Pembangunan Sistem



# Pembangunan Dataset Mask R-CNN



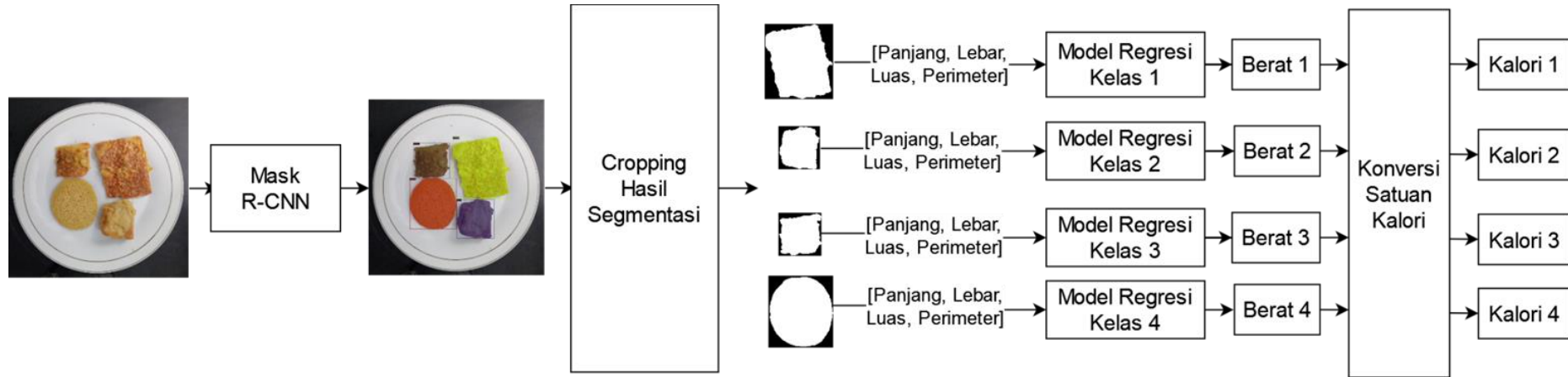
Gambar diambil dari berbagai posisi (baik berjauhan, berdekatan dan tumpang tindih) dan diannotasi secara amodal

# Pembangunan Dataset Multiple Linear Regression



Masing-masing item akan ditimbang beratnya dan gambar diambil sebanyak 10 kali per item dengan posisi yang berbeda-beda

# Alur Sistem



# Skenario Pengujian



Occluded Scenario



Non Occluded Scenario

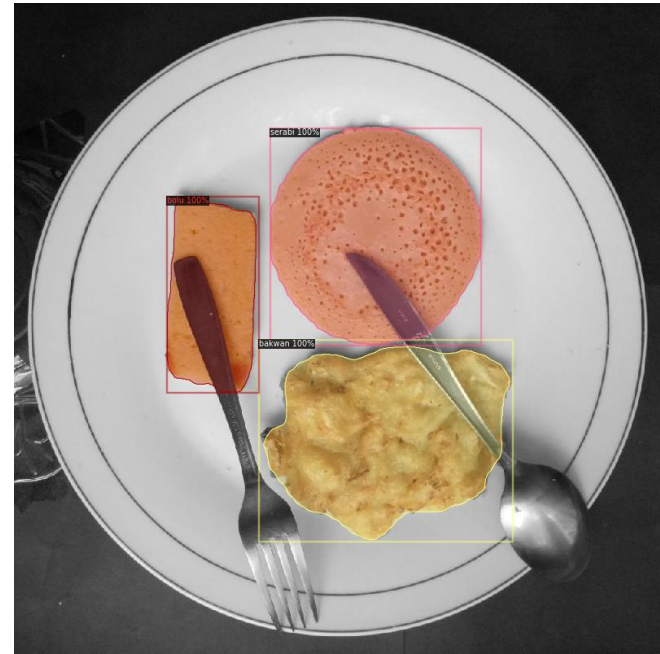
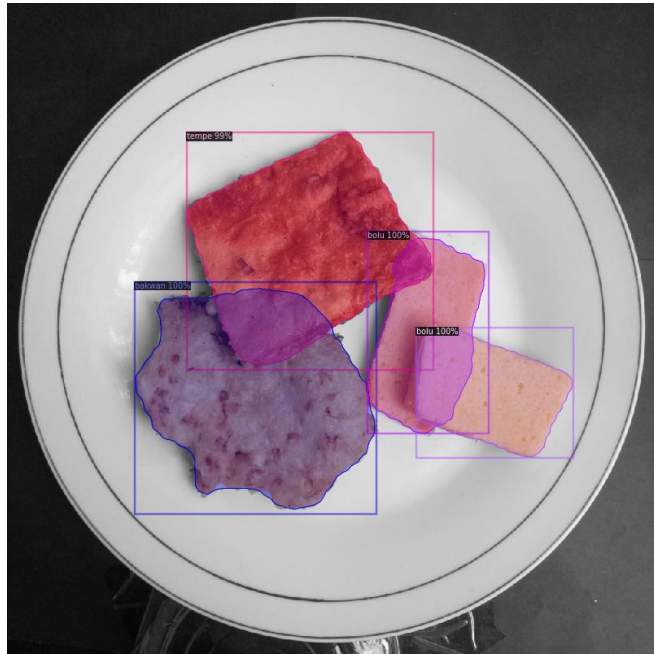
# Performansi Segmentasi Data Test

IoU	Scenario	Backbone Model	Precision	Recall	F1 Score
0.9	Occluded	ResNet-101-FPN	0.697	0.718	0.706
		ResNeXt-101-FPN	0.716	0.722	0.719
	Non-Occluded	ResNet-101-FPN	0.967	0.971	0.966
		ResNeXt-101-FPN	0.992	0.996	0.994
0.85	Occluded	ResNet-101-FPN	0.799	0.823	0.81
		ResNeXt-101-FPN	0.818	0.824	<b>0.821</b>
	Non-Occluded	ResNet-101-FPN	0.967	0.971	0.966
		ResNeXt-101-FPN	0.992	0.996	<b>0.994</b>

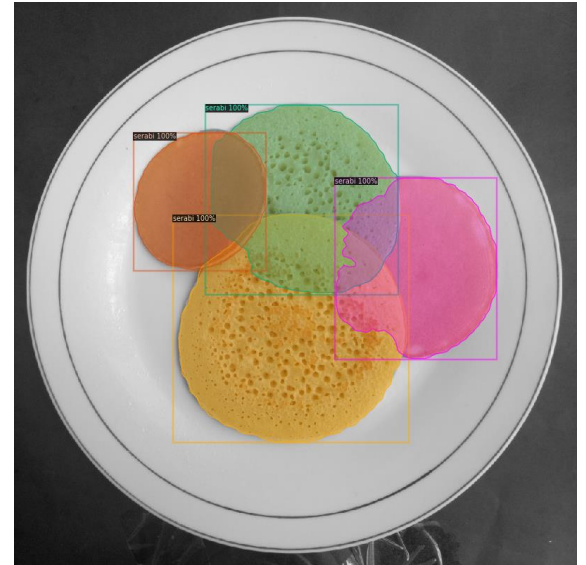
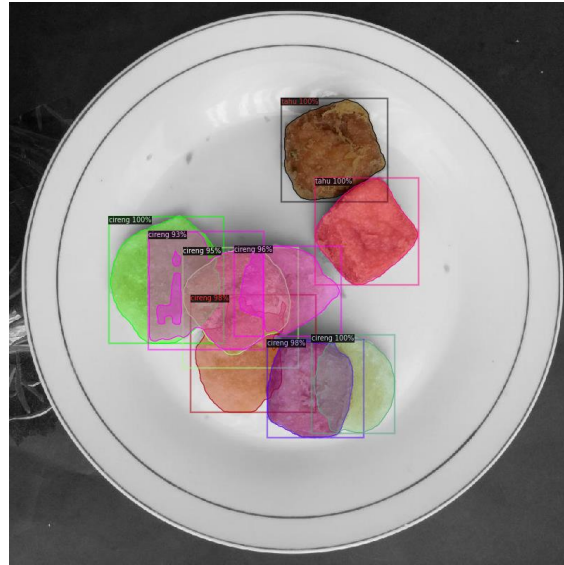
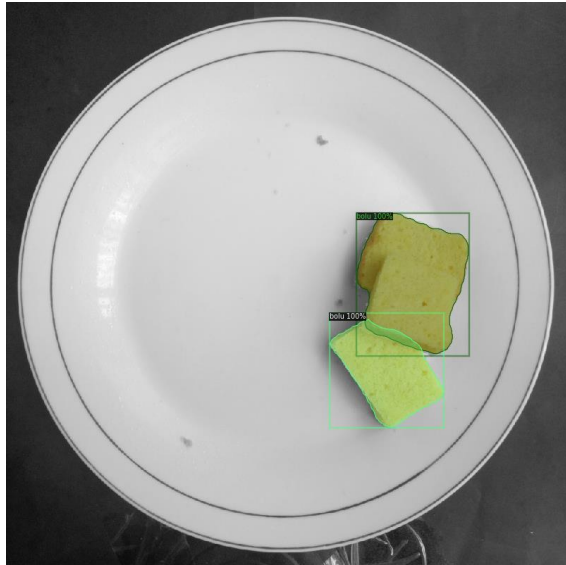
- Performansi Segmentasi dihitung dengan rata-rata precision, recall, f1 score seluruh kelas.
- Dilakukan juga obesrvasi threshold IoU 0.85 dan 0.9



# Contoh Hasil Segmentasi

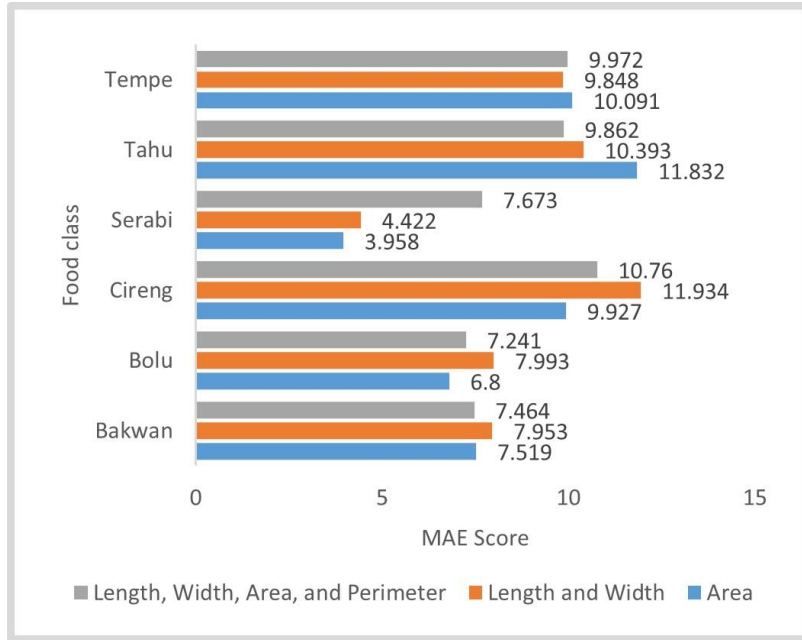


# Kasus Gagal

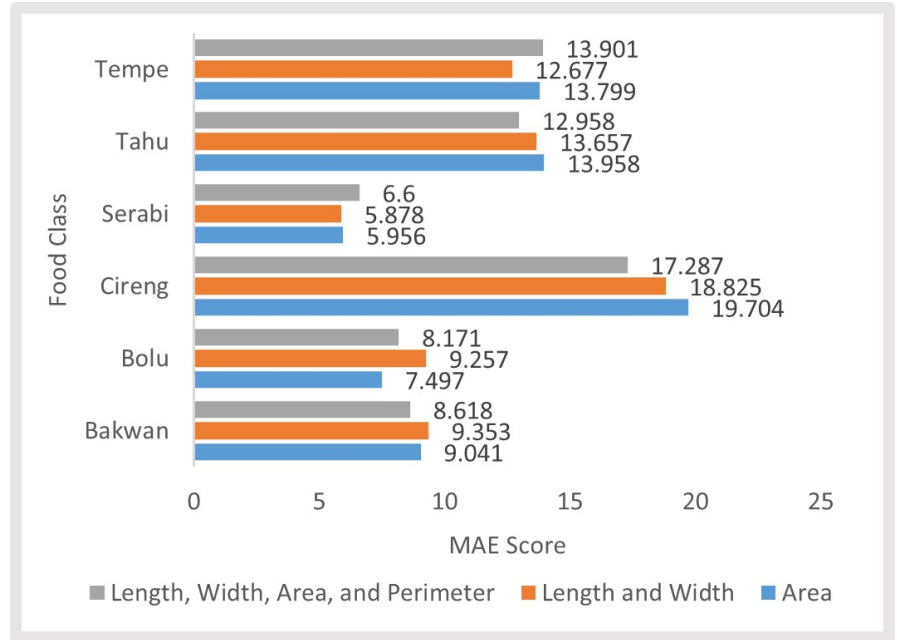


# Hasil Estimasi Kalori

Performansi diukur dengan metrik Mean Absolute Error (MAE)



**Occluded Scenario**



**Non Occluded Scenario**

# Kesimpulan

- Meskipun Mask R-CNN dengan backbone ResNeXt mempunyai performansi yang paling baik, namun masih ada beberapa kesalahan segmentasi.
- Lebih banyak fitur belum tentu dapat menurunkan error pengukuran.

# What Next?

- Pelatihan dengan dataset yang mempunyai lebih banyak variasi posisi tumpang tindih.
- Mencoba teknik instance segmentation yang lain untuk kasus amodal instance segmentation.
- Mencoba teknik amodal instance segmentation yang sudah ada untuk memprediksi bentuk tumpang tindih barang.

# Referensi

- K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.
- T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936-944.
- L. Qi, L. Jiang, S. Liu, X. Shen and J. Jia, "Amodal Instance Segmentation With KINS Dataset," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3009-3018.
- Birchfield, S. (2016): Image Processing and Analysis, Cengage Learning, Amerika Serikat, 2-11.

## Referensi (2)

- S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence.
- C. Chen, B. Wang, C. Xiaoxuan Lu, N. Trigoni, and A. Markham, "A Survey on Deep Learning for Localization and Mapping: Towards the Age of Spatial Machine Intelligence," arXiv, 2020
- A. S. Abdelhady, A. E. Hassanien, Y. M. Awad, M. El-Gayar, and A. Fahmy, "Automatic Sheep Weight Estimation Based on K-Means Clustering and Multiple Linear Regression," Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, pp. 546-555, 2019.

## Referensi (2)

- S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.



# Additional Resources

- <https://www.kdnuggets.com/2021/03/evaluating-object-detection-models-using-mean-average-precision.html>
- <https://programmersought.com/article/46924759663/>

# Demo Program

Let's See How it Works