Machine Learning Engineer Course
Day 26

- U-Net -

DIVE INTO CODE

Thursday August 19, 2021

DIOP Mouhamed

Agenda

1. Check-in
2. Quick Review
3. Image recognition
4. Object detection
5. Semantic segmentation
6. Sample code
7. To do by next class
8. Check-out

© 2020 DIVE INTO CODE Corp.

Check-in

3 minutes Please post the following point to Zoom chat.

Q. What did you learn in the previous week?
(Anything is fine.)

© 2020 DIVE INTO CODE Corp.

Quick Review (Dataset Creation)

Data Creation (Augmentation)

- Doing classification with your own dataset (cats and dogs)
- Generate different images (Augmentation. Flipping, changing colors, cropping, etc.…)
- Creating annotations (labelimg, …)
- Object detection

© 2020 DIVE INTO CODE Corp.

Tasks required to achieve image recognition

Visual recognition task in the computer vision domain:

Image classification (a):

The goal is to recognize the semantic categories class of objects in a particular image.

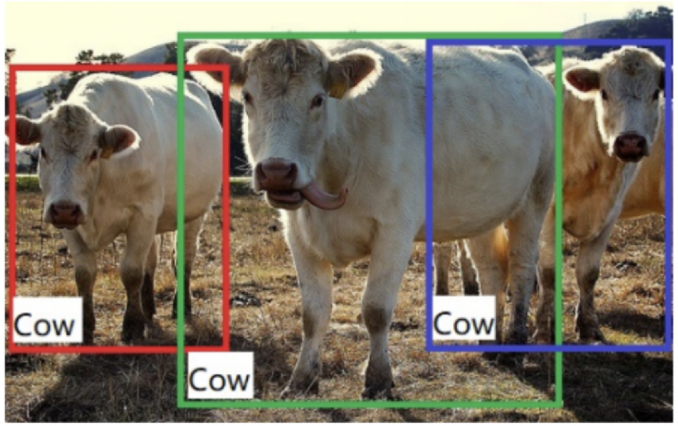
Object detection (b):

The objective is not only to recognize the category class of the object, but also to predict the location (coordinates) of each detected object by means of an imaginary box surrounding the object, called a bounding box.

In addition to box-based methods, there are also keypoint-based methods for object detection, and the latter is currently dominant in benchmarking COCO data sets.

https://paperswithcode.com/sota/real-time-object-detection-on-coco

(a) Image Classification



(b) Object Detection

© 2020 DIVE INTO CODE Corp.

Tasks required to achieve image recognition

Visual recognition task in the computer vision domain:

Semantic Segmentation (c):

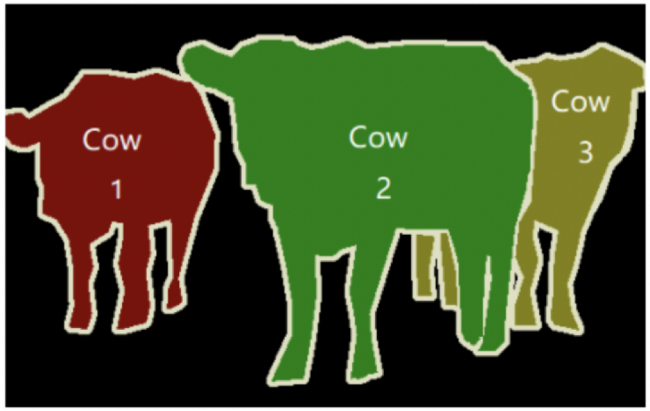
The goal is to perform classification on a pixel-by-pixel basis and assign a specific category label to each pixel. Unlike object detection, semantic segmentation does not distinguish between multiple objects of the same category.

(c) Semantic Segmentation

Instance segmentation (d):

It can be seen as a special form of object detection, where the goal is to identify the position (coordinates) at the pixel level instead of locating the object by a bounding box.

https://arxiv.org/pdf/1908.03673.pdf



(d) Instance Segmentation

© 2020 DIVE INTO CODE Corp.

What was the object detection doing?

Object detection before deep learning:

In the initial stage, the object detection pipeline was divided into three steps.

1. Proposal generation
2. Extraction of feature vectors
3. Domain Classification

The suggestion in the first step is to search regions in the image to find areas that may contain the target. These locations were also referred to as regions of interest (ROI).

What was being done in the search was to resize the input image to different scales and scan the entire image using a multi-scale sliding window.

In the next step, fixed-length feature vectors were obtained from a sliding window at various locations in the image to obtain semantic information that identifies the region.

In the last step, a classifier (typically using SVM) is trained and class labels are assigned to the target regions.

https://arxiv.org/pdf/1908.03673.pdf

© 2020 DIVE INTO CODE Corp.

Some initial thoughts

Let's try object detection with an image classification model.

First, split the image.

(Split the image regardless of the position of the target)

We then apply CNN to all the regions and classify the regions.

Then combine these regions to return to the original image with the detected object.

1. First, we take an image as input:

2. Then we divide the image into various regions:

https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/

© 2020 DIVE INTO CODE Corp.

Some initial thoughts

Oh, no

Problems:

An object may occupy a large part of the space on a given screen (and be cut off from the screen), while the same object may occupy only a few percent of the space on a different screen.

The shape or part of the object may be different in each area depending on how it is photographed.

3. We will then consider each region as a separate image.
4. Pass all these regions (images) to the CNN and classify them into various classes.
5. Once we have divided each region into its corresponding class, we can combine all these regions to get the original image with the detected objects:

© 2020 DIVE INTO CODE Corp.

Some initial thoughts

what to do

Result:

It requires a large number of region parts that require a huge amount of computation. It will take a lot of time if we split the image and apply CNN from one side to the other. We want to reduce this area somehow.

Ideas:

Why don't we just choose an area? That's why there is a step called domain proposal.

3. We will then consider each region as a separate image.
4. Pass all these regions (images) to the CNN and classify them into various classes.
5. Once we have divided each region into its corresponding class, we can combine all these regions to get the original image with the detected objects:

© 2020 DIVE INTO CODE Corp.

Considered methods

R-CNN（2014）

Region Proposals:

Object Recognition using selective search to achieve region proposal.

The CNN architecture is based on AlexNet (2012). The input should be 227x227.

https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

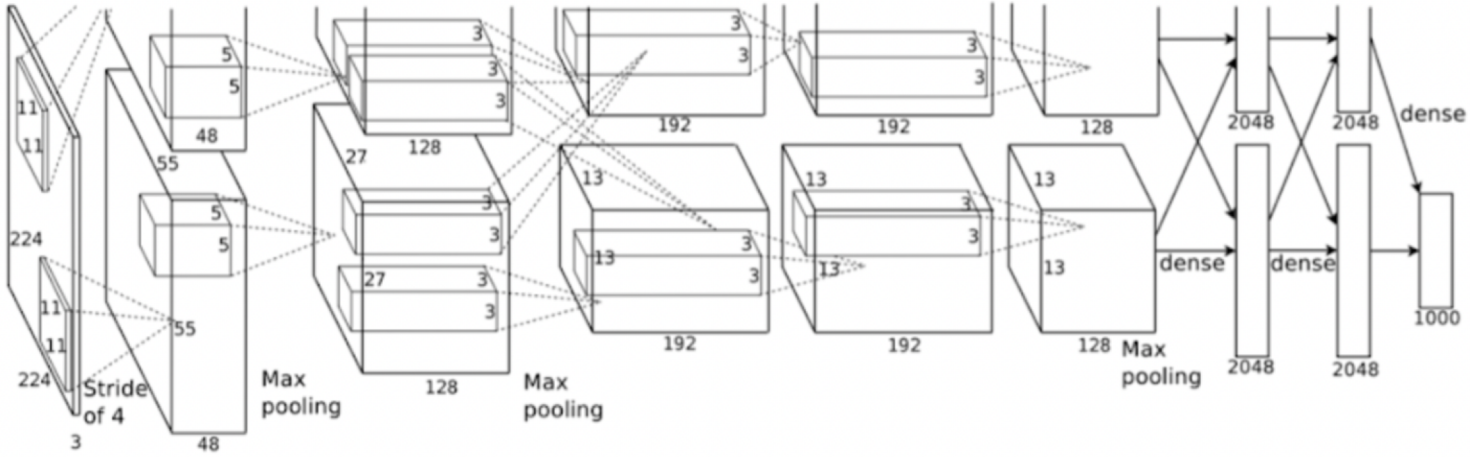


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

© 2020 DIVE INTO CODE Corp.

Considered methods

R-CNN（2014）

Warping (image transformation):

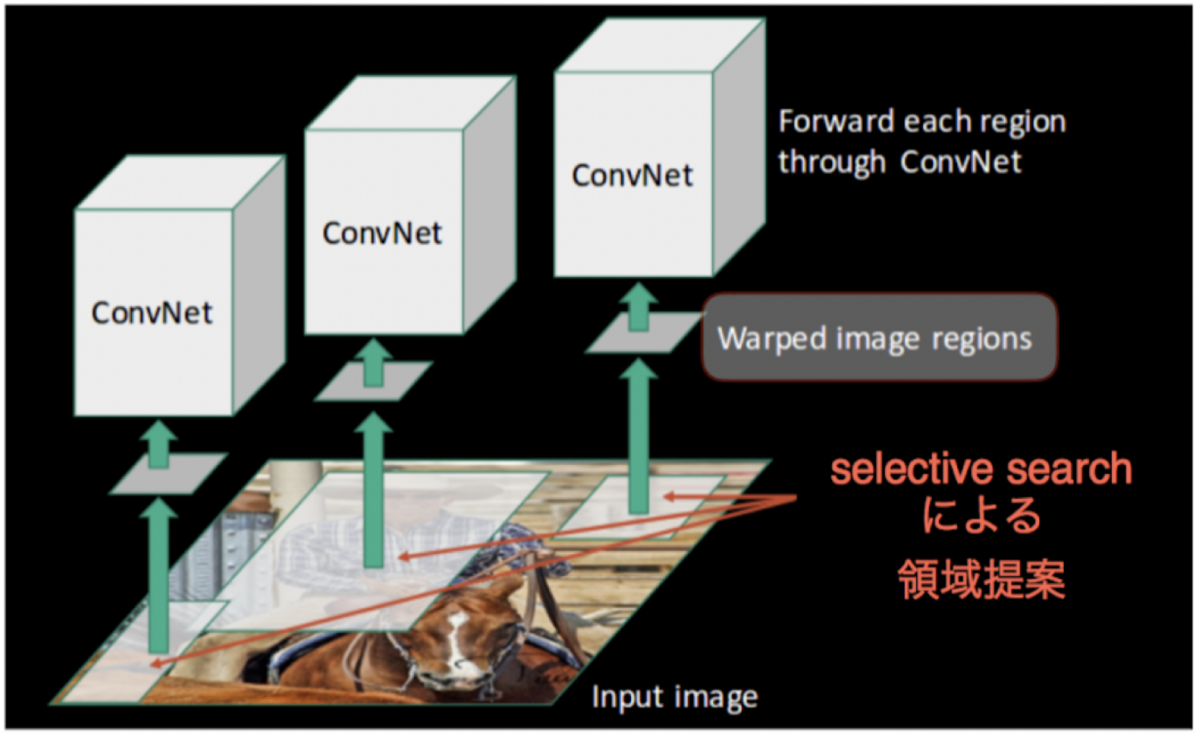
Image Distortion Removal https://www.youtube.com/watch?v=6DtzCZVorSw

CNN (special extract).

If there are 2,000 proposed regions, run the CNN 2,000 times for each image

Result:

Each image takes 40-50 seconds.



© 2020 DIVE INTO CODE Corp.

Considered methods

selective search

A huge number of boxes are proposed for object recognition.

The selective search uses a segmentation algorithm (note: different from DCNN segmentation).

"The most natural way to get (the object's) location at all scales is to use Hierarchical segmentation algorithm.

"Use a greedy algorithm that starts with an initial region and iteratively groups together the two most similar regions.

"It calculates the similarity between this new region and its neighbors.

(from 'Segmentation as Selective Search for Object Recognition')

https://www.koen.me/research/pub/vandesande-iccv2011.pdf

http://www.huppelen.nl/publications/selectiveSearchDraft.pdf

https://ivi.fnwi.uva.nl/isis/publications/bibtexbrowser.php?key=UijlingsIJCV2013&bib=all.bib

- It first takes an image as input:

- Then, it generates initial sub-segmentations so that we have multiple regions from this image:

- The technique then combines the similar regions to form a larger region (based on color similarity, texture similarity, size similarity, and shape compatibility):

- Finally, these regions then produce the final object locations (Region of Interest).

© 2020 DIVE INTO CODE Corp.

Considered methods

Fast R-CNN（2015）

CNN (special extract).

Output a single feature map for an image

Region Proposals:

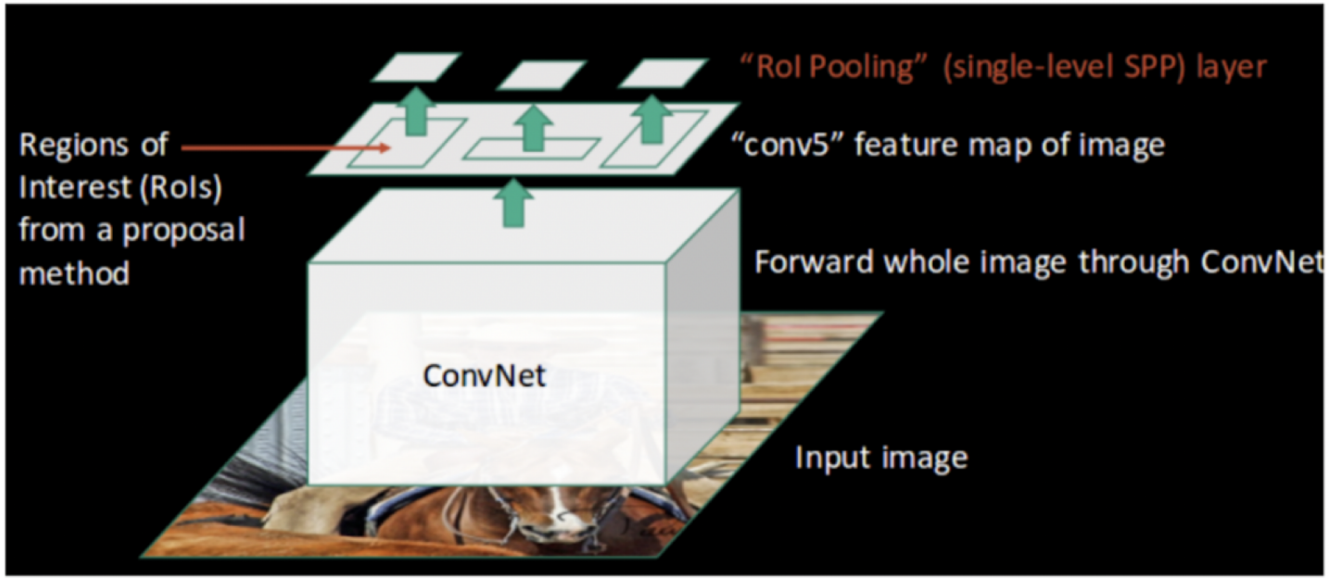
Perform a selective search on the feature map

Result:

Selective search still

takes a long time.

2 seconds per image.



© 2020 DIVE INTO CODE Corp.

Considered methods

Fast R-CNN（2015）

CNN (feature extraction) :

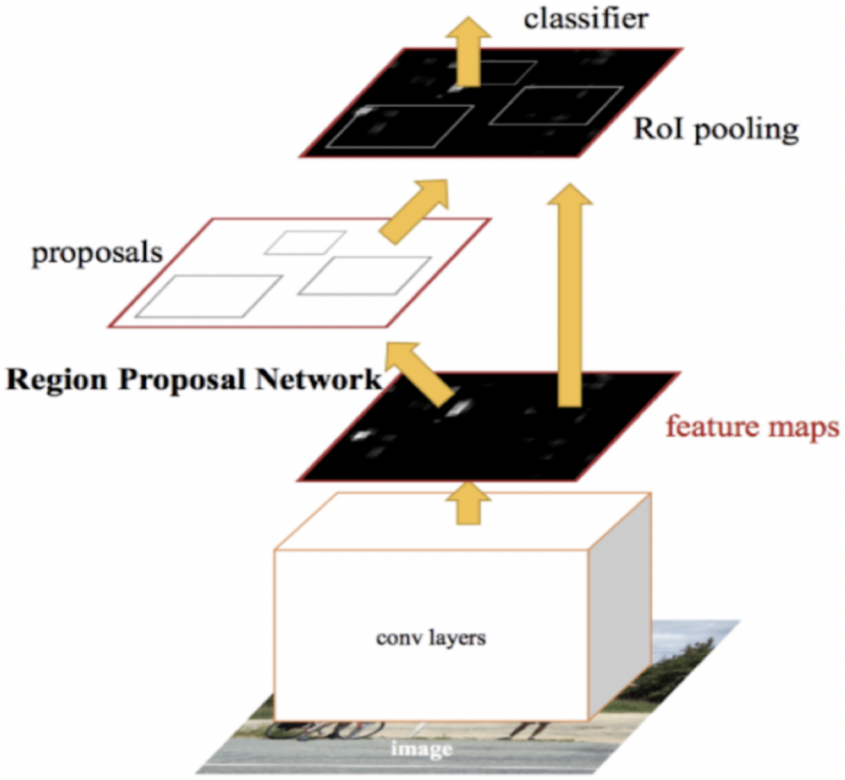
Output a single feature map for an image

Region Proposal Network :

Instead of selective search, we use a sliding window CNN) with k anchor boxes.

Return the probability of an object (not class) and regress the anchor box coordinates to fit the ground truth (using PRN loss)

https://medium.com/lsc-psd/faster-r-cnn%E3%81%AB%E3%81%8A%E3%81%91%E3%82%8Brpn%E3%81%AE%E4%B8%96%E7%95%8C%E4%B8%80%E5%8



RoI pooling layer :

Trimming (scaling) the RPN proposals

https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272

Result:

It is better than selective search, but the calculation of proposals is also rather time consuming: 0.2 seconds per image. It is running on multiple systems, so the performance depends on the other systems.

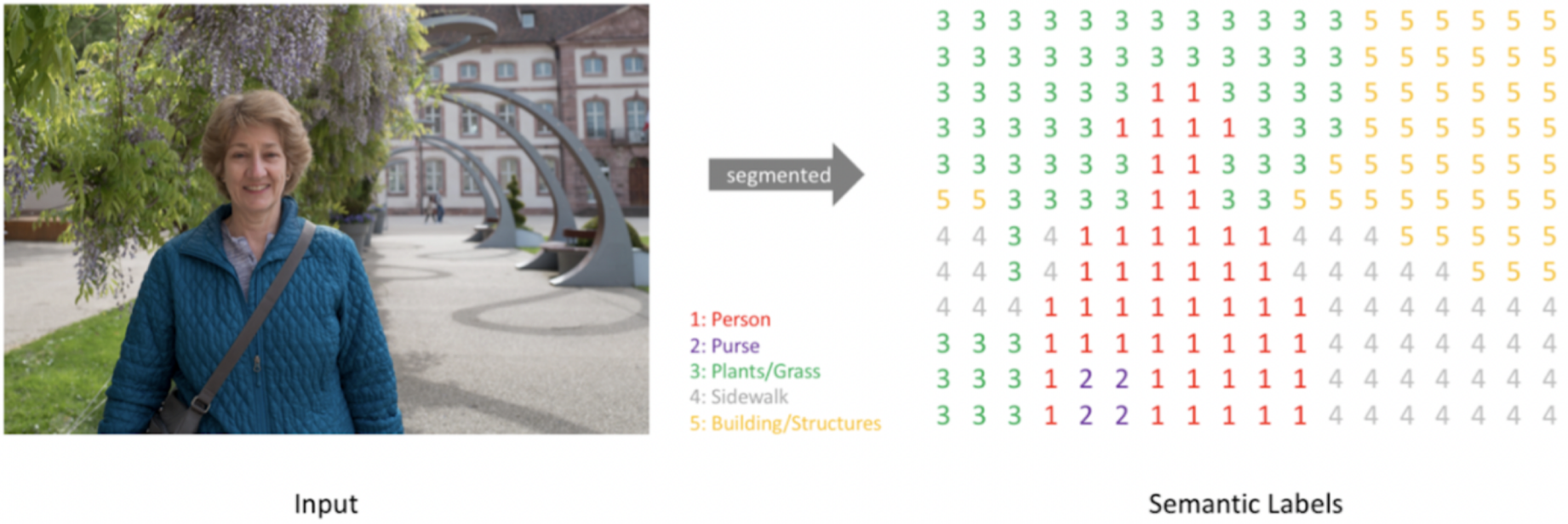
https://medium.com/@whatdhack/a-deeper-look-at-how-faster-rcnn-works-84081284e1cd

© 2020 DIVE INTO CODE Corp.

Semantic segmentation

Semantic segmentation

While Bbox-level algorithms localize the object in a rectangular box, segmenting the object with a mask-level algorithm requires more precise identification of the class (also called dense prediction due to the need for dense prediction) at the pixel level. To segment a target with a mask (i.e., grouping pixels in a meaningful way) level algorithm, the class must be more accurately identified at the pixel level.



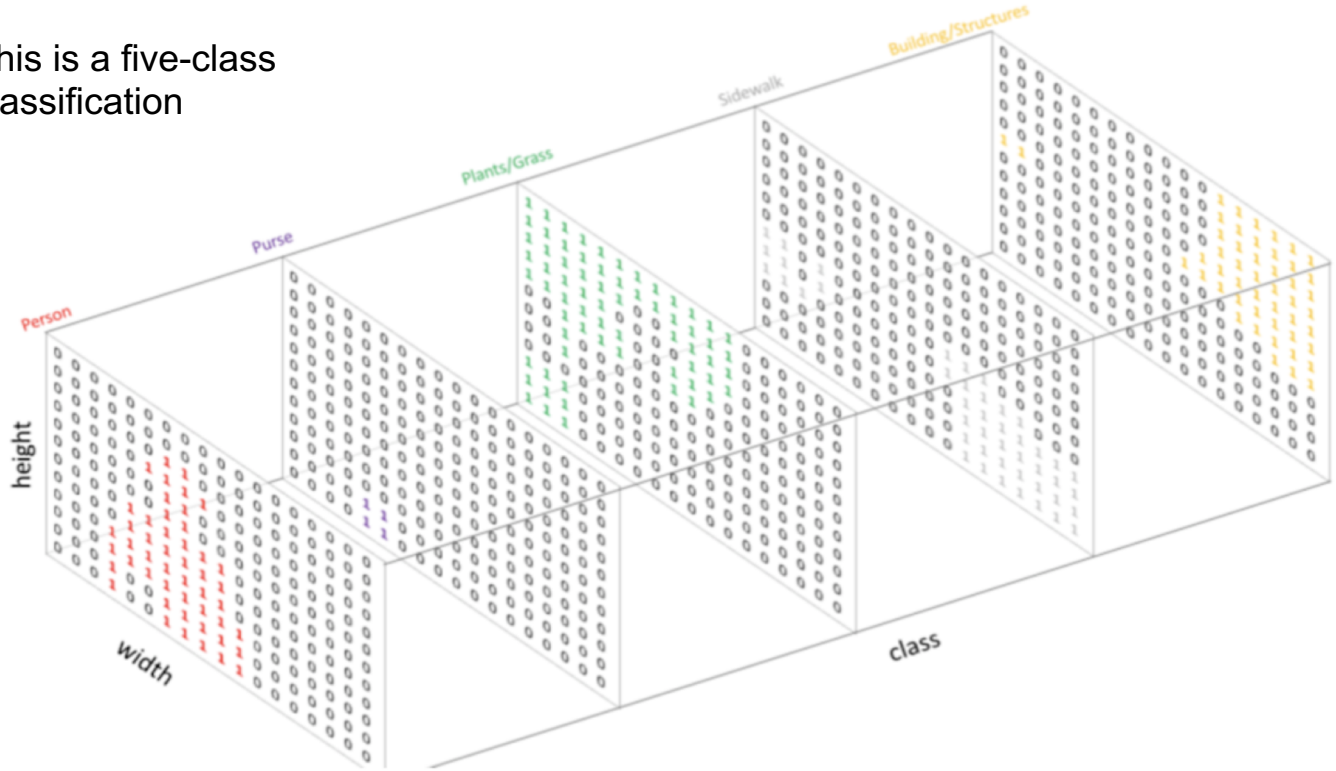
© 2020 DIVE INTO CODE Corp.

Semantic segmentation

The final output feature map has channels for the number of classes.

Using softmax, output the probabilities so that they sum to 1 toward the channel direction for each pixel

This is a five-class classification

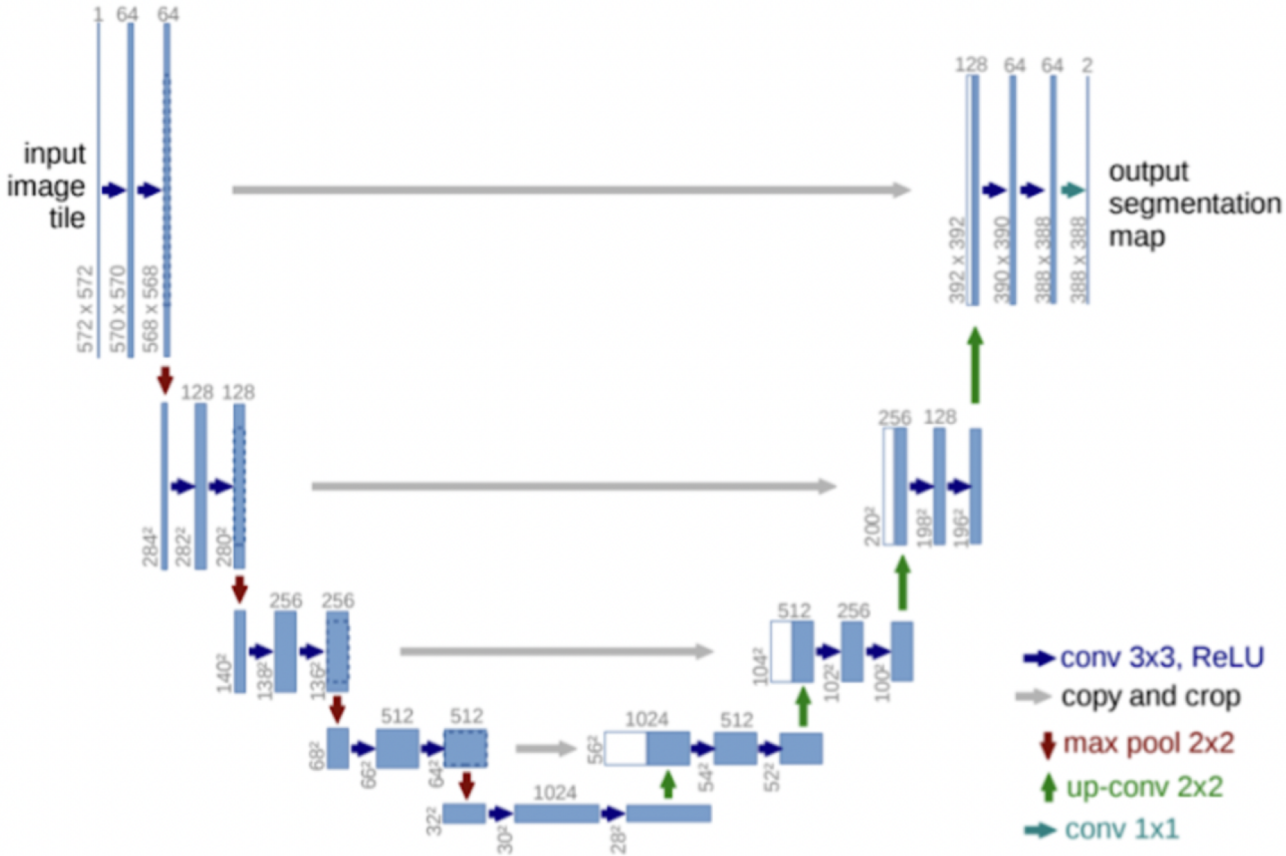


© 2020 DIVE INTO CODE Corp.

Semantic segmentation

U-Net

Read the paper and do some code reading!



© 2020 DIVE INTO CODE Corp.

Sample code

How to solve problems

"U-Net"

[Problem 1] Learning and estimation
[Problem 2] Code reading

© 2020 DIVE INTO CODE Corp.

Sprint 19 – U-net

Explanation about this Sprint is given but please try it on your own first.

Sprint 19 – U-net

Please work on your own after class and submit your assignments on DIVER.

© 2020 DIVE INTO CODE Corp.

Sprint 19 – U-net

A Sample Code of this Sprint is given but please try it on your own.

Sprint 19 – U-net

Please work on your own after class and submit your assignments on DIVER.

© 2020 DIVE INTO CODE Corp.

ToDo by next class

Next class will be Zoom : Thursday August 26, 2021 11:00 ~ 12:00

🚨ToDo: ResNet and VGG
https://diver.diveintocode.jp/curriculums/1963

© 2020 DIVE INTO CODE Corp.

Check-out

3 minutes Please post the following point to Zoom chat.

Q. Current feelings and reflections
(joy, anger, sorrow, anticipation, nervousness, etc.)

© 2020 DIVE INTO CODE Corp.

Thank You For Your Attention

© 2020 DIVE INTO CODE Corp.