Image Processing III: Image segmentation

Research Methods in Data Science – 7PAM2015

Dr Vanessa Graber (based on slides by Dr Gülay Gürkan)

Summary Lecture 2

- Object detection is the field of image processing that is concerned with **creating bounding boxes** around objects and labelling them.
- It is a key research area in computer vision that has numerous applications across many sectors. This is a key skill to have!
- Classical ML is outperformed by DL methods, which can be separa-ted into **two-stage** (R-CNN family) and **single-stage** (YOLO family) **frameworks**. Both have advantages depending on the application.
- Inspired by our brains, **self-attention** (a key concept in transformers) captures contextual relationships between elements in a sequence.
- **Vision transformers** apply self-attention in the context of computer vision by learning semantic relationships between different patches. They are particularly powerful when pretrained on large datasets.

Learning outcomes
After this lecture and the tutorial, you will:

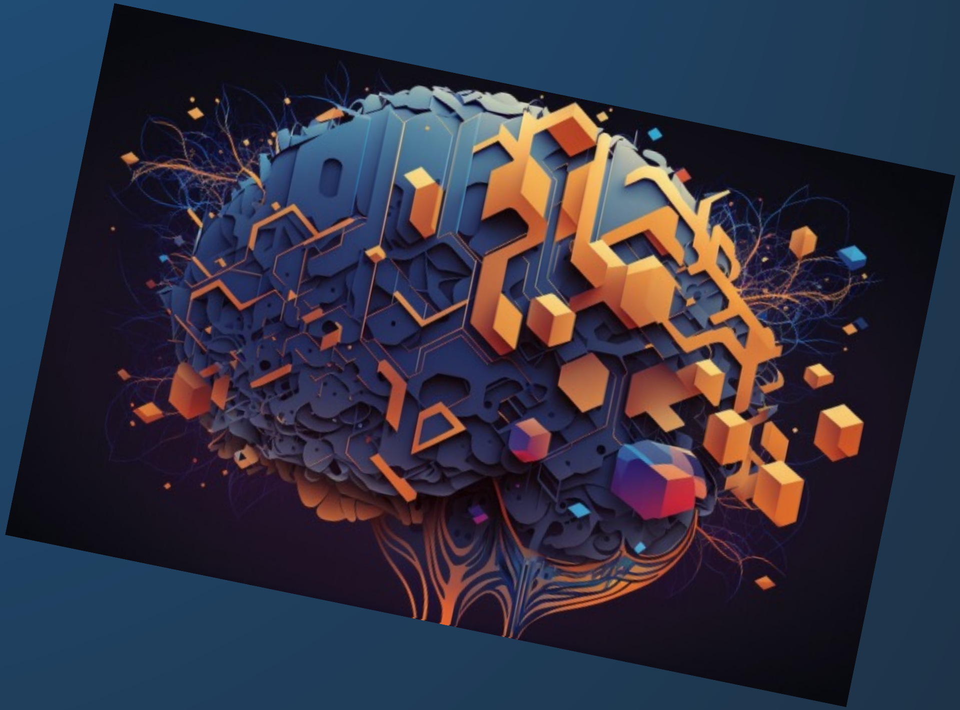
- Understand the main concepts of image segmentation and get to know several applications.
- Understand the key differences between image classification, object detection and image segmentation.
- Have seen various techniques for this type of image processing.
- Have set the foundations to address your first assignment.

Segmentation overview

Segmentation methods

Assignment brief

Summary



Segmentation overview

Segmentation methods

Assignment brief

Summary

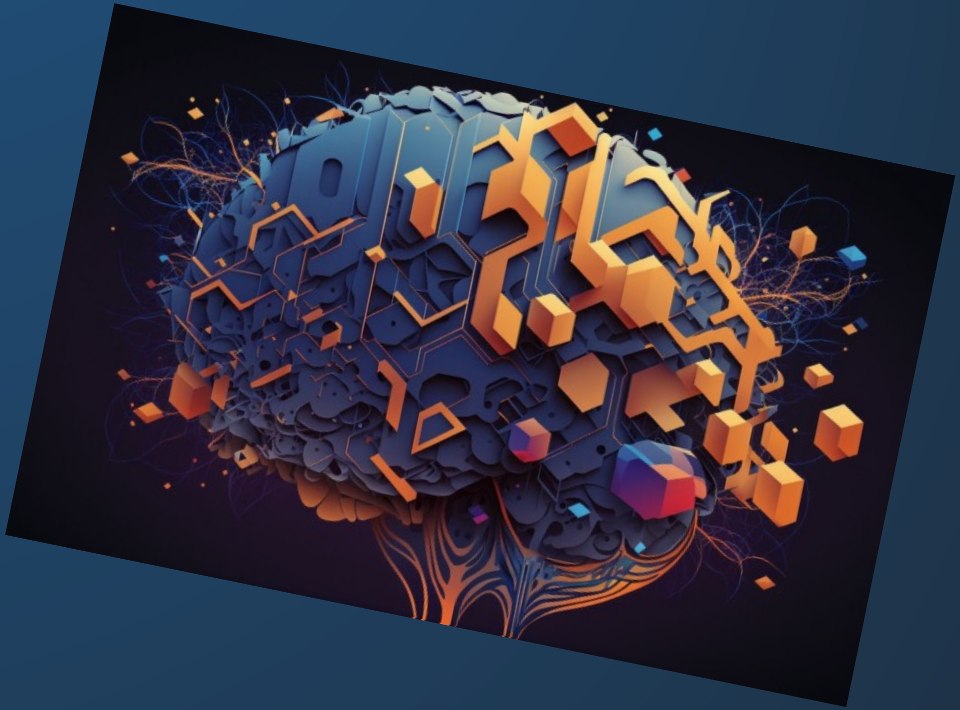


Image recognition
Overview

- Image recognition is a subfield of computer vision that teaches machines how to understand and "see" visual data.

- We typically distinguish the following tasks:
  - Image classification
  - Object detection
  - Image segmentation

In this lecture, we will focus on image segmentation.

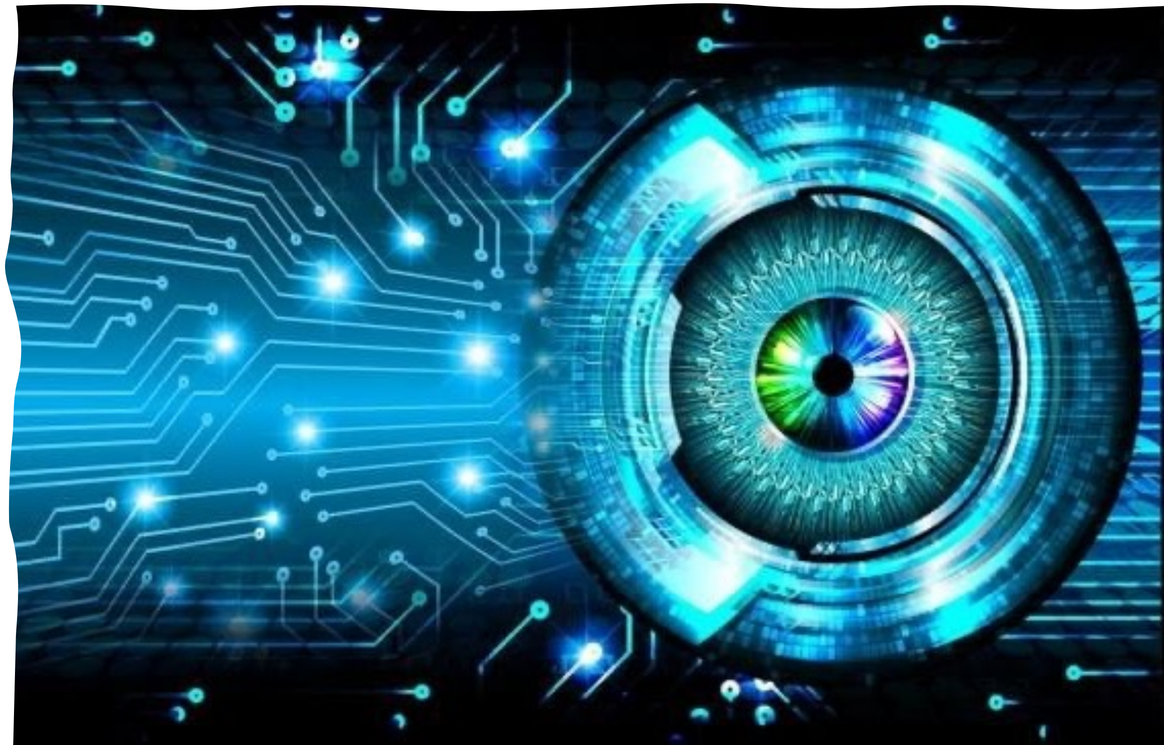


Image segmentation

The central idea

- Image segmentation is the field of computer vision that deals with breaking up an image into several regions (i.e., segments) to reduce the image complexity and allow further processing.

The key idea is to simplify and/or change the representation of an image on a pixel-by-pixel basis into something more meaningful.

Credit: B.Palac

Image segmentation

The central idea

- Image segmentation is the field of computer vision that deals with breaking up an image into several regions (i.e., segments) to reduce the image complexity and allow further processing.

The key idea is to simplify and/or change the representation of an image on a pixel-by-pixel basis into something more meaningful.

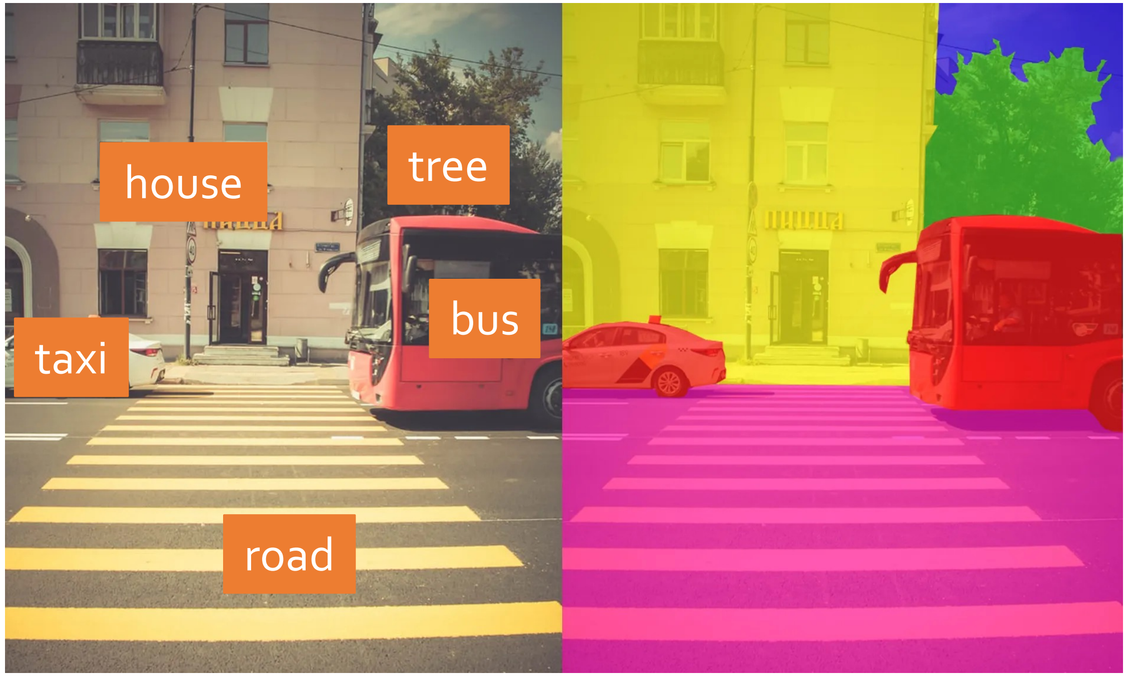
Credit: B.Palac

Image segmentation

The central idea

- Image segmentation is the field of computer vision that deals with breaking up an image into several regions (i.e., segments) to reduce the image complexity and allow further processing.

The key idea is to simplify and/or change the representation of an image on a pixel-by-pixel basis into something more meaningful.



Credit: B.Palac

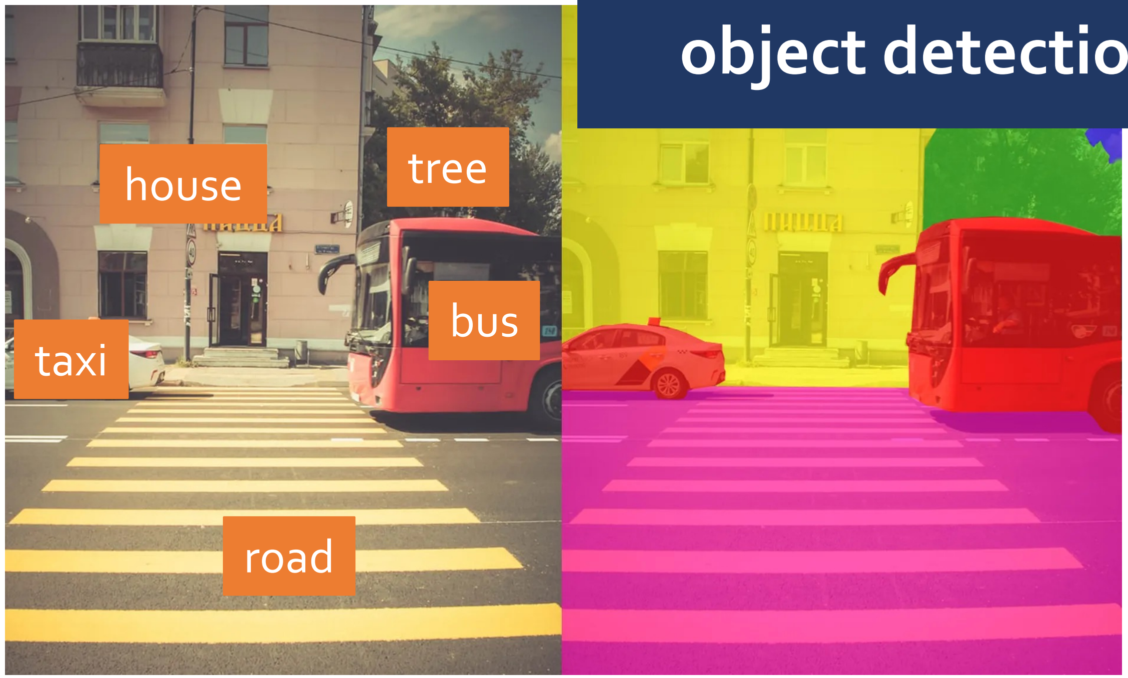
Image segmentation

The central idea

- Image segmentation is the field of computer vision that deals with breaking up an image into several regions to reduce the image complexity and allow furth

This is more complex than image classification and object detection.

The key idea is to simplify and/or change the representation of an image on a pixel-by-pixel basis into something more meaningful.



Credit: B.Palac

Image segmentation

Comparison to object detection

- While **object detection** is focused on defining the location of individual objects with **bounding boxes**, **segmentation** is a pixel-by-pixel based approach that seeks **segmentation masks**.

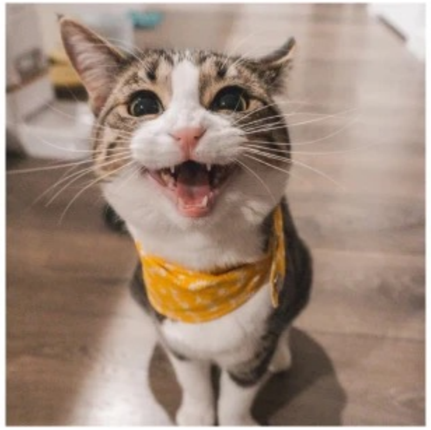
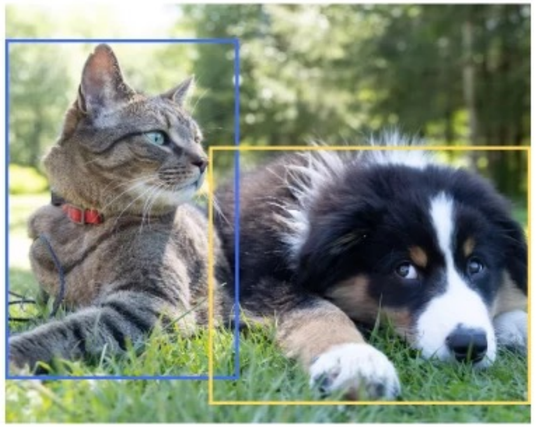


Image Classification



Object Detection

Image Segmentation

Credit: https://www.superannotate.com/blog/image-segmentation-for-machine-learning

Image segmentation
Comparison to object detection

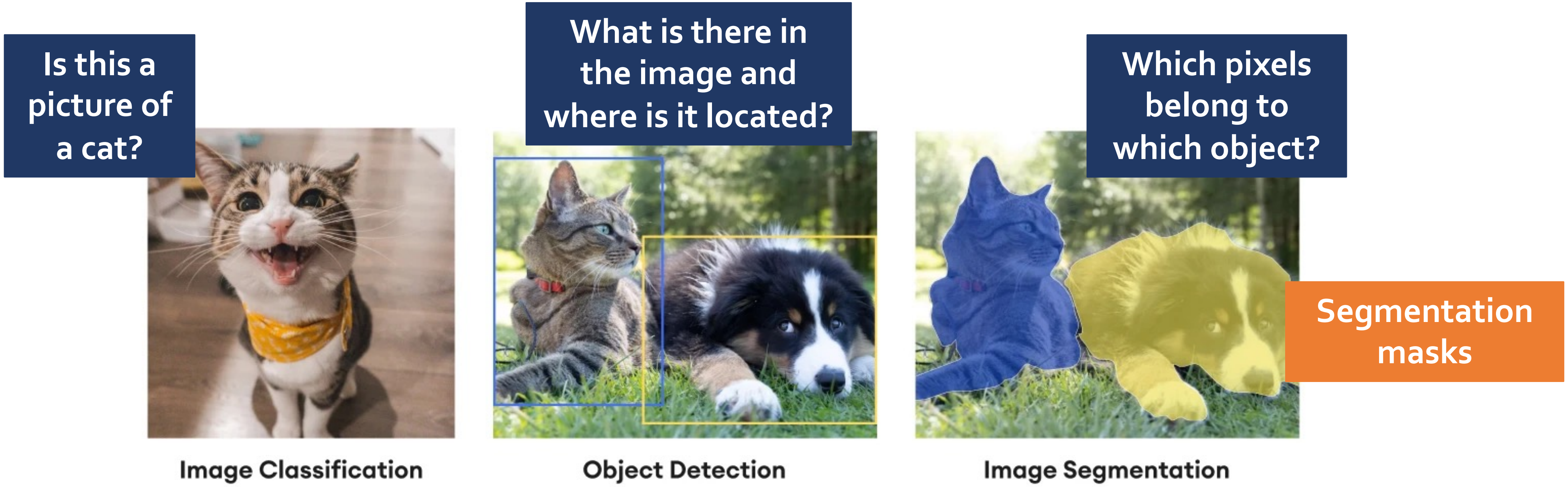
- While **object detection** is focused on defining the location of individual objects with **bounding boxes**, **segmentation** is a pixel-by-pixel based approach that seeks **segmentation masks**.

Credit: https://www.superannotate.com/blog/image-segmentation-for-machine-learning

Image segmentation

Comparison to object detection

- While **object detection** is focused on defining the location of individual objects with **bounding boxes**, **segmentation** is a pixel-by-pixel based approach that seeks **segmentation masks**.



Credit: https://www.superannotate.com/blog/image-segmentation-for-machine-learning

Image segmentation
Three types

- We can distinguish **three types of image** segmentation:
  - Semantic segmentation
  - Instance segmentation
  - Panoptic segmentation

(a) Image

(b) Semantic Segmentation

(c) Instance Segmentation

(d) Panoptic Segmentation

Credit: https://www.v7labs.com/blog/image-segmentation-guide

Image segmentation
Semantic segmentation

- The word “semantic” refers to the meaning of language or logic.
- We classify image pixels into semantic classes so that every pixel belongs to a certain class. The segmentation model does not refer to any additional information or context.

Image segmentation
Semantic segmentation

- The word “semantic” refers to the meaning of language or logic.
- We classify image pixels into **semantic classes** so that every pixel belongs to a certain class. The segmentation model does not refer to any additional information or context.

We do not get information about the type of vehicle or type of building in the image.

Semantic segmentation does not provide in-depth information. This can be a problem for complex images.

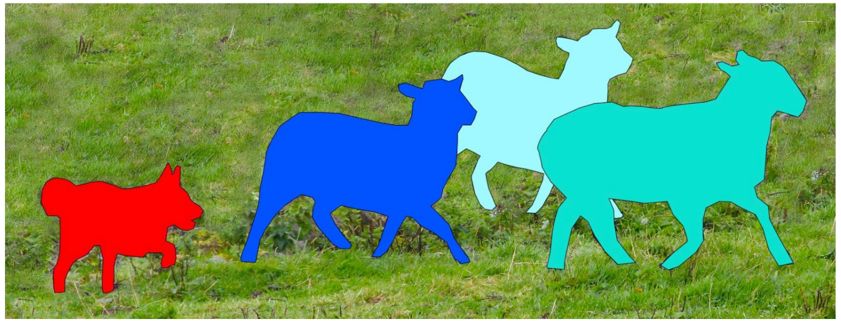
Credit: B.Palac

Image segmentation

Instance segmentation

- Here, we classify image pixels based on the instances of an object. In contrast to semantic segmentation, we do not use object classes.
- Instance segmentation does not know the class of a classified region, so it can only segregate similar or overlapping regions based on object boundaries.

This approach allows us to not only distinguish different animals (e.g., dog vs sheep) but also identify different instances within the same "class" (e.g., three different sheep).



Credit: Russ Tedrake

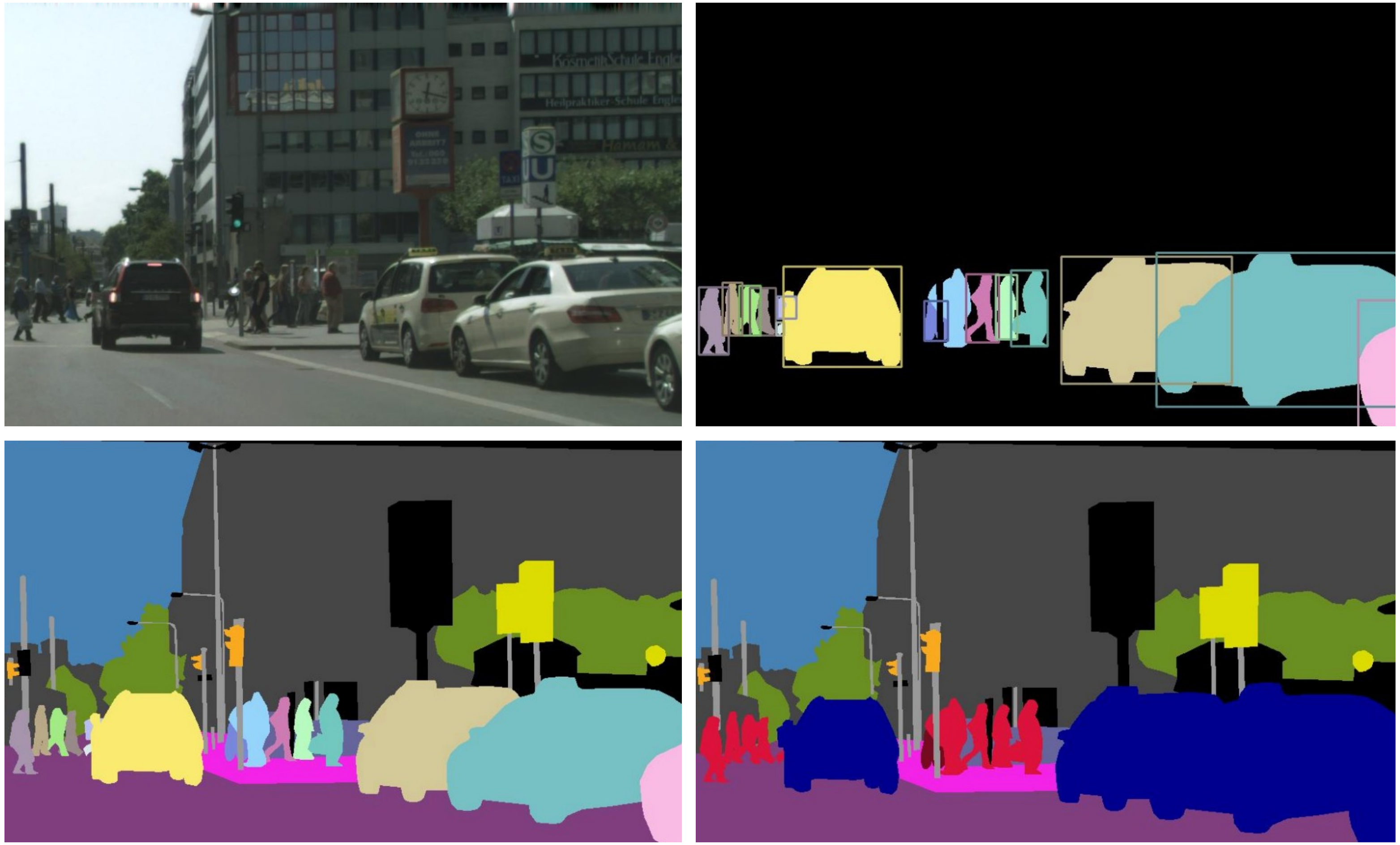
Image segmentation
Panoptic segmentation

- “Panoptic” means to give a comprehensive or panoramic view.
- Panoptic segmentation combines semantic and instance segmentation, so that each instance of an object in an image is segmented and its identity predicted.

Credit: https://medium.com/@prakhar.bansal/panoptic-segmentation-explained-5fa7313591a3

Image segmentation
Another example

Original image



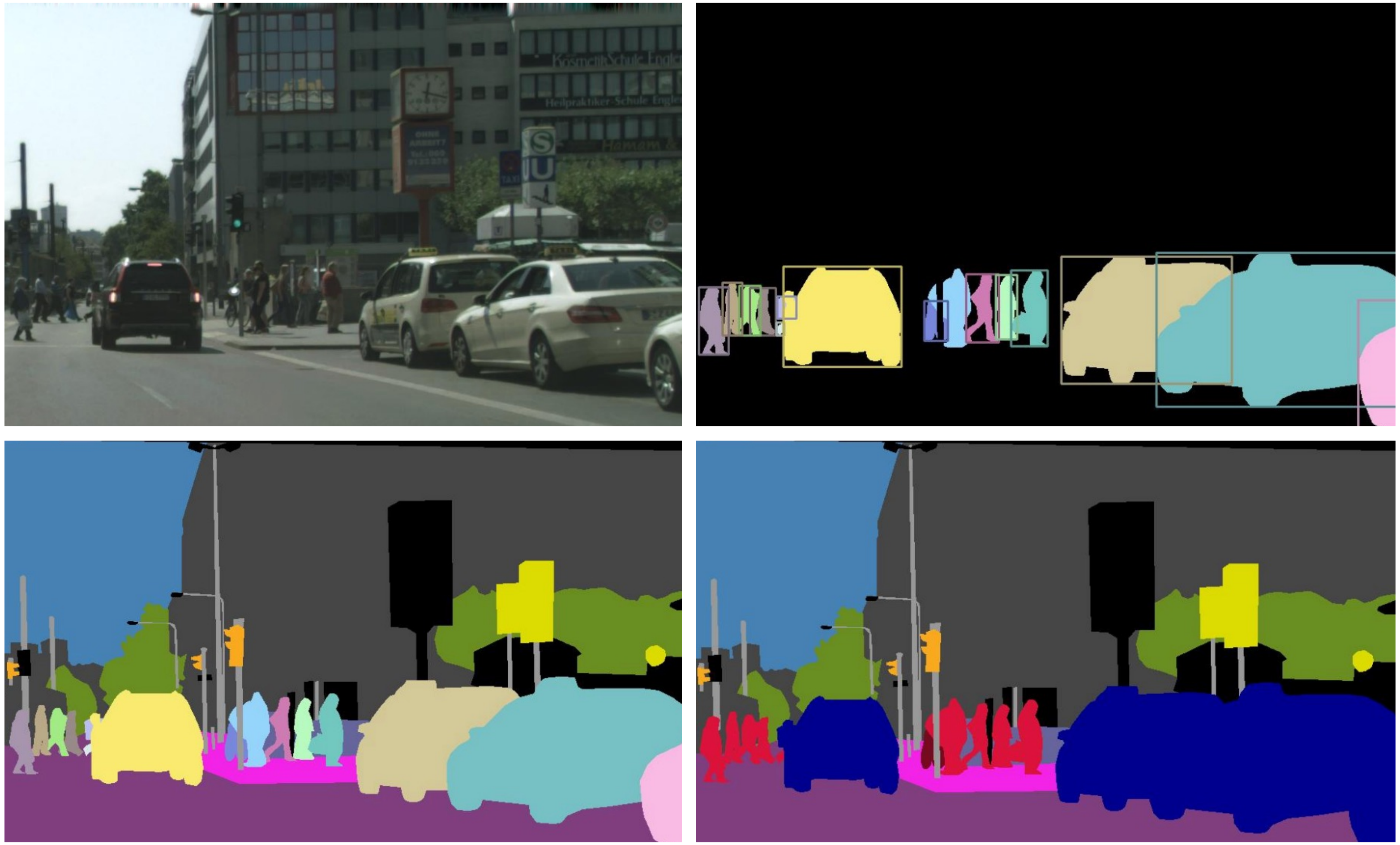
Credit: Kirillov et al. (2019)

Image segmentation

Another example

Original image

Instance segmentation



Panoptic segmentation

Semantic segmentation

Credit: Kirillov et al. (2019)

Image segmentation
Applications

- Image segmentation has a **wide range of applications**, including
  - **Transportation** (self-driving cars, e.g., lane segmentation and pedestrian tracking; traffic management, e.g., number plate identification)
  - **Agriculture** (locating objects in satellite images, e.g., tracking animals, estimating crop yields, determining weed spread)
  - **Medical imaging** (disease diagnosis, e.g., precisely locating cancer cells or tumours, brain segmentation)
  - **Gaming** (e.g., creating virtual realities)
  - **Security** (face recognition, e.g., phone security, tracking)

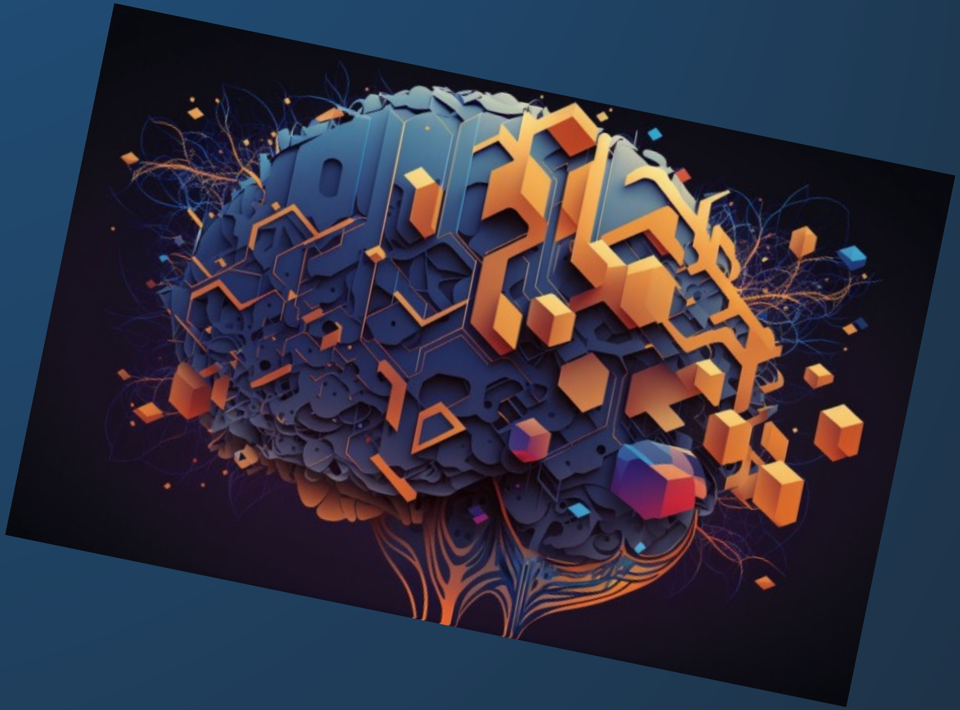
Credit: Li et al. (2023)

Segmentation overview

Segmentation methods

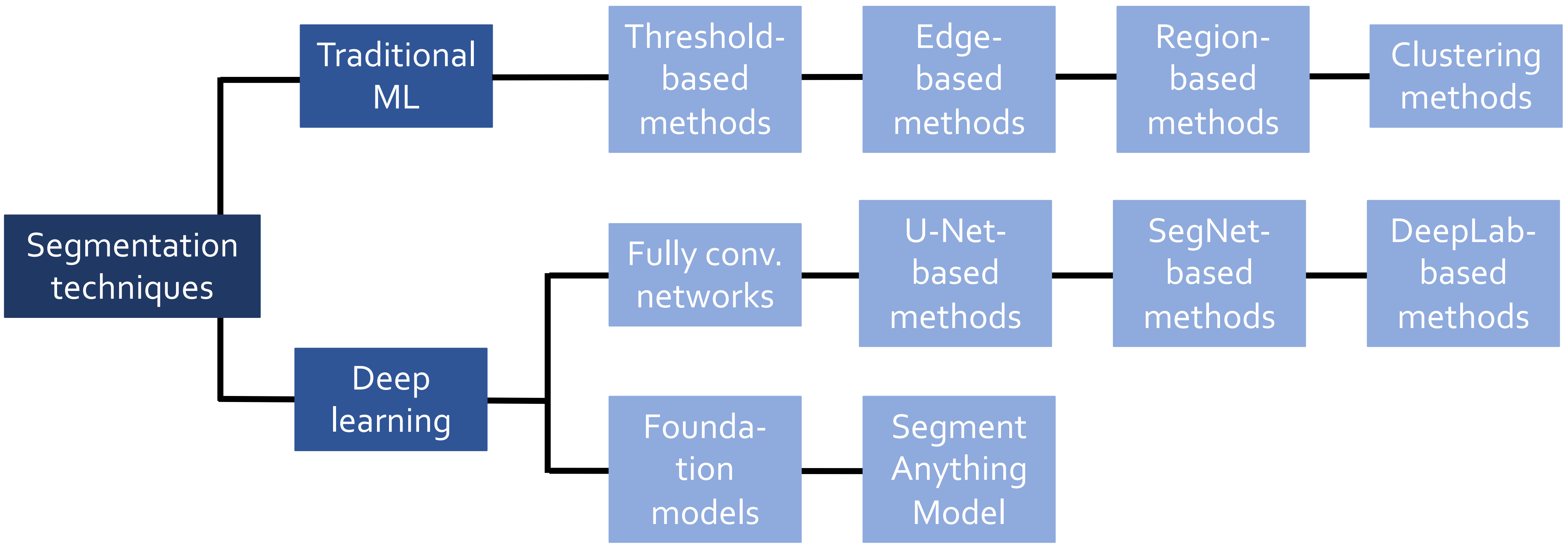
Assignment brief

Summary



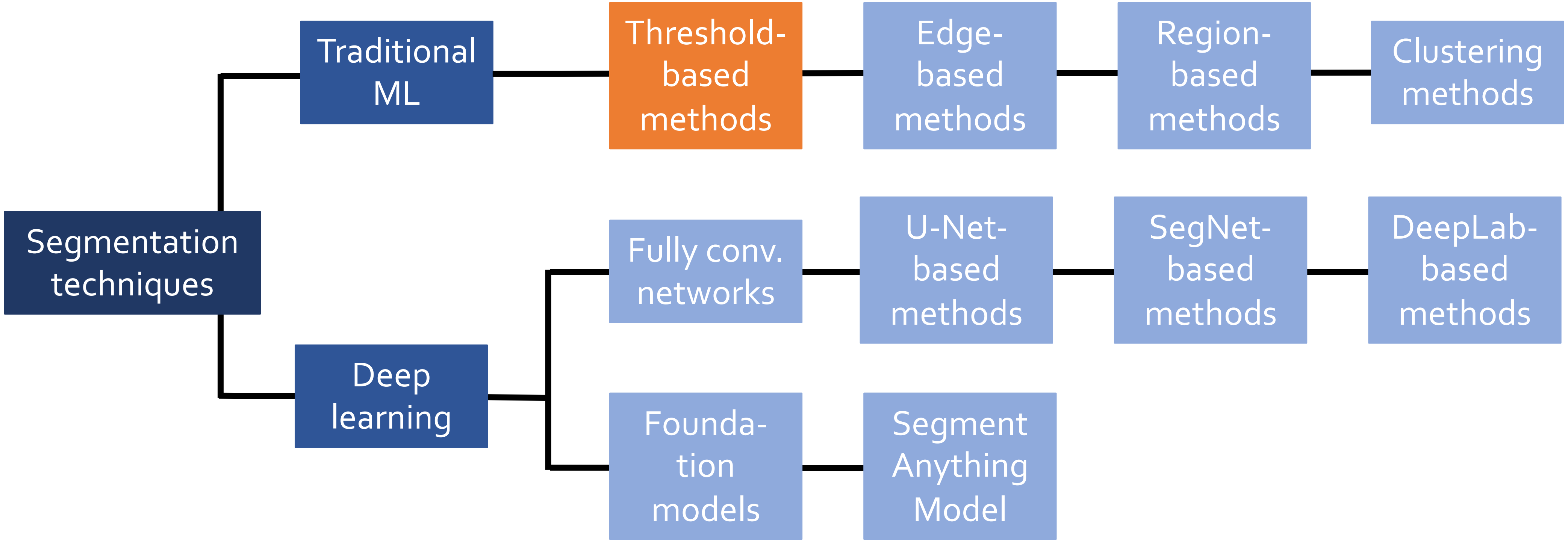
Segmentation methods
Overview of techniques

- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Segmentation methods
Overview of techniques

- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Traditional segmentation methods

Threshold-based methods I

- For this basic method we **set a threshold** to separate pixel values and assign pixels **either a 0 or 1** if the pixel intensity value is below the threshold or above, respectively. This results in a binary map.

- We generally distinguish **two variants**:
    - Global thresholding
    - Adaptive thresholding

Traditional segmentation methods

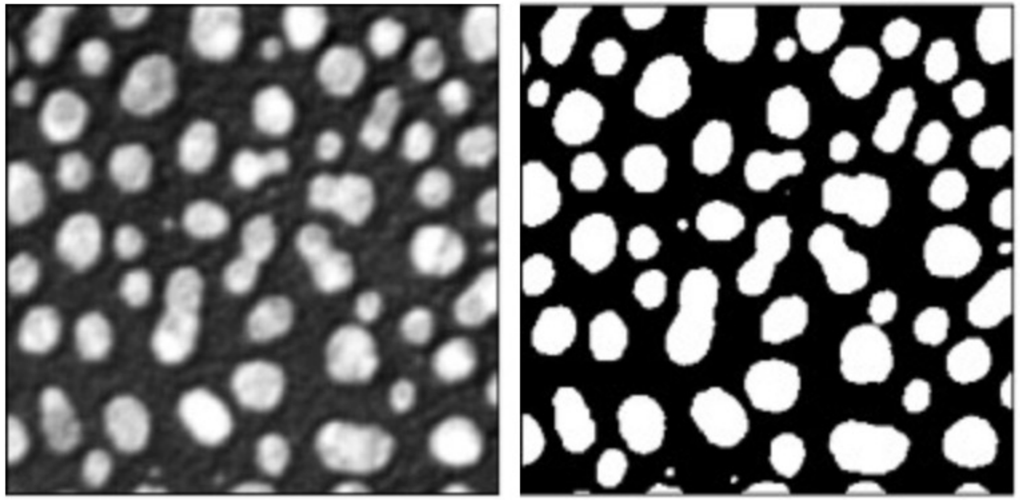
Threshold-based methods I

- For this basic method we **set a threshold** to separate pixel values and assign pixels **either a 0 or 1** if the pixel intensity value is below the threshold or above, respectively. This results in a **binary map**.

- We generally distinguish **two variants**:
  - Global thresholding
  - Adaptive thresholding

In global thresholding, we separate foreground (intensity > threshold) and background (intensity < threshold) based on the pixel value.

Global thresholding segmentation



Traditional segmentation methods

Threshold-based methods I

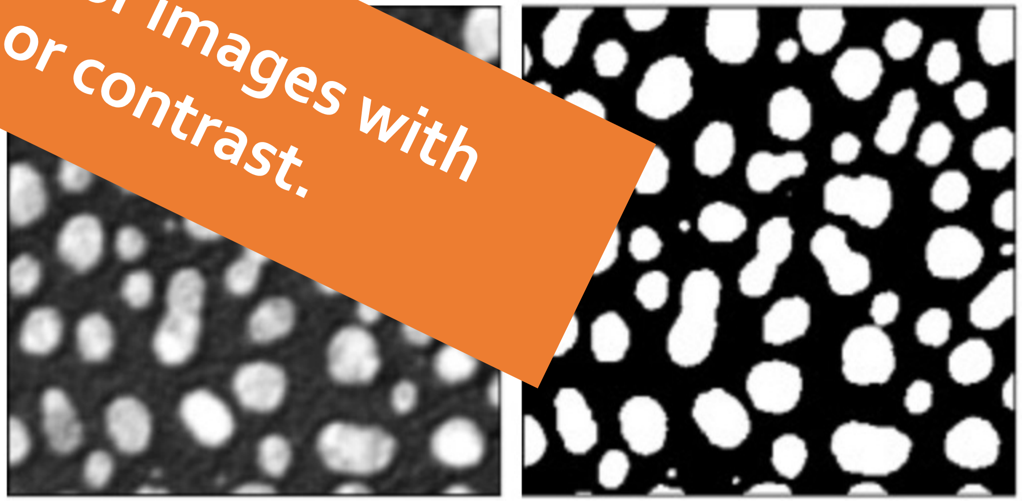
- For this basic ... d we **set a threshold** to separate pixel values and assign pixels ... r 1 if the pixel intensity value is below the threshold o ... ly. This results in a binary map.

- We generally disting...
  - Global thresholding
  - Adaptive thresholding

This approach is not ideal for images with varying illumination or contrast.

In global thresholding, we separate foreground (intensity > threshold) and background (intensity < threshold) based on the pixel value.

Global thresholding segmentation

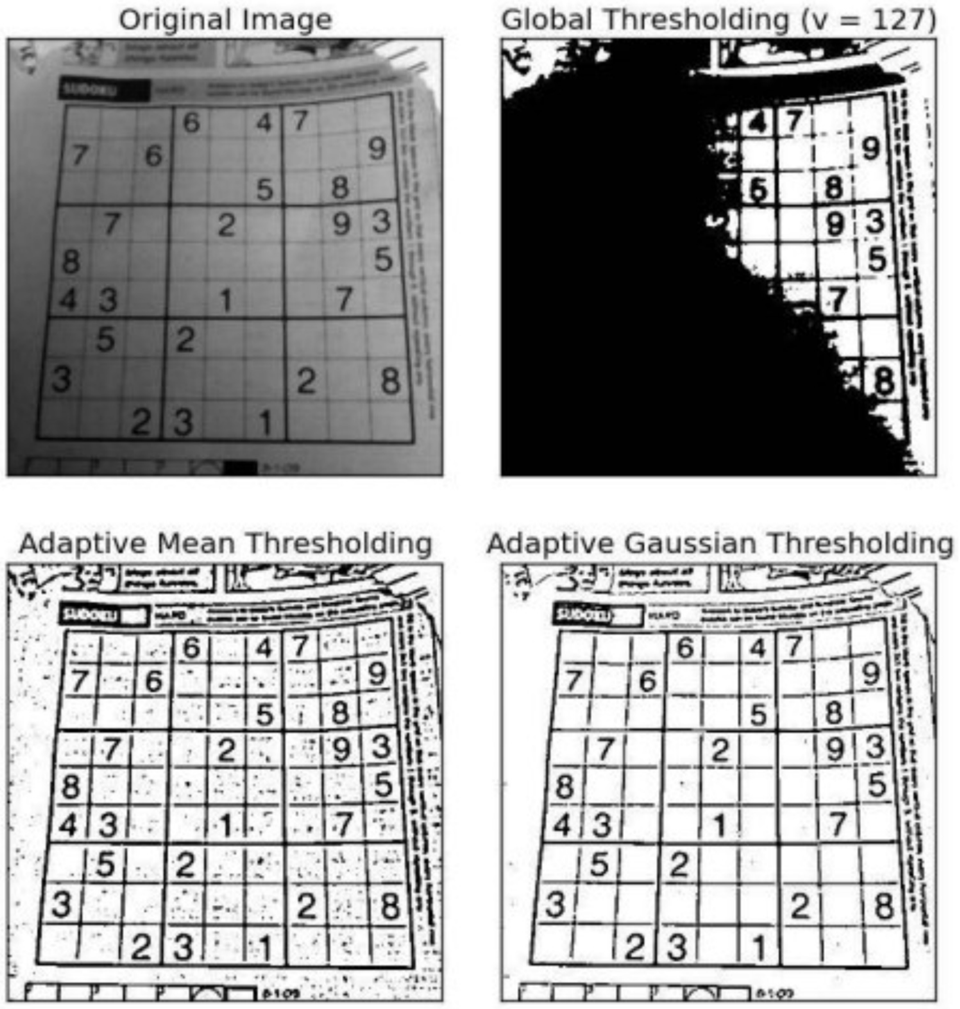


Traditional segmentation methods

Threshold-based methods II

- To overcome this issue, in adaptive thresholding, we separate foreground and background regions by adjusting the threshold value locally based on certain image characteristics.

Here, the threshold value is determined for the subregions based on a chosen statistic (e.g., the local mean or 2D Gaussian-weighted average) of pixel intensities within the region.

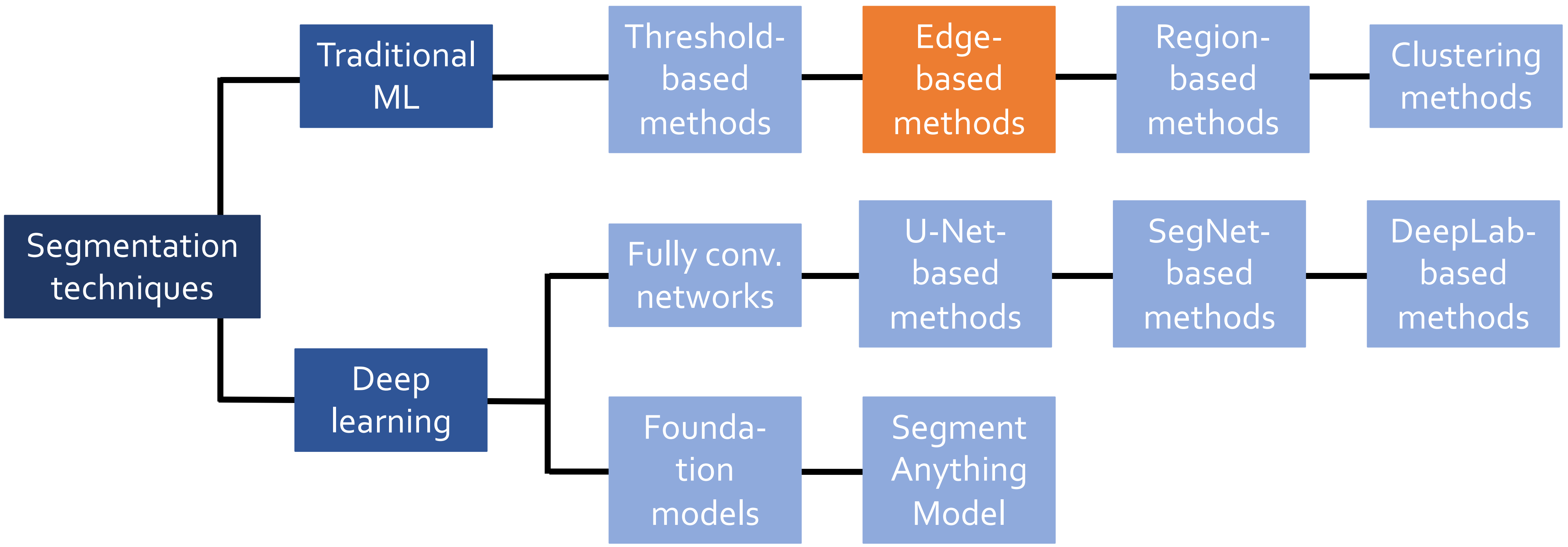


Credit: https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html

Segmentation methods

Overview of techniques

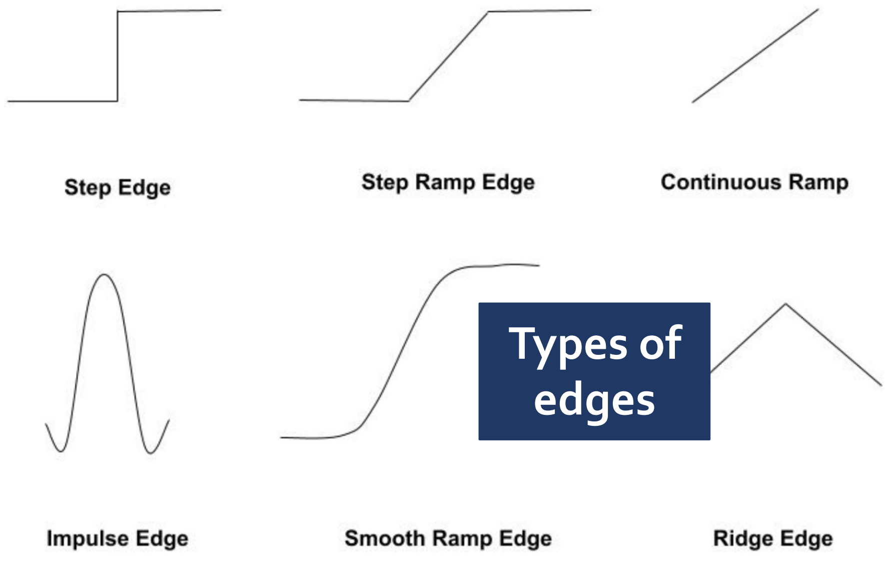
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Traditional segmentation methods

Edge-based methods

- These methods are based on existing **edge-detection operators** (e.g., Kirsch, Sobel, Laplace, Canny, Roberts, Prewitt, etc.) that look for locations of discontinuity in colours, grey levels, texture, etc. by comparing adjacent regions. The resulting image is then further processed to combine the edge segments into similar areas.



Types of edges

Credit: Mrinal Tyagi

1	1	1
1	-8	1
1	1	1

0	1	0
1	-4	1
0	1	0

Laplace

-1	0	1
-2	0	2
1	0	1

1	2	1
0	0	0
-1	-2	1

Sobel

1	0
0	-1

0	1
-1	0

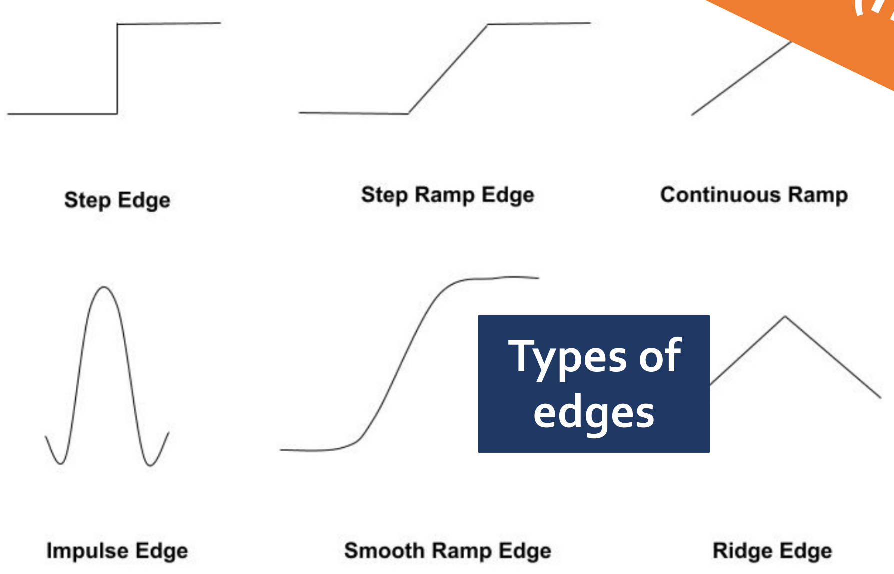
Roberts

Traditional segmentation methods

Edge-based methods

- These methods a[illegible]based on existing **edge-detection operators** (e.g., Kirsch, Sobel, L[illegible]anny, Roberts, Prewitt, etc.) that look for locations of dis[illegible]lours, grey levels, texture, etc. by comparing ad[illegible]sulting image is then further processed to co[illegible]ts into similar areas.

Edge-detection relies on performing convolutional operations of the selected operator (matrix) with our image.



Types of edges

[illegible]	0	1
[illegible]	[illegible]	[illegible]
1	[illegible]	[illegible]

1	0
0	-1

Laplace

0	1	0
1	-4	1
0	1	0

[illegible]

1	2	1
0	0	0
-1	-2	1

Roberts

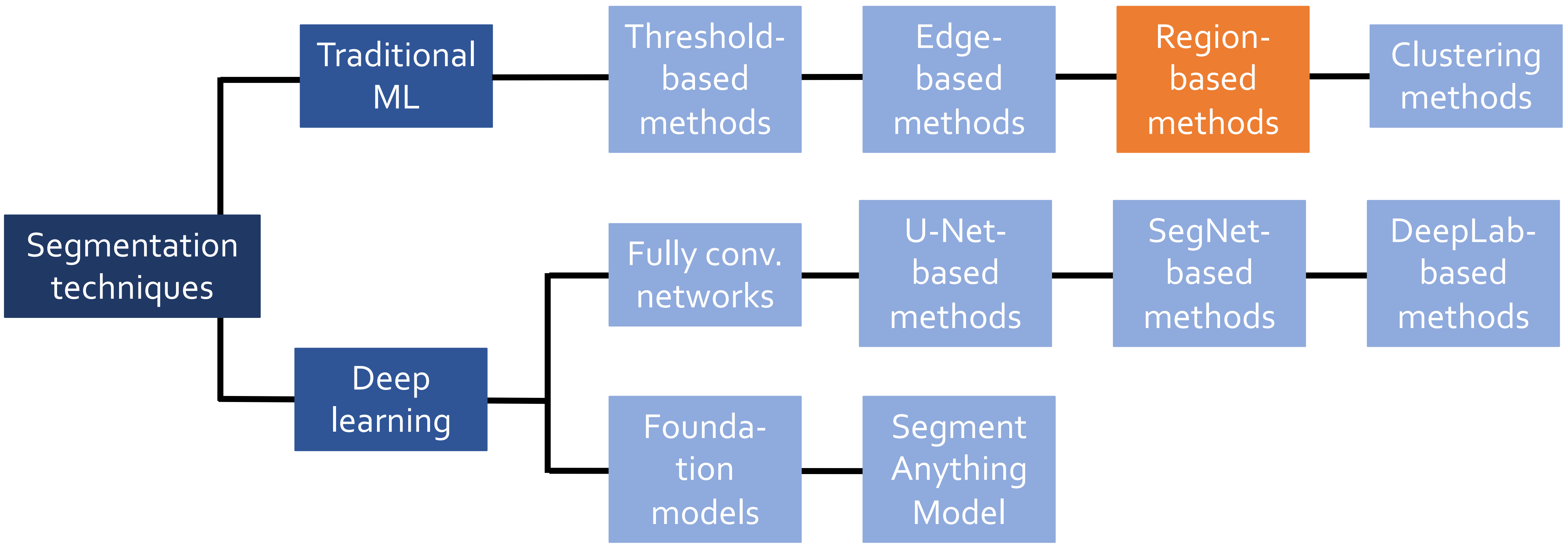
0	1
-1	0

Credit: Mrinal Tyagi

Segmentation methods

Overview of techniques

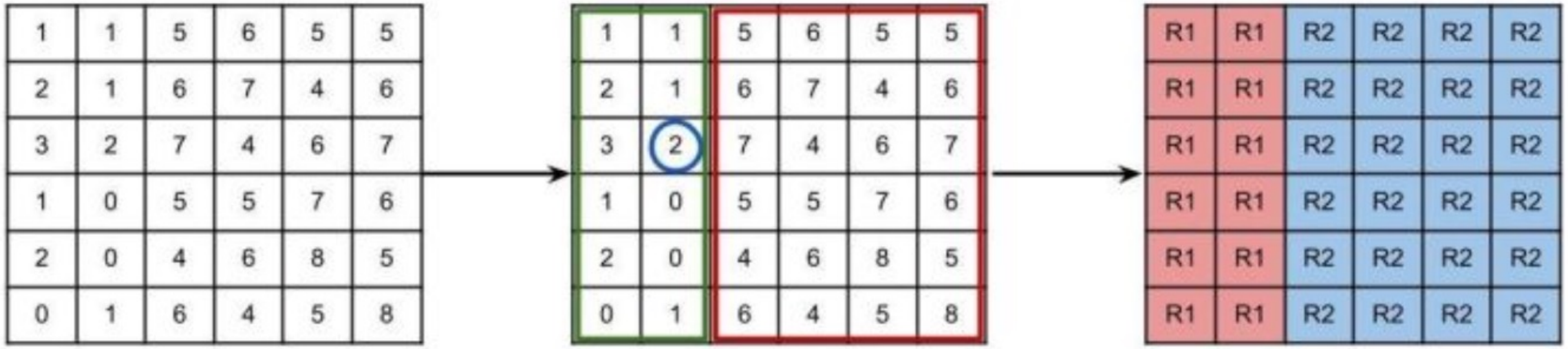
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Traditional segmentation methods
Region-based methods I

- In this case, a certain number N of **pixels with similar properties** (e.g., intensity or colour) are grouped together into a region. Numerous prescriptions exist for determining these regions. For example:

Method 1: Start with a seed pixel and threshold (both are random) and grow the region around the seed based on the chosen threshold.



Original Image

Region growing process with 2 as the seed pixel.

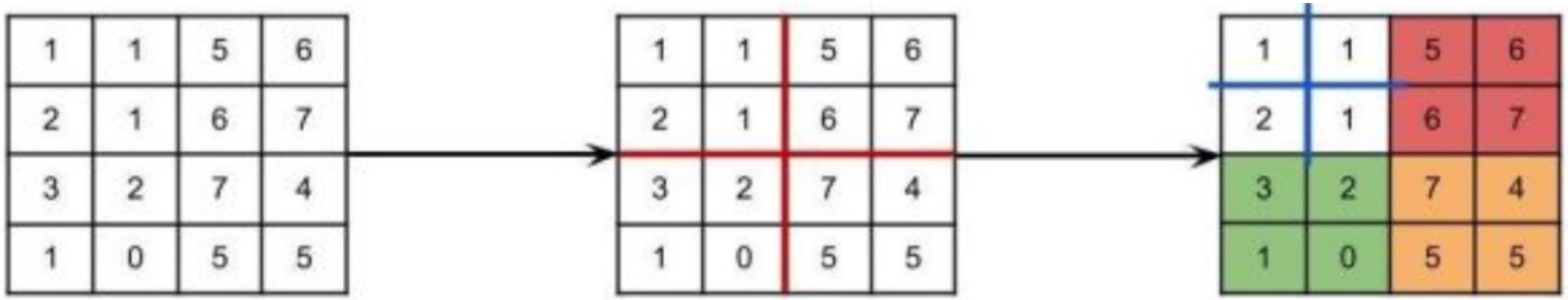
Splitting image into two regions based on a threshold.

Credit: Mrinal Tyagi

Traditional segmentation methods

Region-based methods II

Method 2: Pre-define some rule of what makes a region. Then, take the entire image and check if the rule is satisfied. If not split the image (typically into quadrants) and check the individual subsections. If condition is satisfied, classify as a region, otherwise proceed iteratively with the splitting.



Original Image

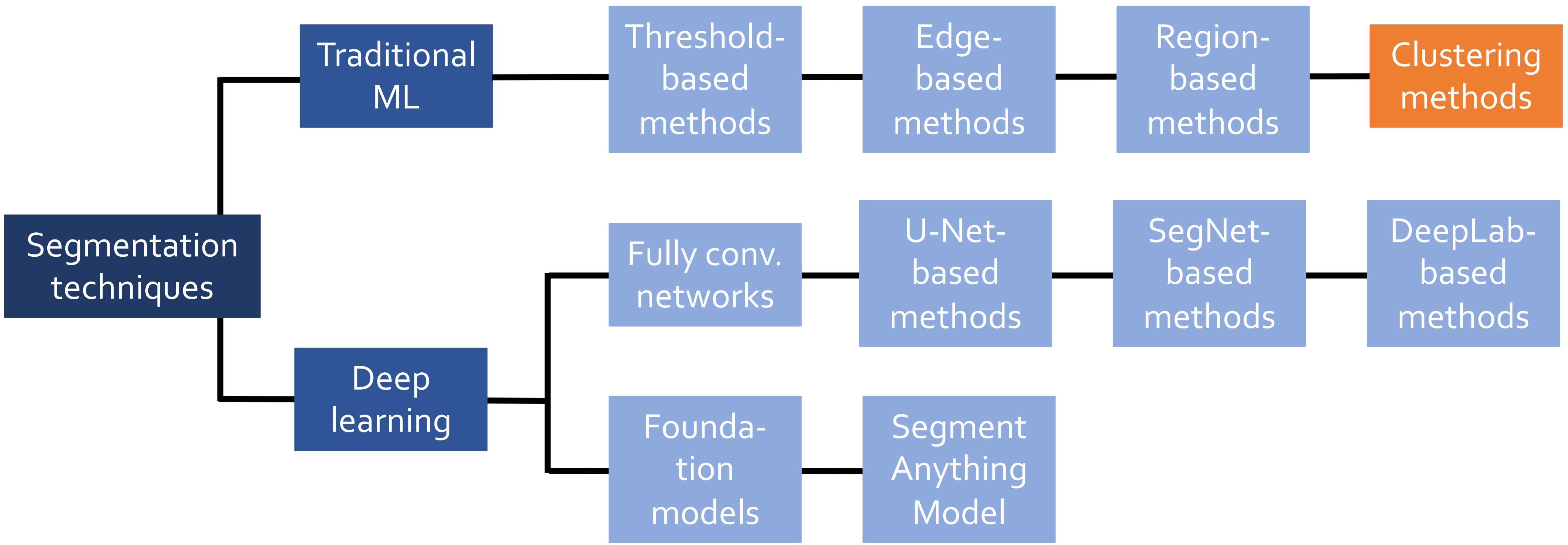
Region splitting into 4 quadrant

Classifying a quadrant as a region if it satisfies condition else performing further splitting

Credit: Mrinal Tyagi

Segmentation methods
Overview of techniques

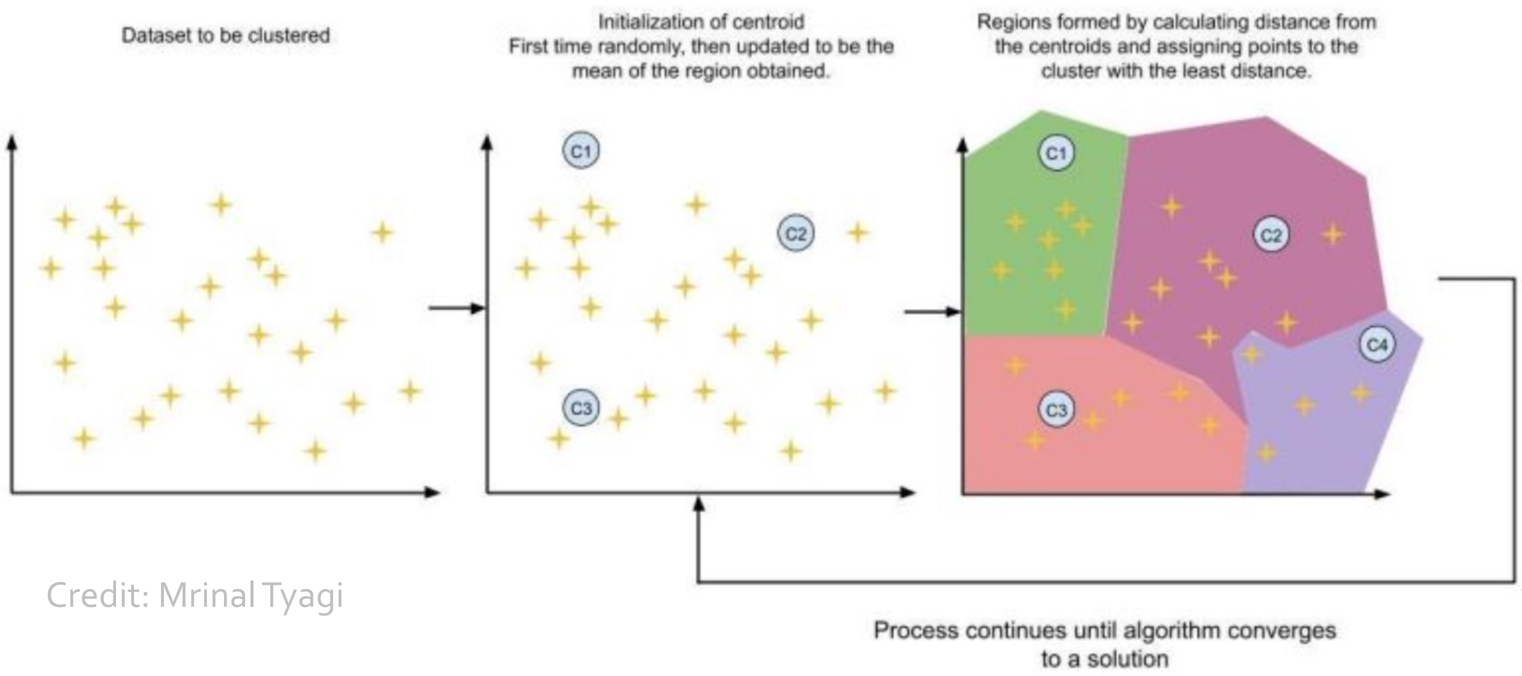
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Traditional segmentation methods

Clustering methods I

- The most common clustering approach in image segmentation is **k-means clustering**. This unsupervised method works as follows:



Credit: Mrinal Tyagi

Step 1: Initialise K random cluster centres and calculate distances to each point.

Step 2: Assign points to cluster centre with the closest mean.

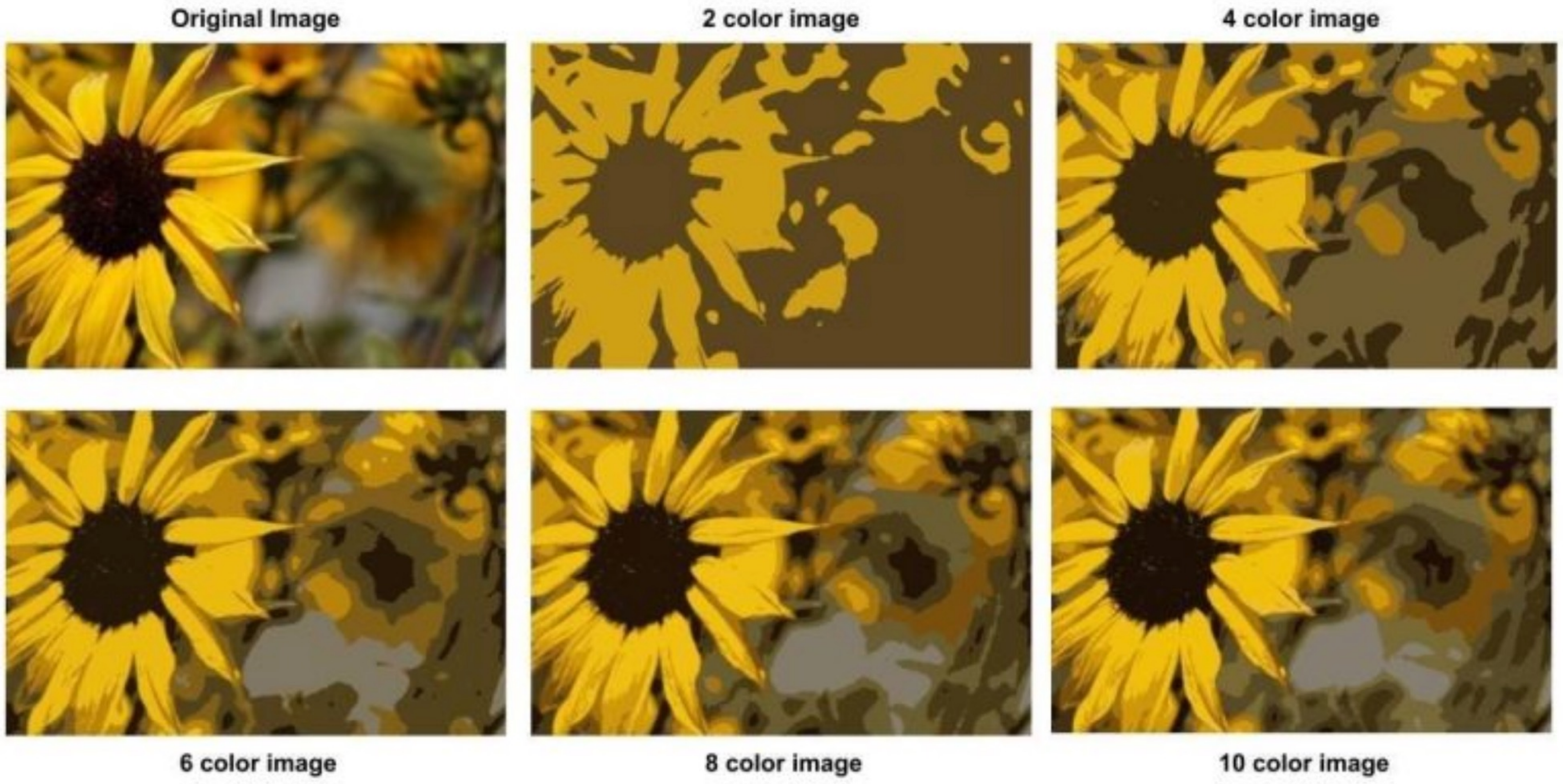
Step 3: For each cluster recalculate the cluster centre by taking mean of all points.

Step 4: Repeat steps 2 and 3 until convergence is reached.

Traditional segmentation methods
Clustering methods II

- The result of such a segmentation approach is shown below:

Individual colours correspond to the number of clusters inputted into the K-means segmentation algorithm. I.e., for a two-colour image, we cluster into two types of regions, and so on.



Credit: Mrinal Tyagi

Question time

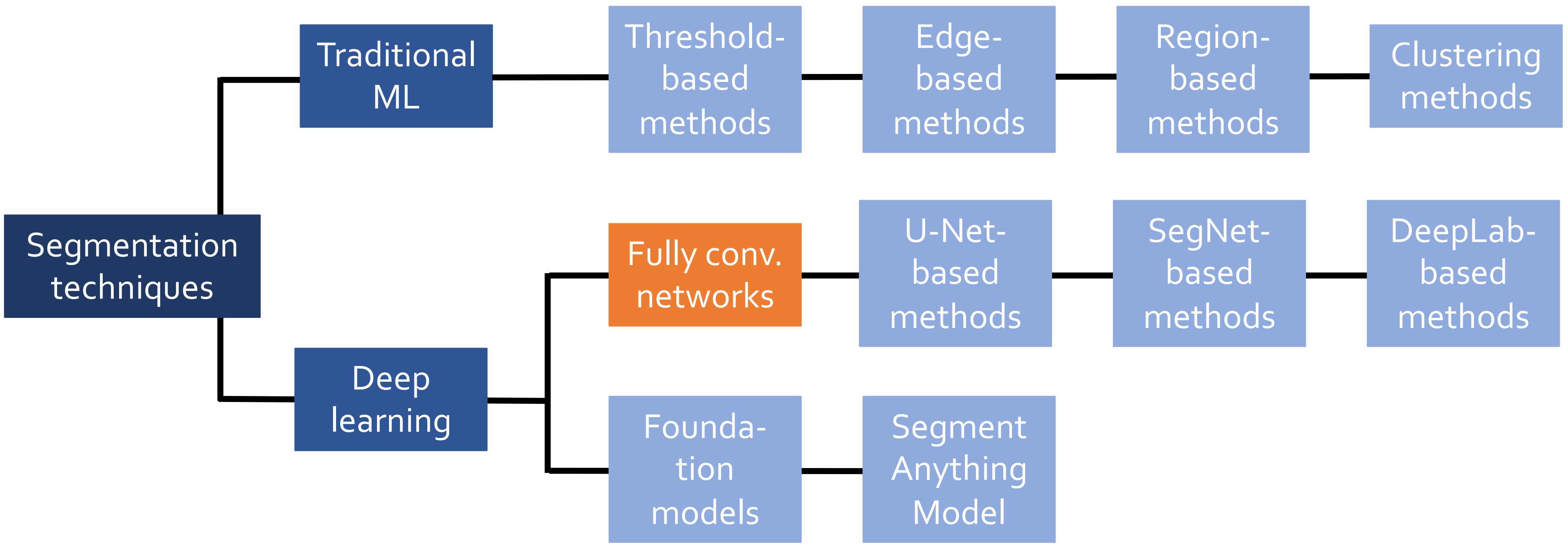
Mentimeter quiz

Let's take a few minutes to recap what we have discussed so far:

Go to menti.com and enter the code 4241 0388.

Segmentation methods
Overview of techniques

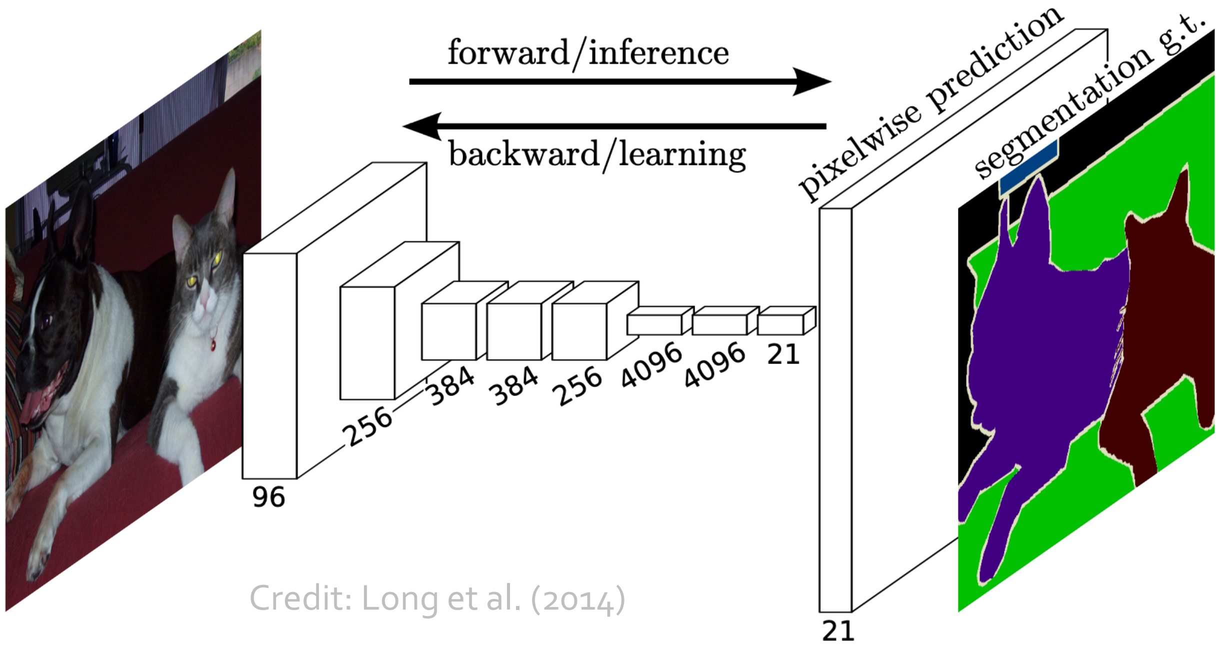
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



DL segmentation methods

Fully convolutional networks I

- The first approaches was developed by Long et al. (2014). Up to that point, CNNs had been used for image classification. Segmentation, however, is more complex as we want to separate object instances from each other.



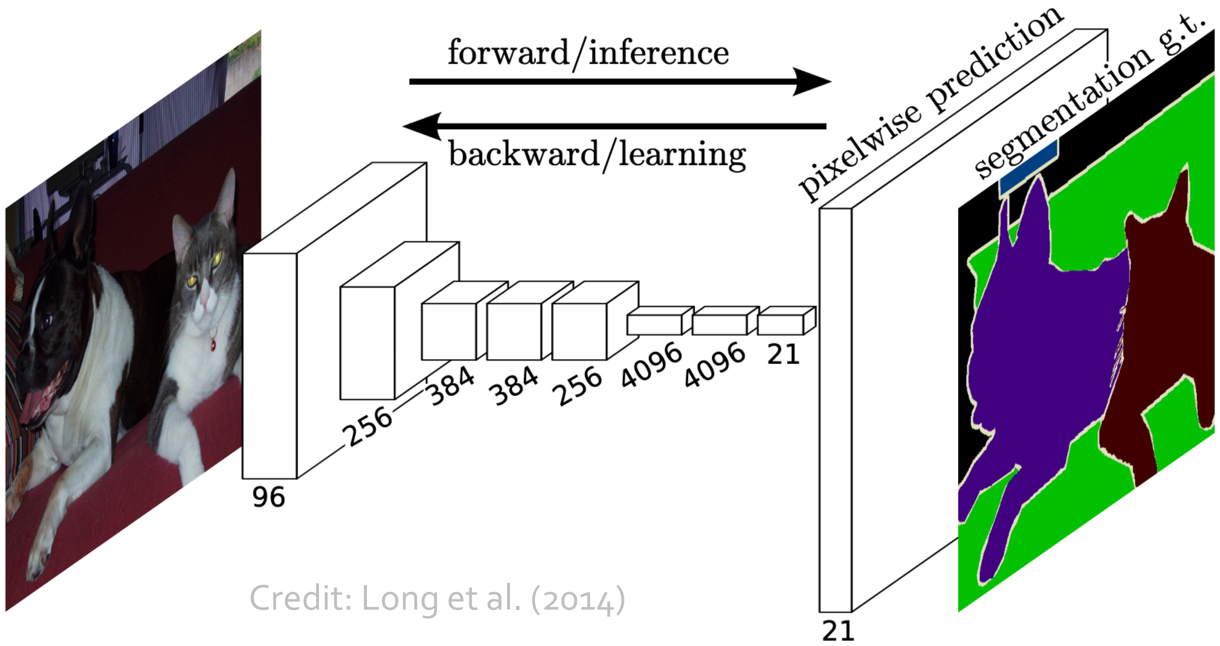
Credit: Long et al. (2014)

Transfer learning is crucial in this area because obtaining segmentation mask labels (ground truths) is very time consuming.

DL segmentation methods

Fully convolutional networks I

- The first approaches was developed by Long et al. (2014). Up to that point, CNNs had been used for image classification. Segmentation, however, is more complex as we want to separate object instances from each other.



Credit: Long et al. (2014)

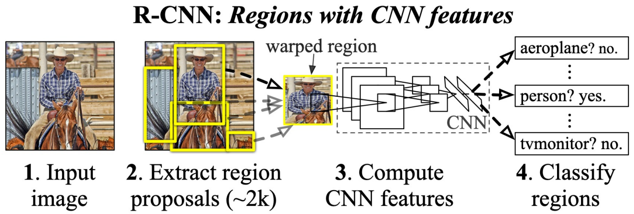
Transfer learning is crucial in this area because obtaining segmentation mask labels (ground truths) is very time consuming.

The problem: Pretrained networks for image classification learn global features. How can we modify NNs to also capture local information?

DL segmentation methods

Fully convolutional networks II

Credit: Girshick et al. (2014)

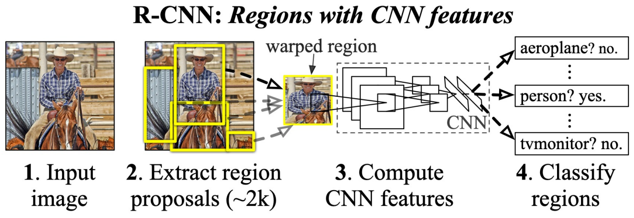


- In Lecture 2 on Object Detection, we saw that people used **pretrained networks on Regions of Interest** (RoIs) to extract features from image parts and output a **class probability** for what we see in that RoI.

DL segmentation methods

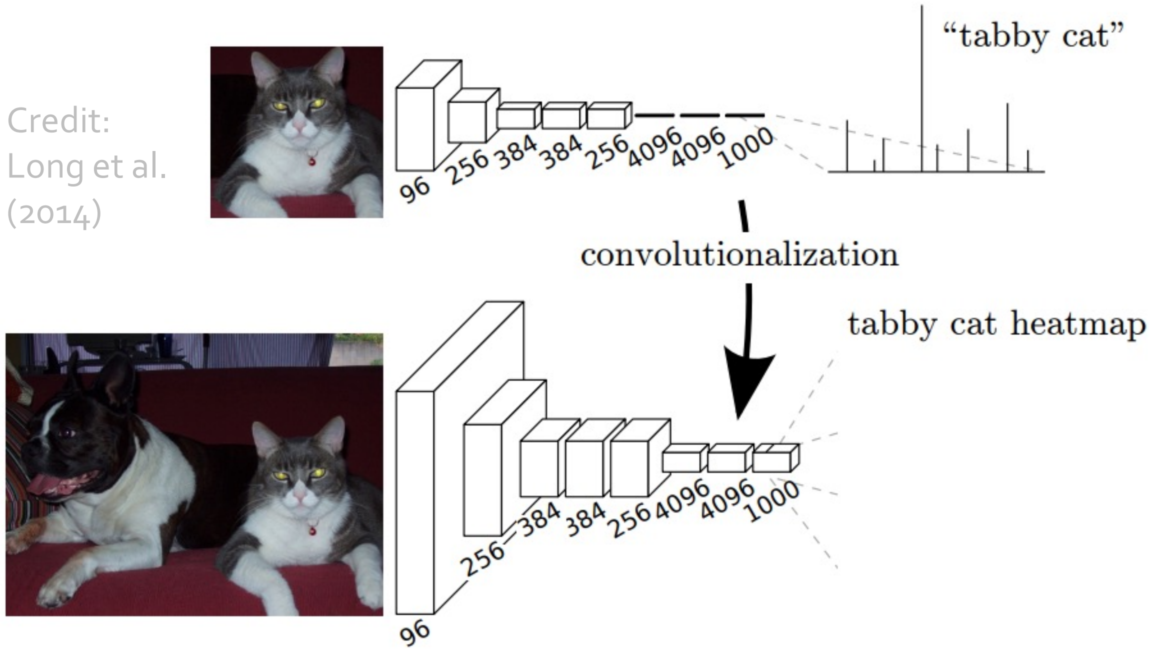
Fully convolutional networks II

Credit: Girshick et al. (2014)



- In Lecture 2 on Object Detection, we saw that people used **pretrained networks on Regions of Interest** (RoIs) to extract features from image parts and output a **class probability** for what we see in that RoI.

Credit: Long et al. (2014)

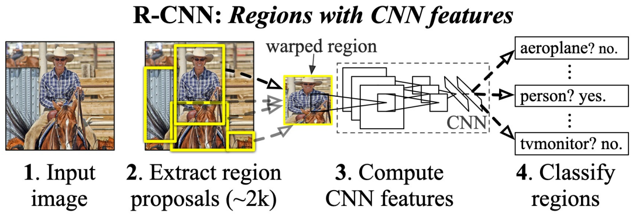


- The key idea of Long et al. (2014) is to **"reinterpret" the final fully connected layers** outputting the class score as convolutions by "reshaping" weights into a matrix.

- As convolutions are resolution independent, we can use a kernel that covers the entire image.

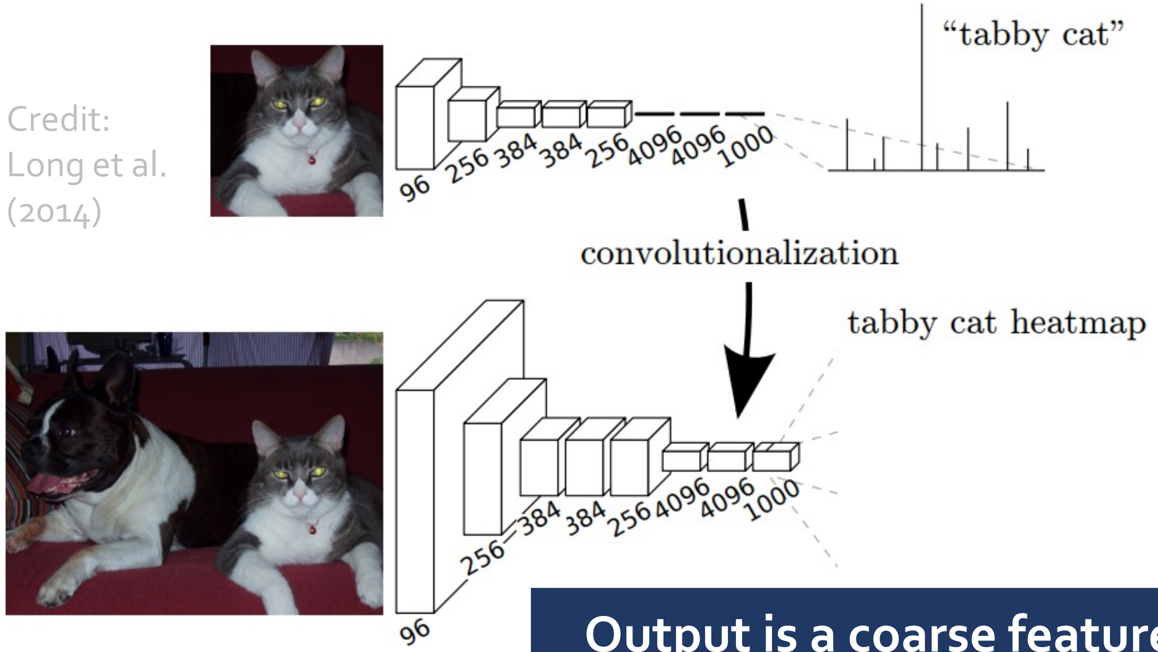
DL segmentation methods
Fully convolutional networks II

Credit: Girshick et al. (2014)



- In Lecture 2 on Object Detection, we saw that people used **pretrained networks on Regions of Interest** (RoIs) to extract features from image parts and output a **class probability** for what we see in that RoI.

Credit: Long et al. (2014)



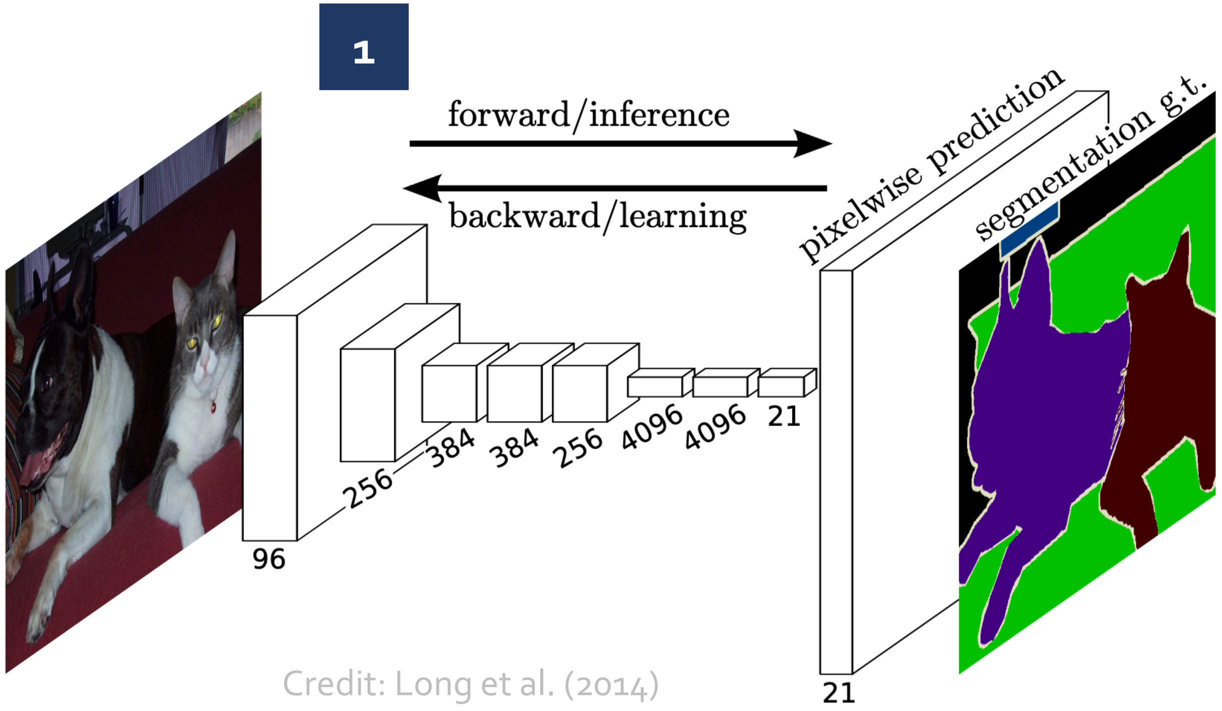
Output is a coarse feature (heat) map of the object.

- The key idea of Long et al. (2014) is to **"reinterpret" the final fully connected layers** outputting the class score as convolutions by "reshaping" weights into a matrix.
- As convolutions are resolution independent, we can use a kernel that covers the entire image.

DL segmentation methods

Fully convolutional networks III

- Using transfer learning, we adjust final convolutional layer to not represent 1000 classes but a number suitable for our target task, e.g., 21.

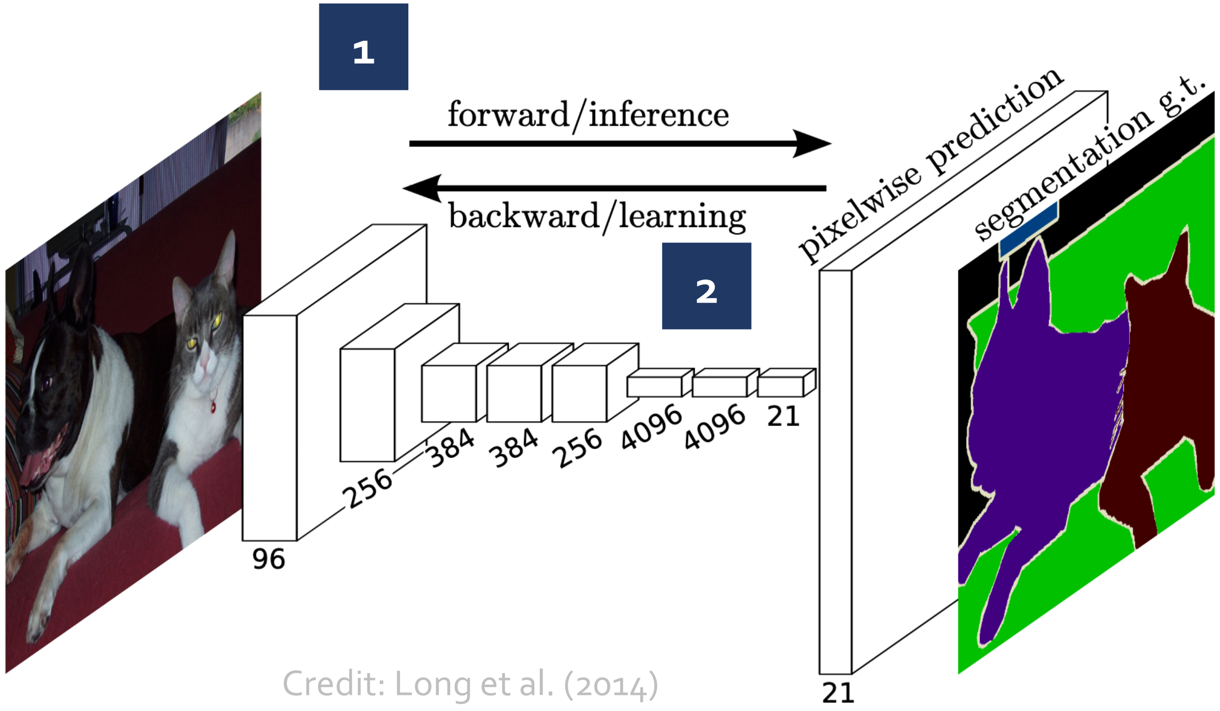


Credit: Long et al. (2014)

Step 1: Forward pass (inference) via convolutions decreases the size of feature maps and gives global information about WHAT we see.

DL segmentation methods
Fully convolutional networks III

- Using transfer learning, we adjust final convolutional layer to not represent 1000 classes but a number suitable for our target task, e.g., 21.



Credit: Long et al. (2014)

Step 1: Forward pass (inference) via convolutions decreases the size of feature maps and gives global information about WHAT we see.

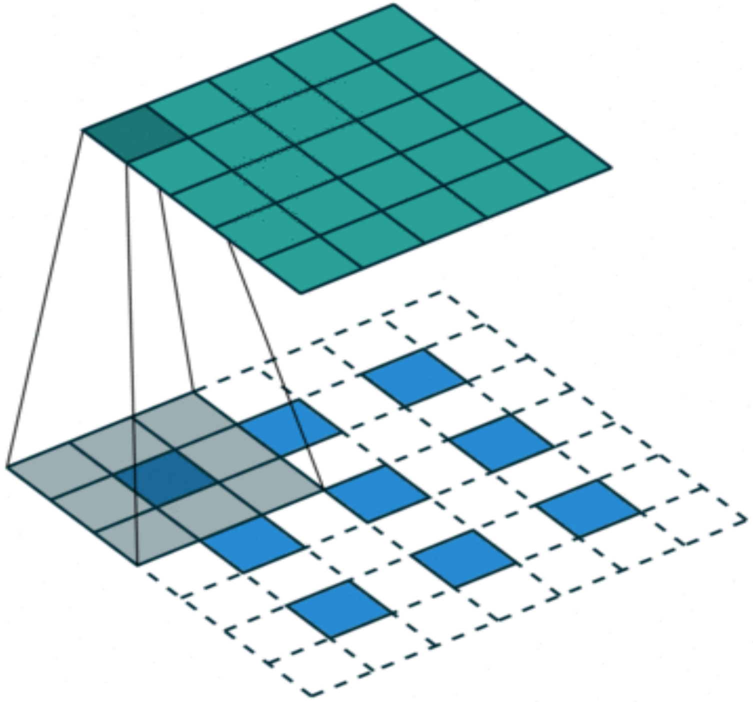
Step 2: Backward pass (learning) performs backward convolutions (i.e., deconvolutions or upsampling) to increase feature map size and provide local information about WHERE object is located in image.

DL segmentation methods

Fully convolutional networks III

- Using transfer learning, we adjust final convolutional layer to not represent 1000 classes but a number suitable for our target task, e.g., 21.

The blue map corresponds to the input, the green one to the output.



Credit: https://github.com/vdumoulin/conv_arithmetic

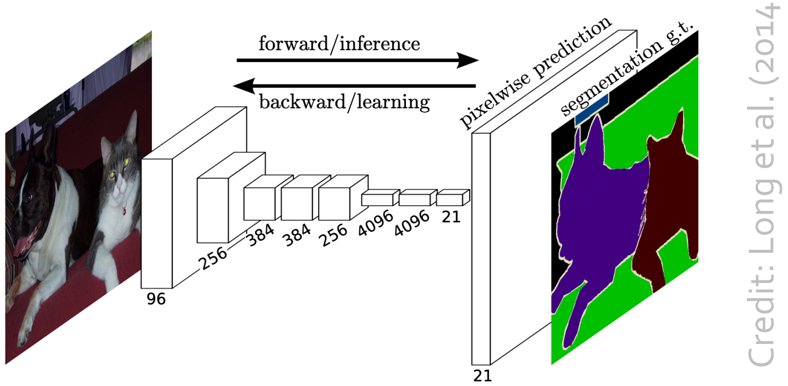
Step 1: Forward pass (inference) via convolutions decreases the size of feature maps and gives global information about WHAT we see.

Step 2: Backward pass (learning) performs backward convolutions (i.e., deconvolutions or upsampling) to increase feature map size and provide local information about WHERE object is located in image.

DL segmentation methods

Fully convolutional networks IV

- The fully convolutional network is optimised using a pixel-based loss that is computed with respect to the segmentation ground truth.

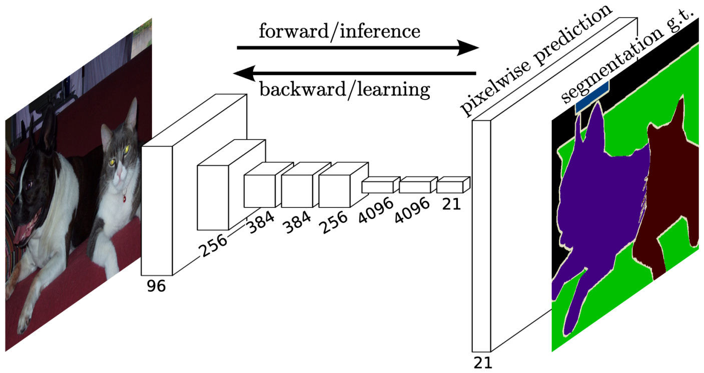


Credit: Long et al. (2014)

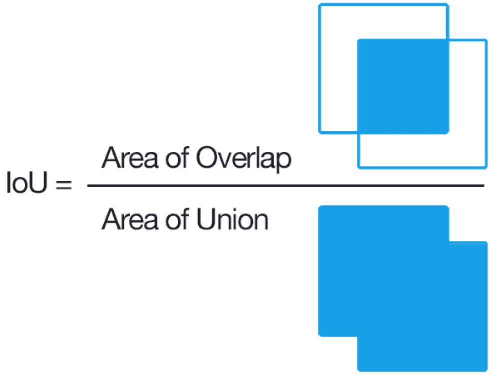
DL segmentation methods

Fully convolutional networks IV

- The fully convolutional network is optimised using a pixel-based loss that is computed with respect to the segmentation ground truth.



Credit: Long et al. (2014)



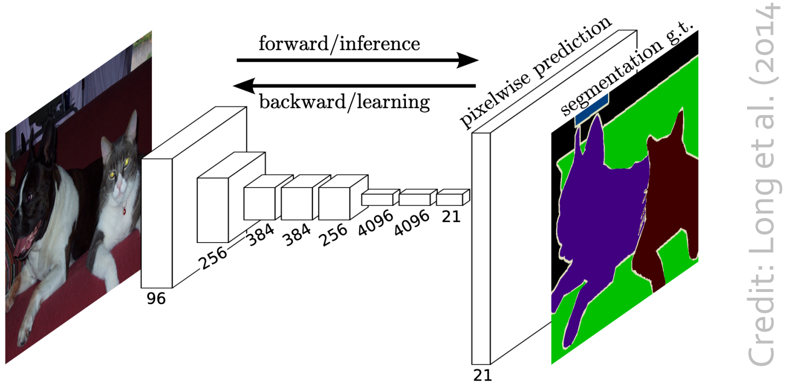
Credit: https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

- Several metrics exist for evaluation
  - Pixel accuracy
  - Mean accuracy
  - Mean IoU (Intersection over union)
  - Frequency weighted IoU

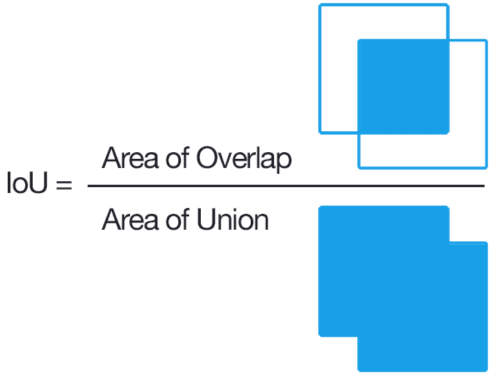
DL segmentation methods

Fully convolutional networks IV

- The fully convolutional network is optimised using a pixel-based loss that is computed with respect to the segmentation ground truth.



Credit: Long et al. (2014)



Credit: https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

- Several metrics exist for evaluation
  - Pixel accuracy
  - Mean accuracy
  - Mean IoU (Intersection over union)
  - Frequency weighted IoU

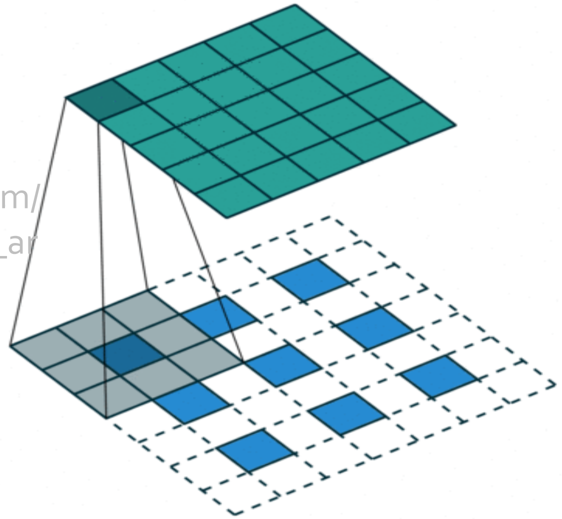
	FCN-AlexNet	FCN-VGG16	FCN-GoogLeNet[4]
mean IU	39.8	**56.0**	42.5
forward time	50 ms	210 ms	59 ms
conv. layers	8	16	22
parameters	57M	134M	6M
rf size	355	404	907
max stride	32	32	32

Long et al. performed experiments with different pretrained NNs (backbones). VGG16 performs best in terms of mean IoU but is slower than the other two.

DL segmentation methods

Fully convolutional networks V

Credit: https://github.com/vdumoulin/conv_arithmetic



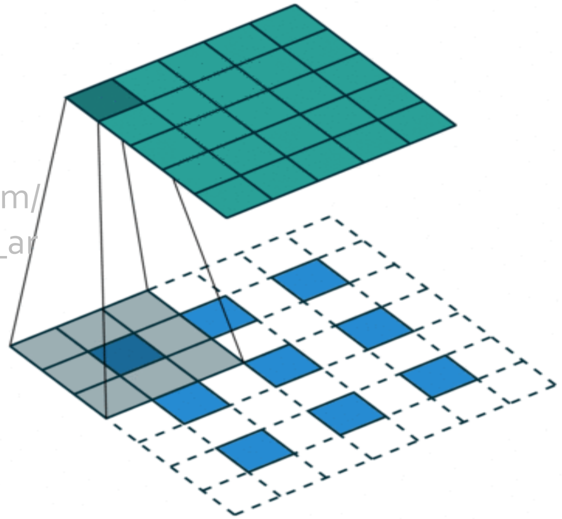
- The upsampling procedure discussed previously has a key disadvantage: The resulting heatmaps are **very coarse**.
- Long et al. develop a procedure to enhance the output by adding links combining final prediction layers with lower layers with finer strides.

DL segmentation methods

Fully convolutional networks V

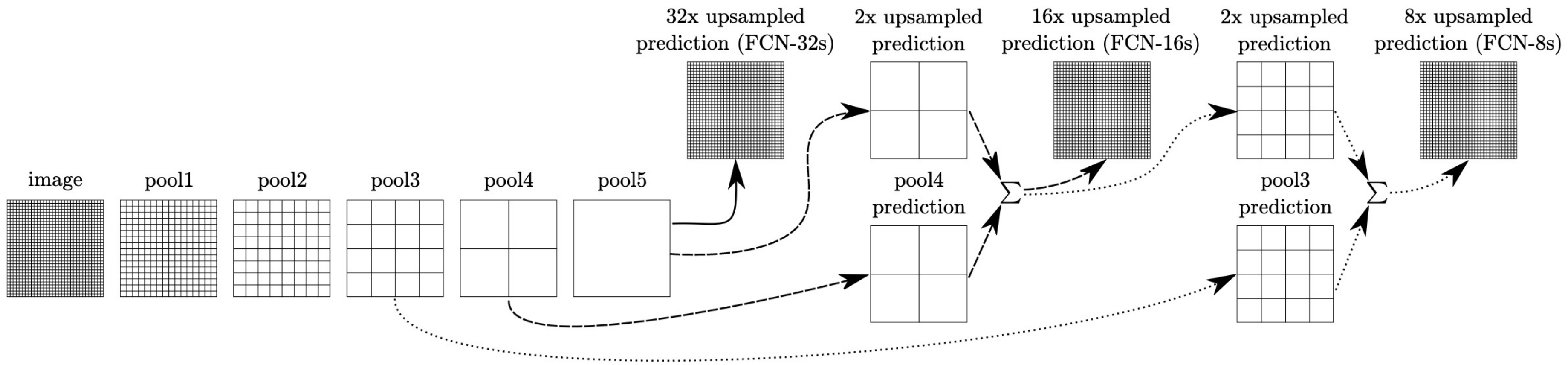
- The upsampling procedure discussed previously has a key disadvantage: The resulting heatmaps are **very coarse**.

Credit: https://github.com/vdumoulin/conv_arithmetic



- Long et al. develop a procedure to enhance the output by adding links combining final prediction layers with lower layers with finer strides.

This combines fine (low layer) and coarse (high layer) information to make local predictions that respect global structure.

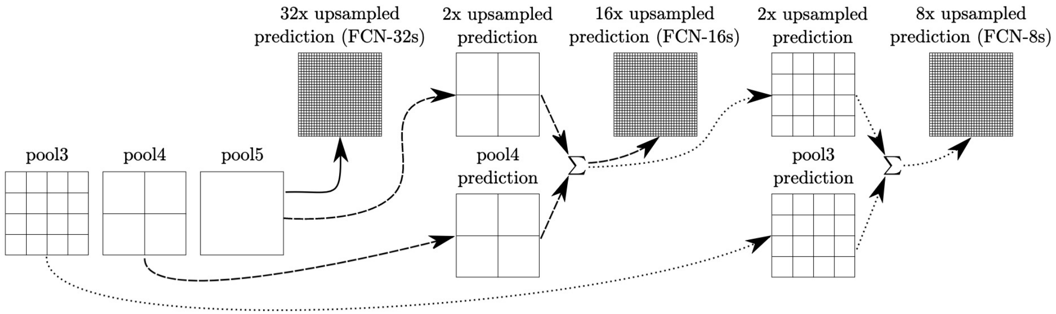


Different ways to fuse low-level and high-level information for image segmentation. Credit: Long et al. (2014)

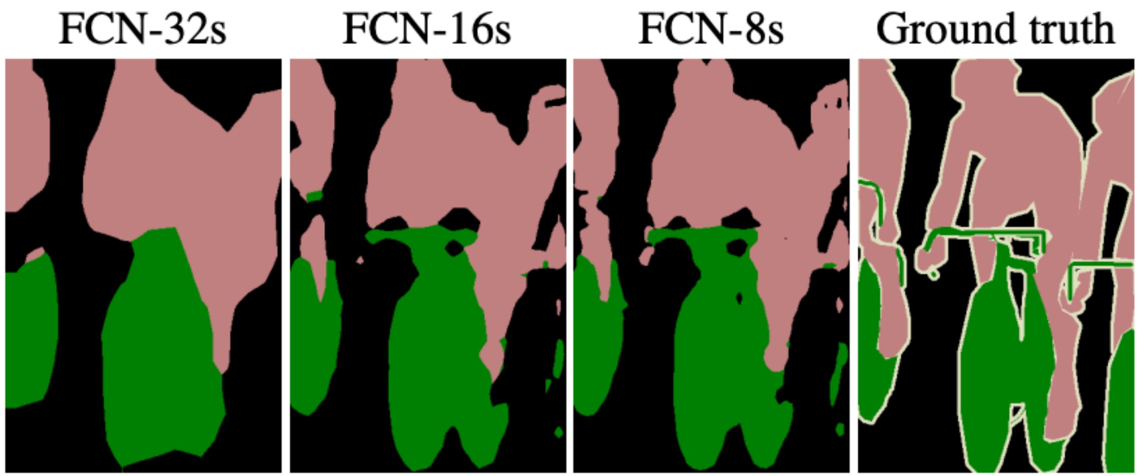
DL segmentation methods

Fully convolutional networks VI

- The impact of these different fusion prescriptions are illustrated below:



Credit: Long et al. (2014)



Credit: Long et al. (2014)

	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s-fixed	83.0	59.7	45.4	72.0
FCN-32s	89.1	73.3	59.4	81.4
FCN-16s	90.0	75.7	62.4	83.0
FCN-8s	**90.3**	**75.9**	**62.7**	**83.2**

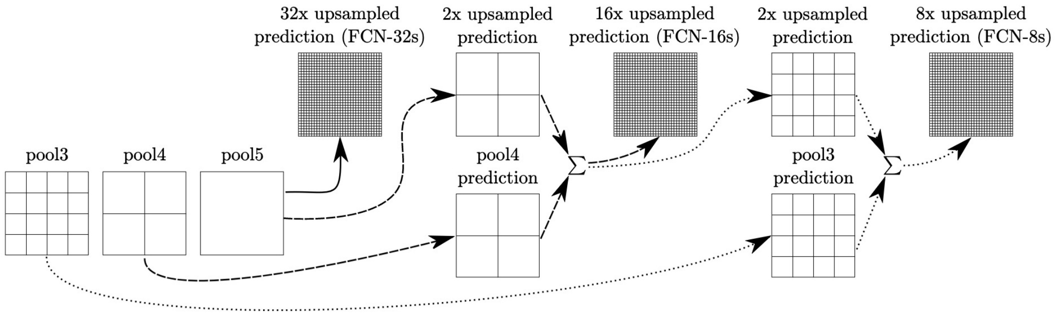
By adding information from lower layers, we recover finer structures in our segmentation masks.

DL segmentation methods

Fully convolutional networks VI

- The impact of th[illegible]ifferent fusion prescriptions a[illegible]d below:

Credit: Long et al. (2014)



Credit: Long et al. (2014)

	pixel acc.	mean acc.	mean IU	f.w. IU
[illegible]	[illegible]	59.7	45.4	72.0
[illegible]	[illegible]	[illegible]	59.4	81.4
[illegible]	[illegible]	[illegible]	62.4	83.0
FC[illegible]	[illegible]	[illegible]	**62.7**	**83.2**

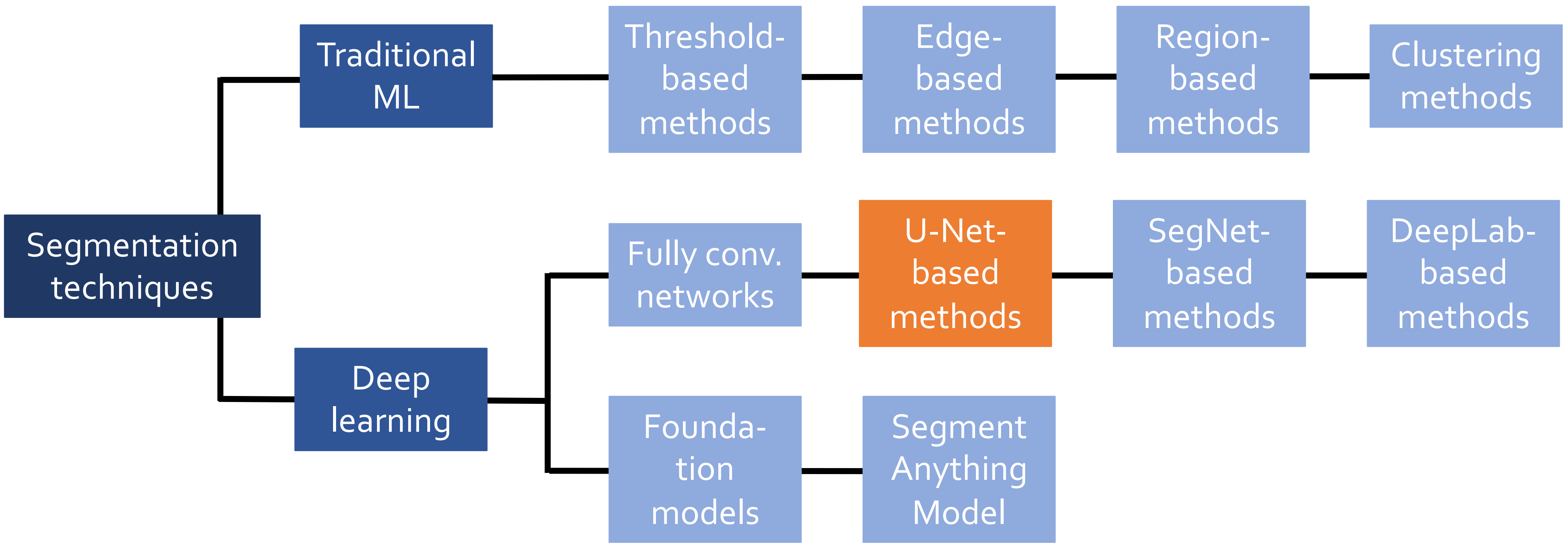
This approach outperformed traditional methods and inferences are much faster.

By adding information from lower layers, we recover more fine structures in our segmentation masks.

Segmentation methods

Overview of techniques

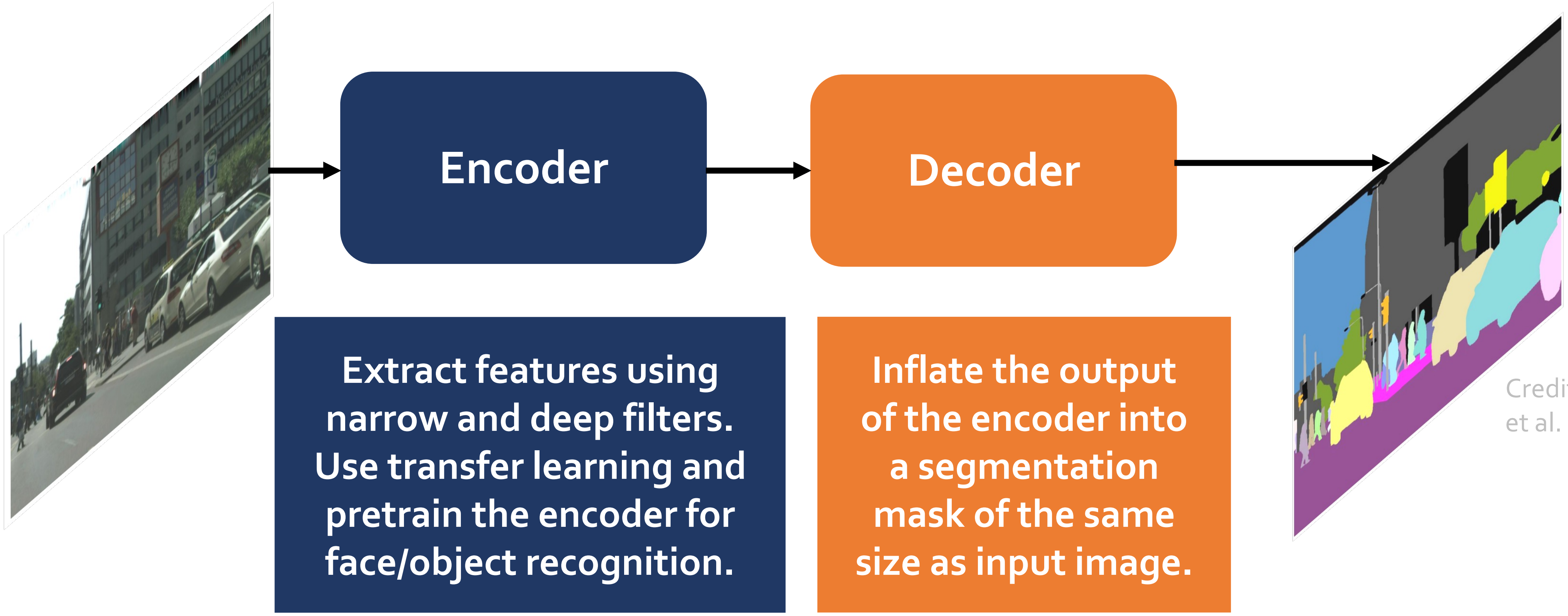
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



DL segmentation methods

Encoder-decoder networks

- A second class of methods is based on the idea of first encoding the information within an image and subsequently decoding it:

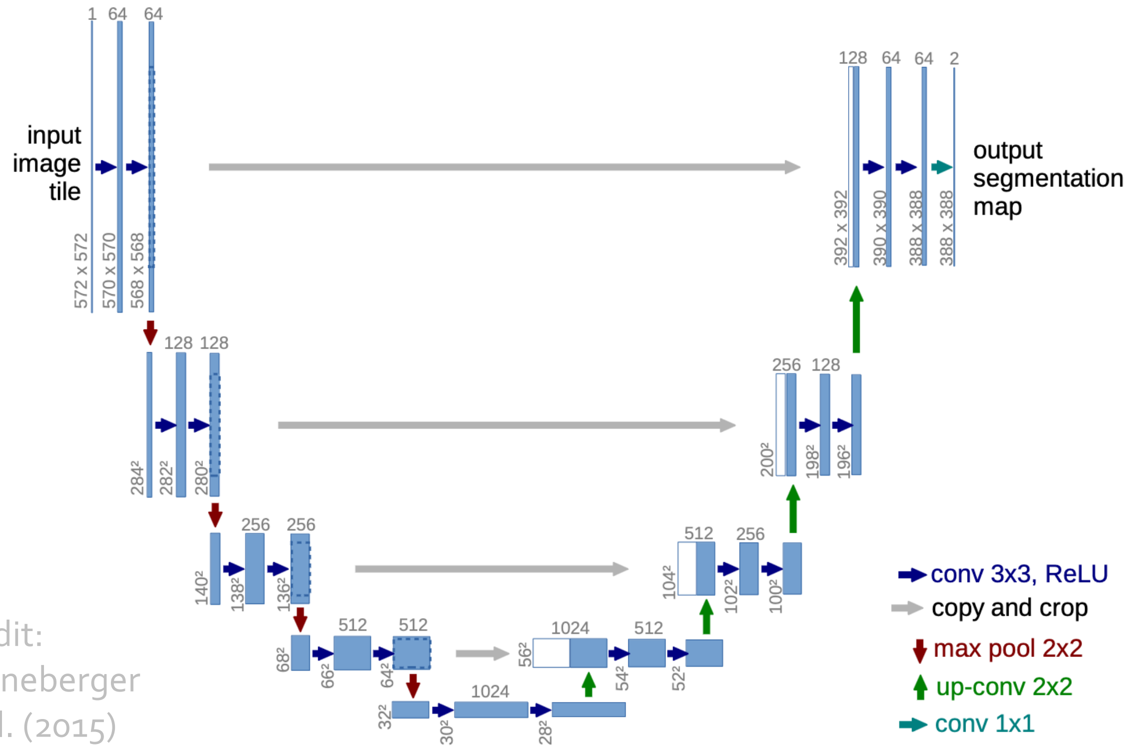


Credit: Kirillov et al. (2019)

DL segmentation methods

U-Net-based methods I

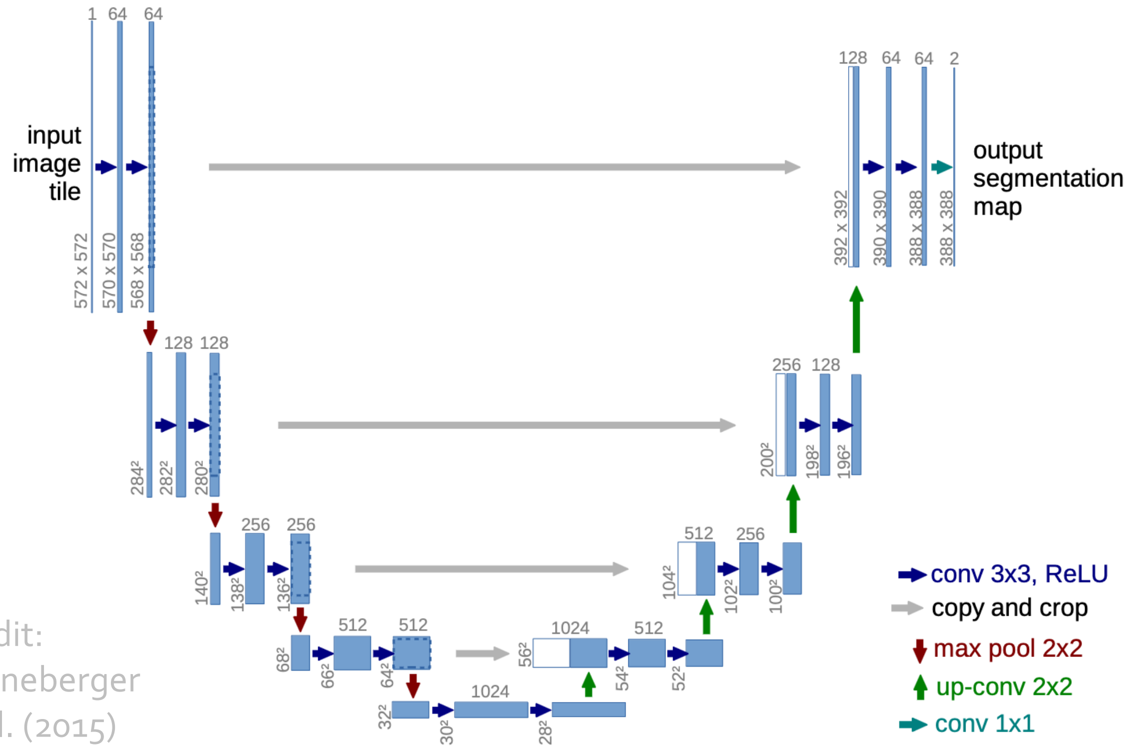
- Ronneberger et al. (2015) combine this idea with that of Fully Convolutional Networks for biomedical image segmentation, where labelled training data is very limited. Their **U-Net architecture** is as follows:



Credit: Ronneberger et al. (2015)

DL segmentation methods
U-Net-based methods I

- Ronneberger et al. (2015) combine this idea with that of Fully Convolutional Networks for biomedical image segmentation, where labelled training data is very limited. Their **U-Net architecture** is as follows:



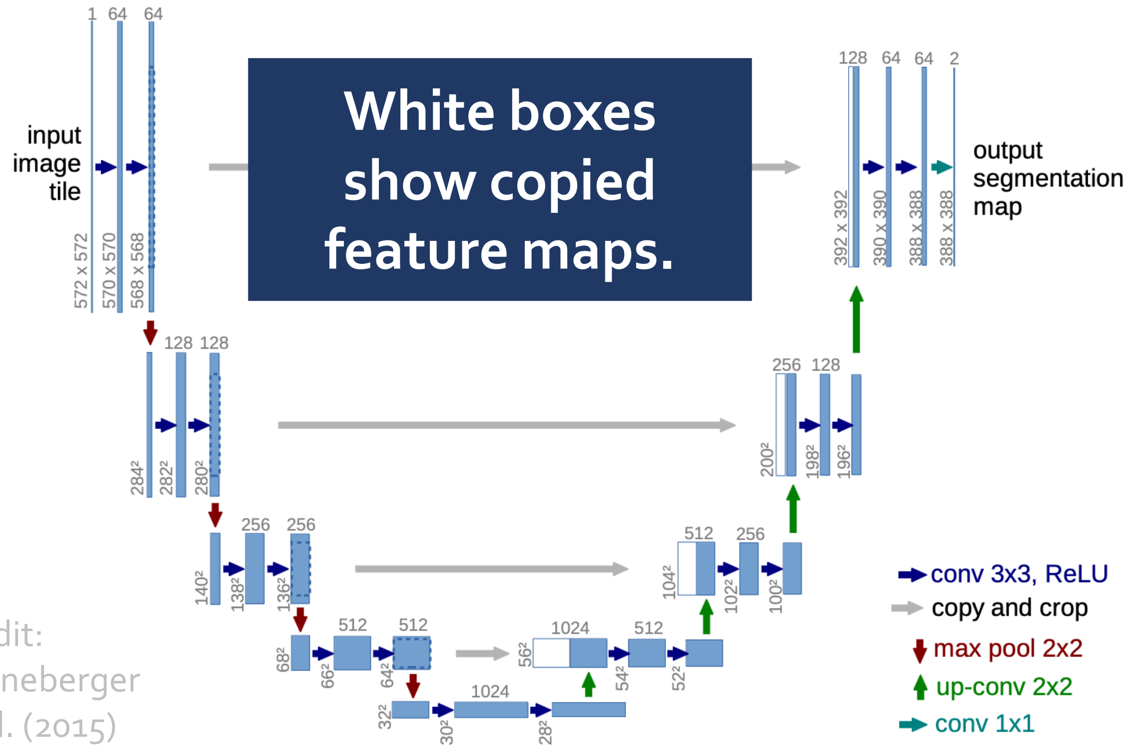
Credit: Ronneberger et al. (2015)

Each blue box corresponds to a multi-channel feature map. The number of channels is shown at the top, the x-y dimensions at the bottom left of each box.

DL segmentation methods

U-Net-based methods I

- Ronneberger et al. (2015) combine this idea with that of Fully Convolutional Networks for biomedical image segmentation, where labelled training data is very limited. Their **U-Net architecture** is as follows:



White boxes show copied feature maps.

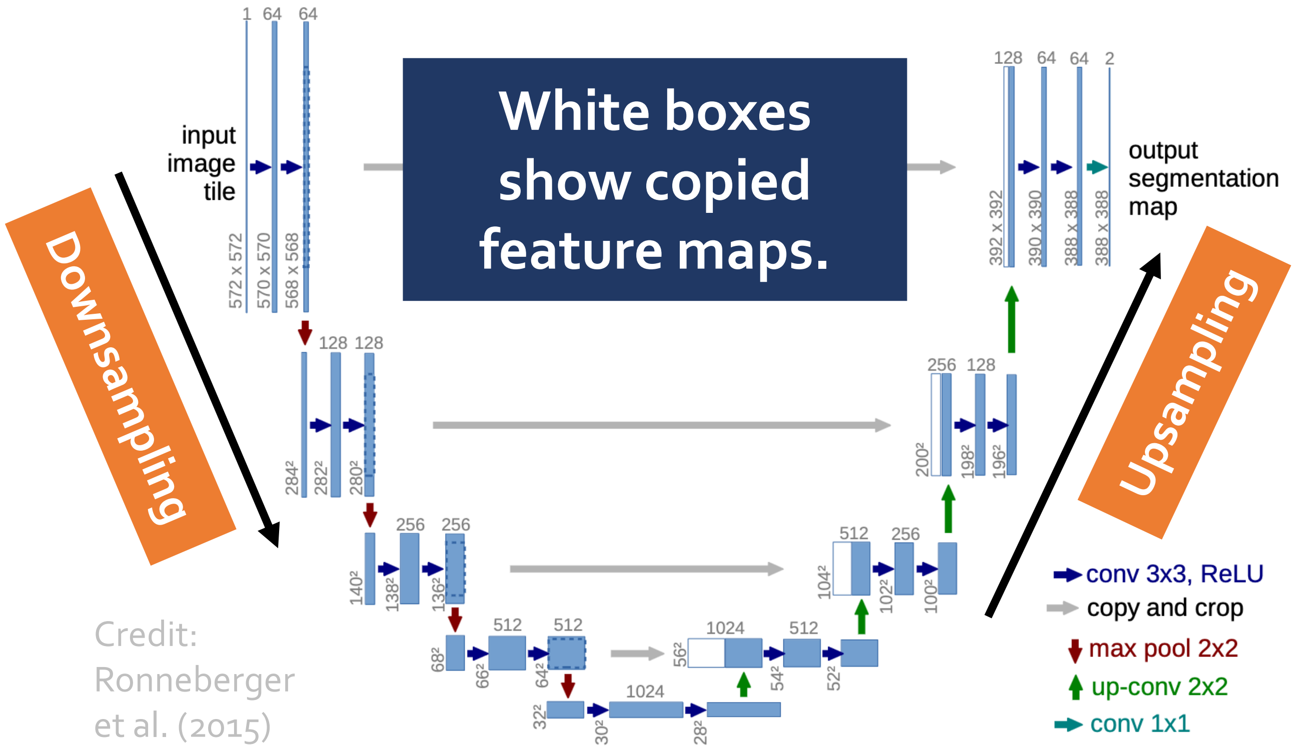
Each blue box corresponds to a multi-channel feature map. The number of channels is shown at the top, the x-y dimensions at the bottom left of each box.

Credit: Ronneberger et al. (2015)

DL segmentation methods

U-Net-based methods I

- Ronneberger et al. (2015) combine this idea with that of Fully Convolutional Networks for biomedical image segmentation, where labelled training data is very limited. Their **U-Net architecture** is as follows:



Credit: Ronneberger et al. (2015)

Each blue box corresponds to a multi-channel feature map. The number of channels is shown at the top, the x-y dimensions at the bottom left of each box.

Compared to Long et al., the upsampling branch in U-Net has many more feature channels to give a symmetric architecture.

DL segmentation methods

U-Net-based methods II

- The network is optimised by minimising the soft-max function $\mathrm{p_l}(\mathbf{x})$ of the final feature map combined with the cross-entropy loss function for each pixel $\mathbf{x} \in \Omega$ in our input image Ω with respect to ground truth $\mathrm{l}(\mathbf{x})$:

$$\mathrm{L} = -\sum_{\mathbf{x}\in\Omega} \mathrm{w}(\mathbf{x})\log[\mathrm{p}_{\mathrm{l}(\mathbf{x})}(\mathbf{x})]$$

$$\mathrm{w}(\mathbf{x}) = \text{weight map}$$

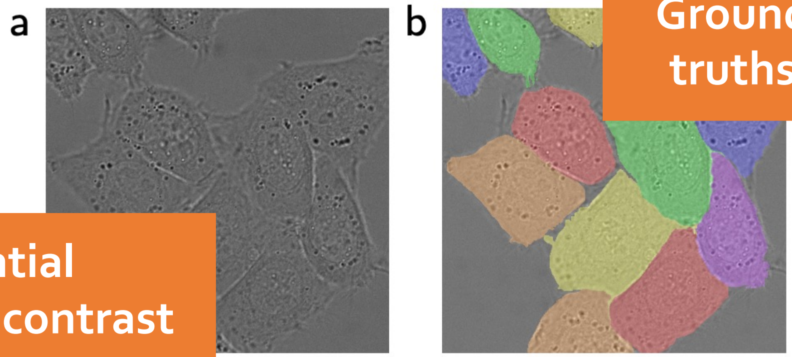
DL segmentation methods

U-Net-based methods II

- The network is optimised by minimising the soft-max function $\mathrm{p_l}(\mathbf{x})$ of the final feature map combined with the cross-entropy loss function for each pixel $\mathbf{x} \in \Omega$ in our input image Ω with respect to ground truth $\mathrm{l}(\mathbf{x})$:

$$\mathrm{L} = -\sum_{\mathbf{x}\in\Omega} \mathrm{w}(\mathbf{x})\log[\mathrm{p}_{\mathrm{l}(\mathbf{x})}(\mathbf{x})]$$

$\mathrm{w}(\mathbf{x}) = \text{weight map}$



Ground truths

Differential interference contrast microscopy images of HeLa cells

Credit: Ronneberger et al. (2015)

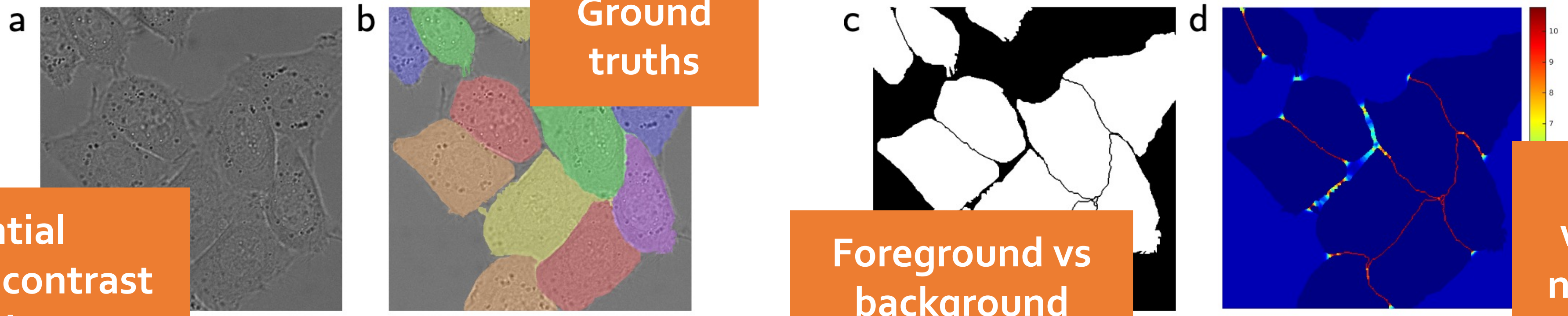
DL segmentation methods

U-Net-based methods II

- The network is optimised by minimising the soft-max function $p_l(\mathbf{x})$ of the final feature map combined with the cross-entropy loss function for each pixel $\mathbf{x} \in \Omega$ in our input image Ω with respect to ground truth $l(\mathbf{x})$:

$$L = -\sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \log[p_{l(\mathbf{x})}(\mathbf{x})]$$

$$w(\mathbf{x}) = \text{weight map}$$

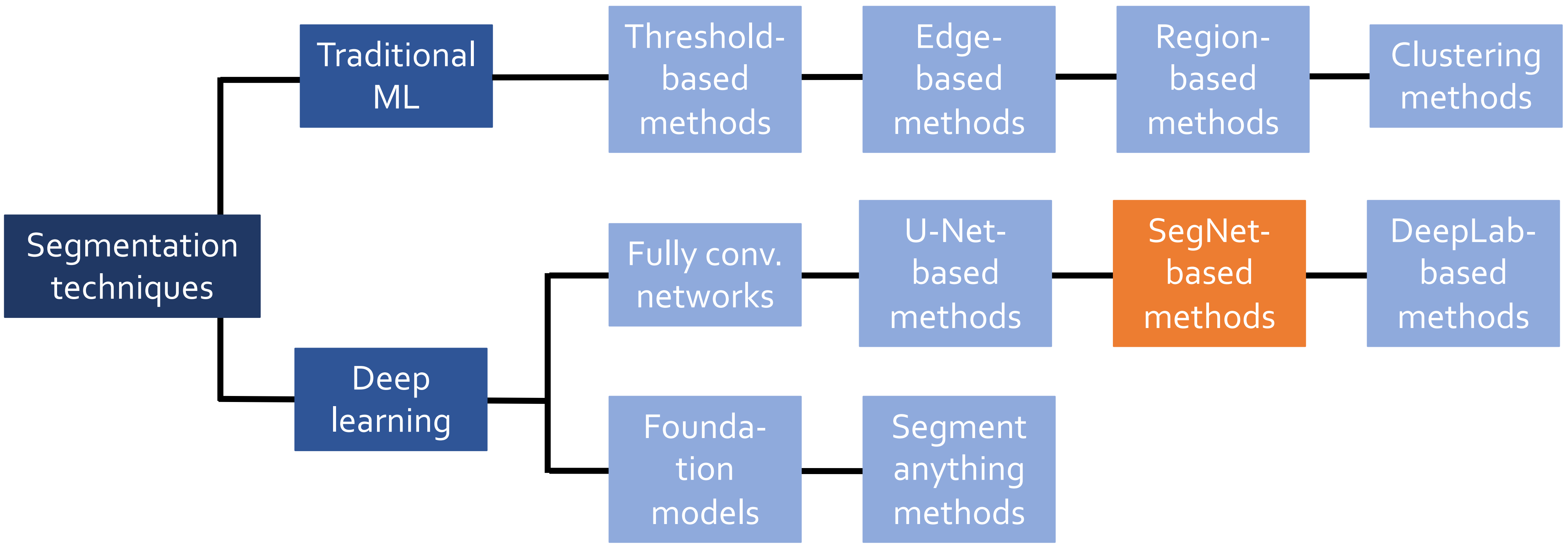


Credit: Ronneberger et al. (2015)

Segmentation methods

SegNet-based methods

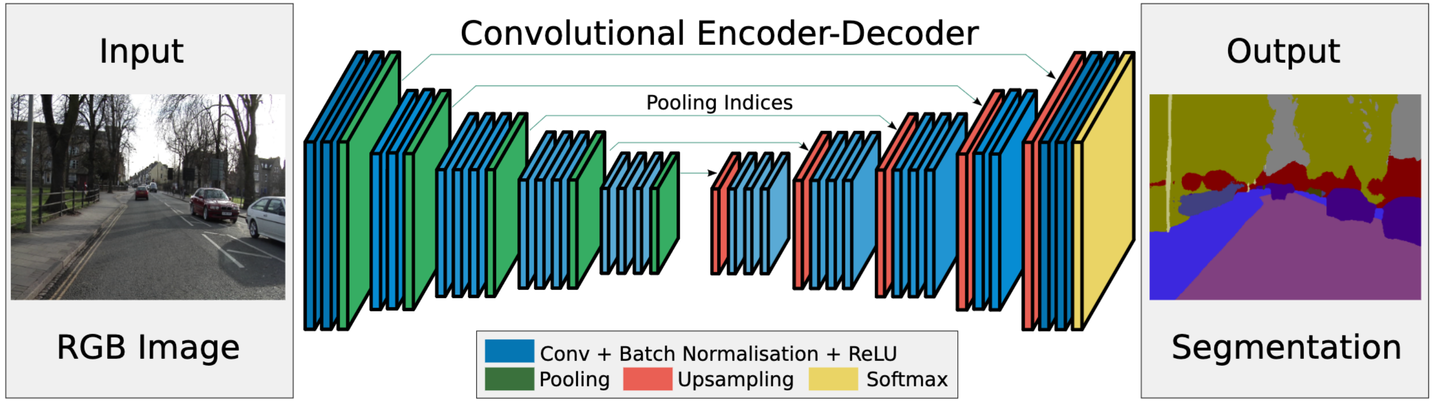
- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.

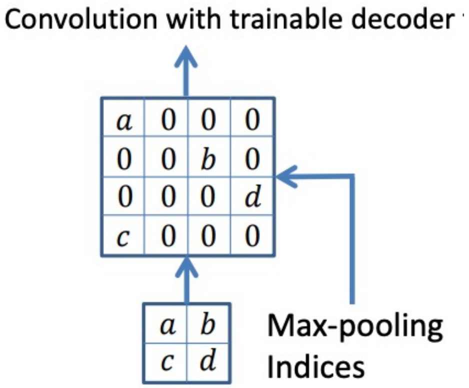


DL segmentation methods
SegNet-based methods

- SegNet was developed by Badrinarayanan et al. (2016) to perform semantic segmentation. It again is fully convolutional (does not contain any fully connected / dense layers).

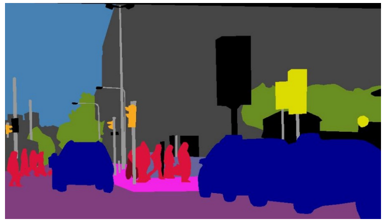
Credit: Badrinarayanan et al. (2016)





DL segmentation methods

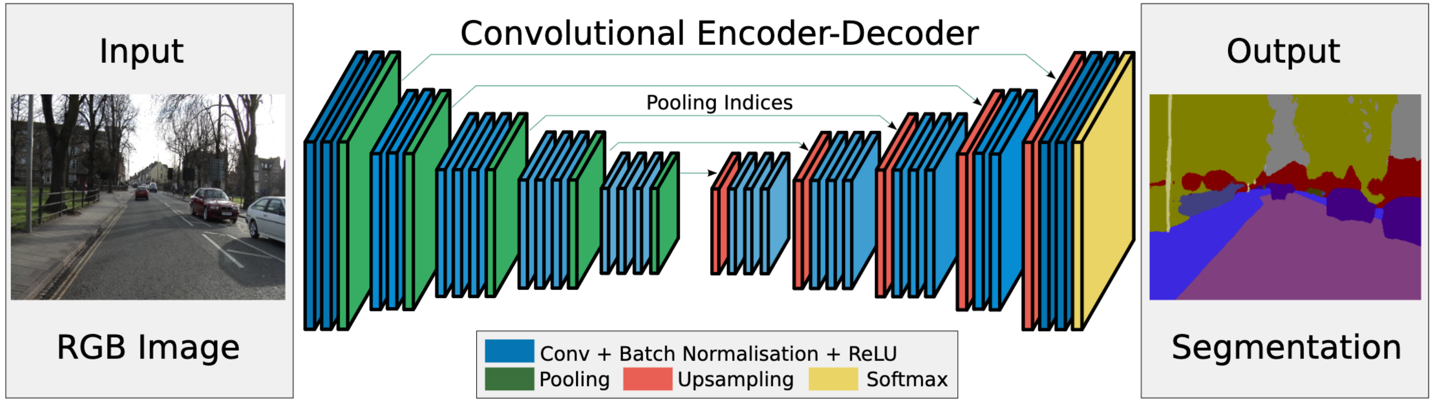
SegNet-based methods

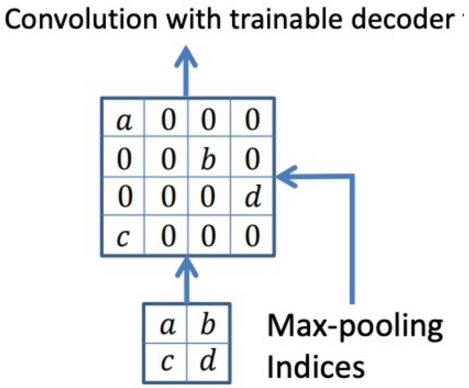


Semantic segmentation

- SegNet was developed by Badrinarayanan et al. (2016) to perform semantic segmentation. It again is fully convolutional (does not contain any fully connected / dense layers).

Credit: Badrinarayanan et al. (2016)

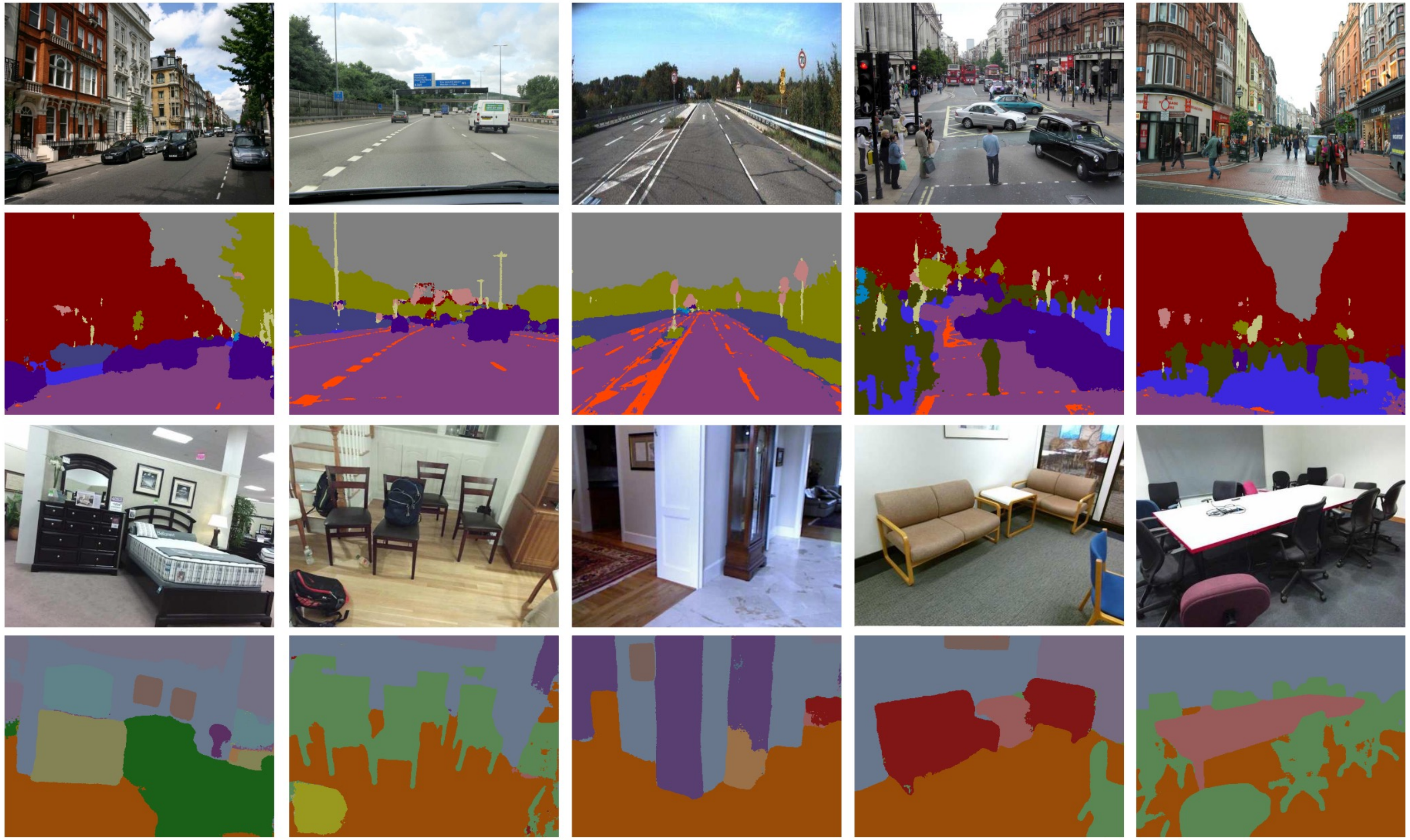




Compared to U-Net, the decoder in SegNet upsamples its input using pool indices transferred from the encoder branch to produce (not full but) sparse feature maps.

DL segmentation methods
SegNet-based methods

- SegNet performs well for “**scene” understanding applications**, where efficiency in terms of memory and computation time are needed.

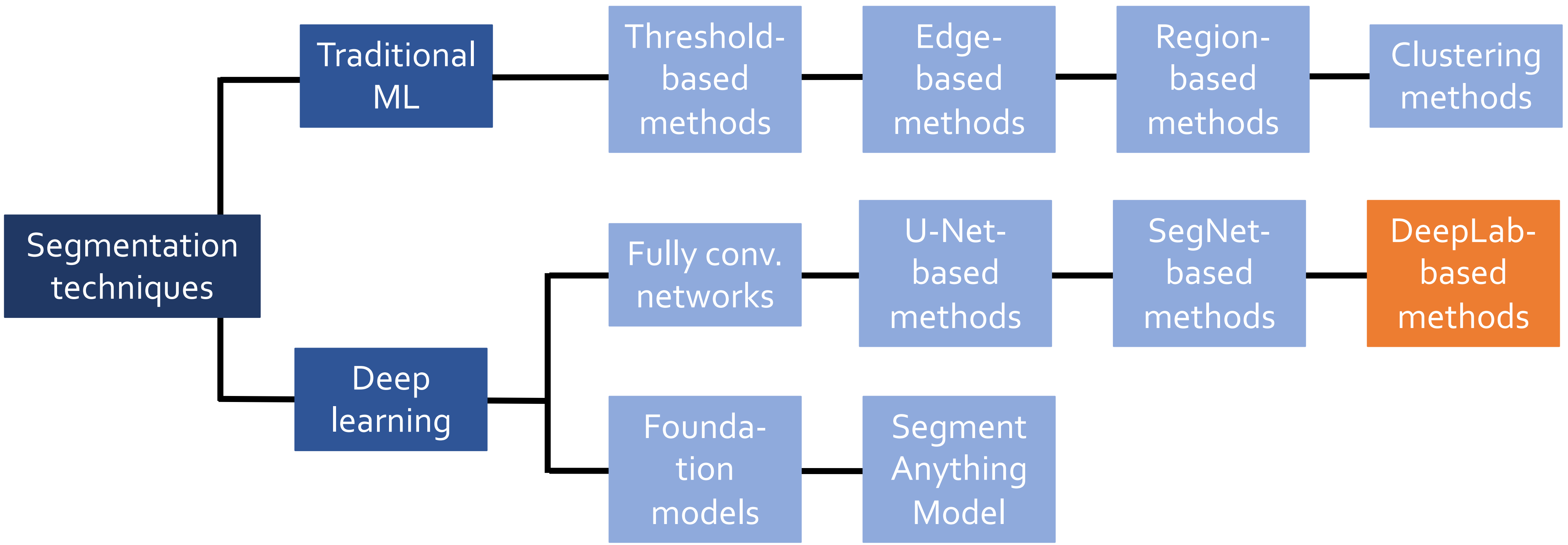


Credit: Badrinarayanan et al. (2016)

Segmentation methods

SegNet-based methods

- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



DL segmentation methods
DeepLab-based methods I

- Also focused on semantic image segmentation, DeepLab was developed by Chen et al. (2017) to address **several issues**:

  a.) Reduced feature resolution in upsampled images.

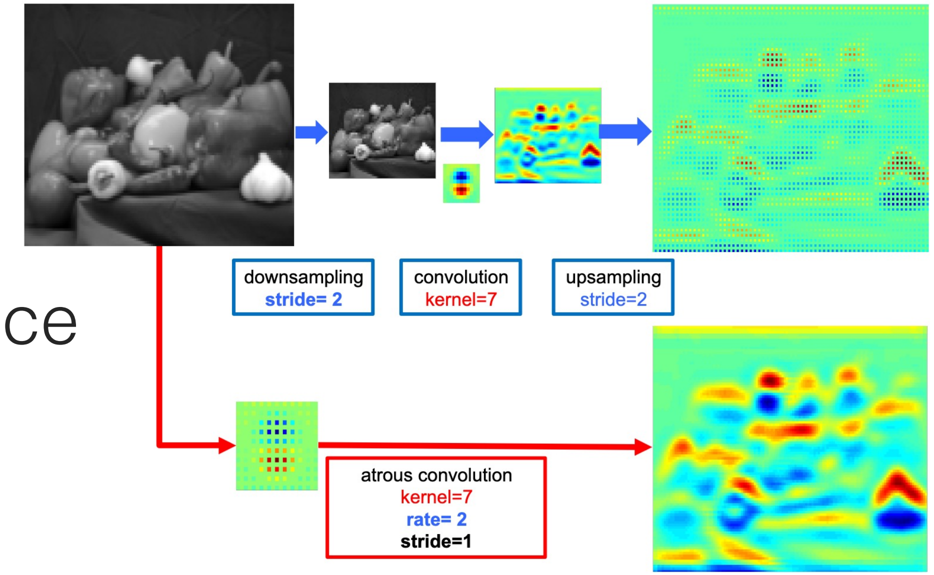
  b.) Existence of objects at multiple scales.

  c.) Reduced localisation accuracy due to translational invariance.

DL segmentation methods
DeepLab-based methods I

- Also focused on semantic image segmentation, DeepLab was developed by Chen et al. (2017) to address **several issues**:
  - a.) Reduced feature resolution in upsampled images.
  - b.) Existence of objects at multiple scales.
  - c.) Reduced localisation accuracy due to translational invariance.

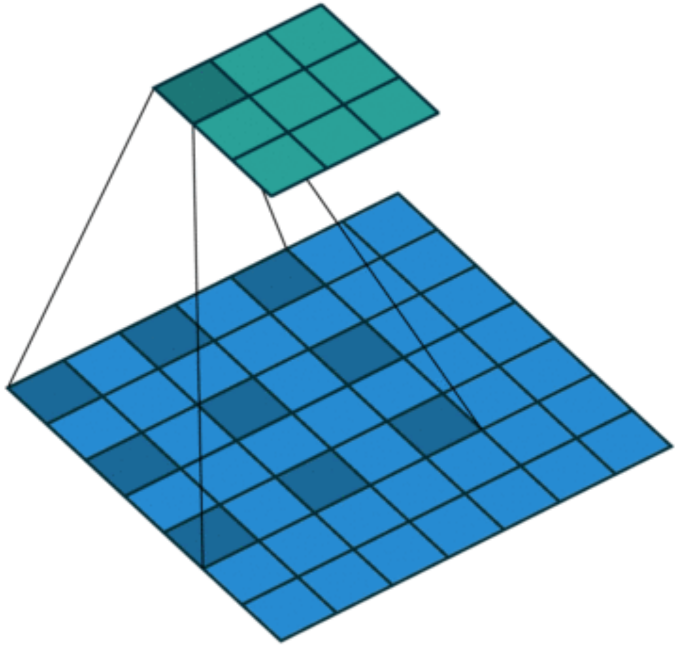
- **Issue a.)**: During downsampling, we reduce the resolution due to striding and pooling. By upsampling the images with random kernels, we introduce holes ("zeros") into our images. We get **sparse features**.



Credit: Chen et al. (2016)

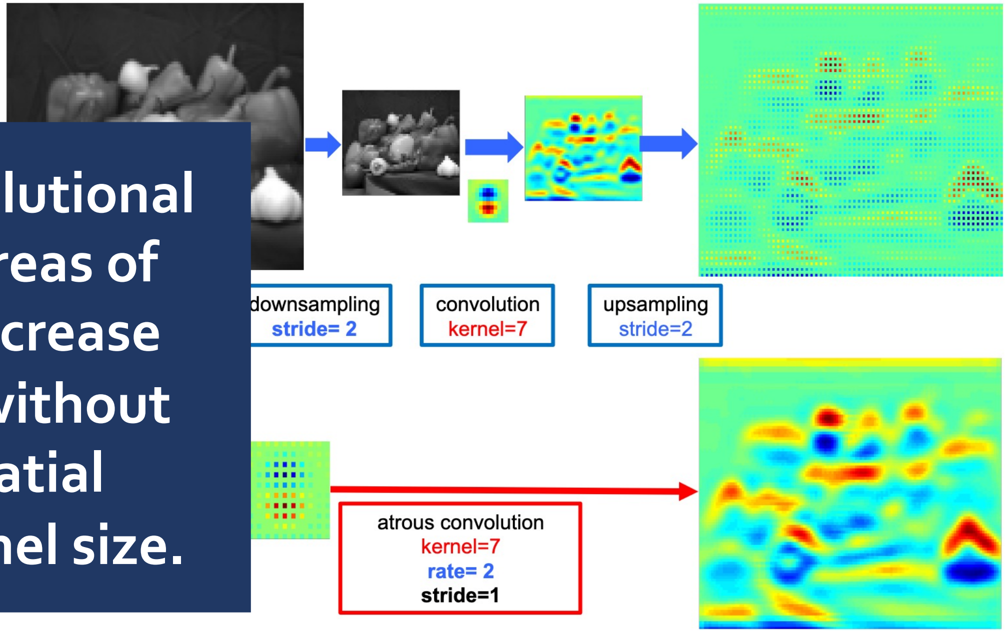
DL segmentation methods
DeepLab-based methods II

- To overcome the problem of holes, Chen et al. used so-called **atrous or dilated convolutions**. The gif below shows this convolution in action for a dilated 3x3 kernel with rate r=2.



Credit: https://github.com/vdumoulin/conv_arithmetic

This allows the convolutional filter to see larger areas of the input (i.e., we increase the receptive field) without decreasing the spatial resolution or the kernel size.



Credit: Chen et al. (2016)

DL segmentation methods

DeepLab-based methods III

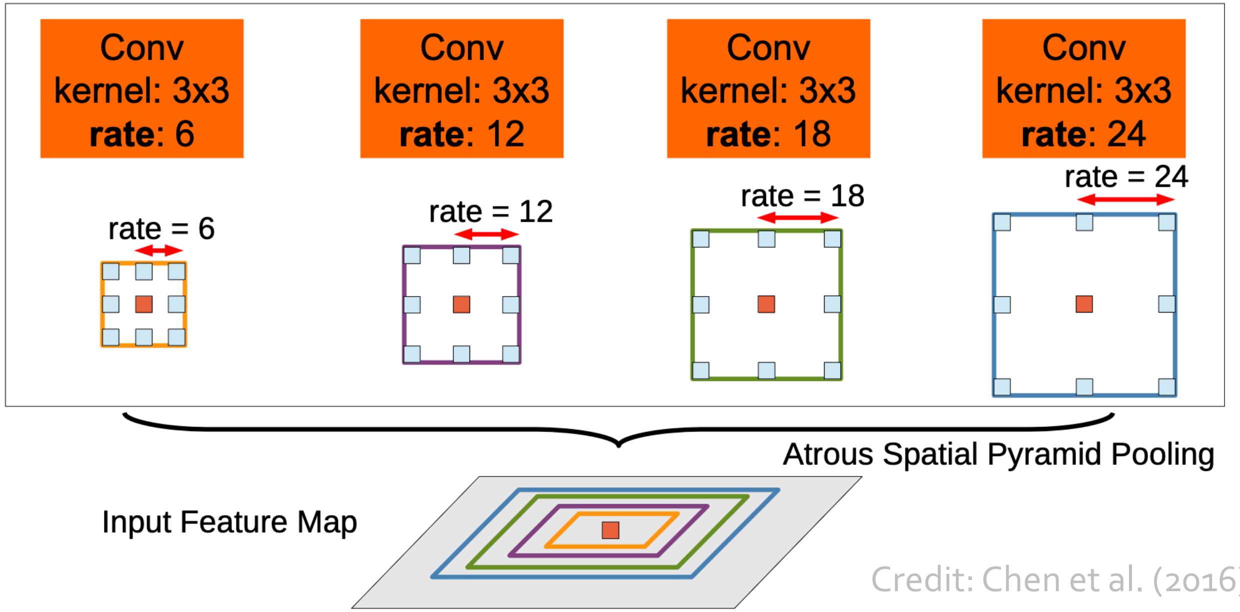
- Issue b.): Chen et al. take the idea of dilated convolution further to also successfully segment objects that exist at different scales. To this end, they use different rate factors for the same fixed 3x3 kernel to construct fields of view of different sizes around a centre pixel:

DL segmentation methods
DeepLab-based methods III

- **Issue b.):** Chen et al. take the idea of dilated convolution further to also successfully segment **objects** that exist at **different scales**. To this end, they use different rate factors for the same fixed 3x3 kernel to construct fields of view of different sizes around a centre pixel:

Smaller fields of view are sensitive to smaller objects, while larger fields of view probe larger objects.

Various fields of view for a centre pixel are concatenated to probe multiple scales.



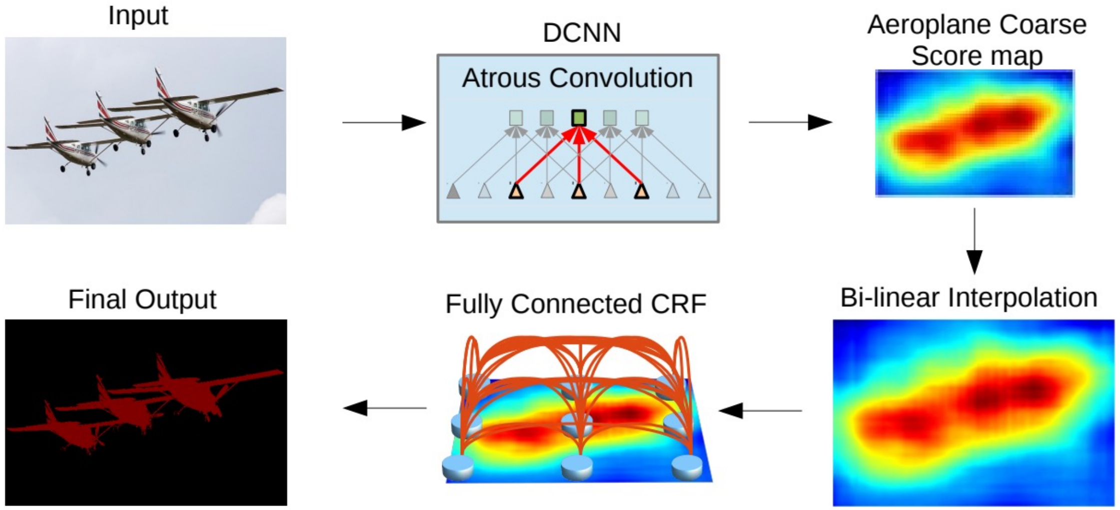
Credit: Chen et al. (2016)

DL segmentation methods
DeepLab-based methods IV

- Issue c.): To deal with the translational invariance of convolutions and the resulting complication in localising objects, Chen et al. apply so-called fully connected Conditional Random Fields (CRFs).

CRFs are probabilistic filters for pattern recognition that capture the long-range dependencies in data, i.e., contextual relationships.

CRFs improve the segmentation results and refine object boundaries based on labels of neighbouring pixels.



Credit: Chen et al. (2016)

DL segmentation methods

DeepLab-based methods IV

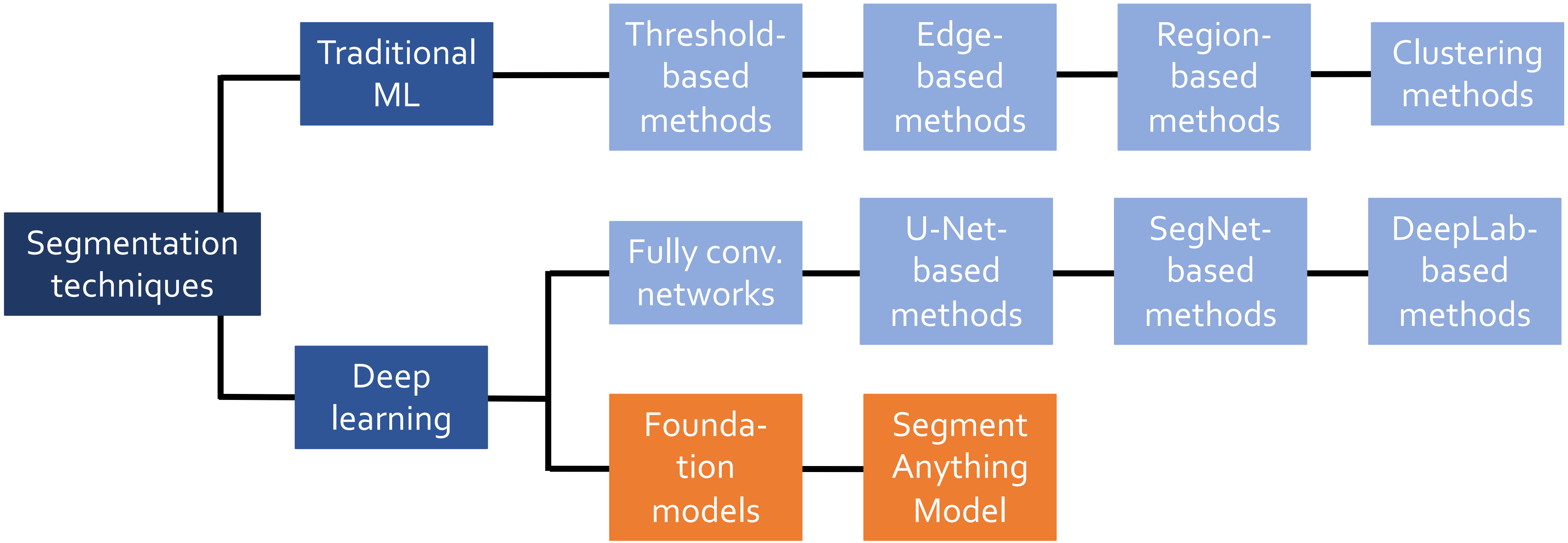
- Some results for the semantic segmentation with DeepLab:

(a) Image (b) Before CRF (c) After CRF (a) Image (b) Before CRF (c) After CRF (a) Image (b) Before CRF (c) After CRF

Credit: Chen et al. (2016)

Segmentation methods
SegNet-based methods

- As for other image processing tools, segmentation can be performed with classical ML approaches or neural-network based Deep Learning.



Segmentation methods
State-of-the-art DL approaches

- Foundation models in DL are those that are trained on huge numbers of diverse data to capture features that are universal to many different domains and then fine-tuned for different tasks (see block on LLMs).

Developed by Meta AI Research Segment Anything Model (SAM) is such a foundation model trained on over 1 billion masks that can segment any object in an image or video with high quality and efficiency.

Credit: Kirillov et al. (2023)

Credit: Meta AI Research

Question time

Mentimeter quiz

Let’s take a few minutes to recap what we have discussed so far:

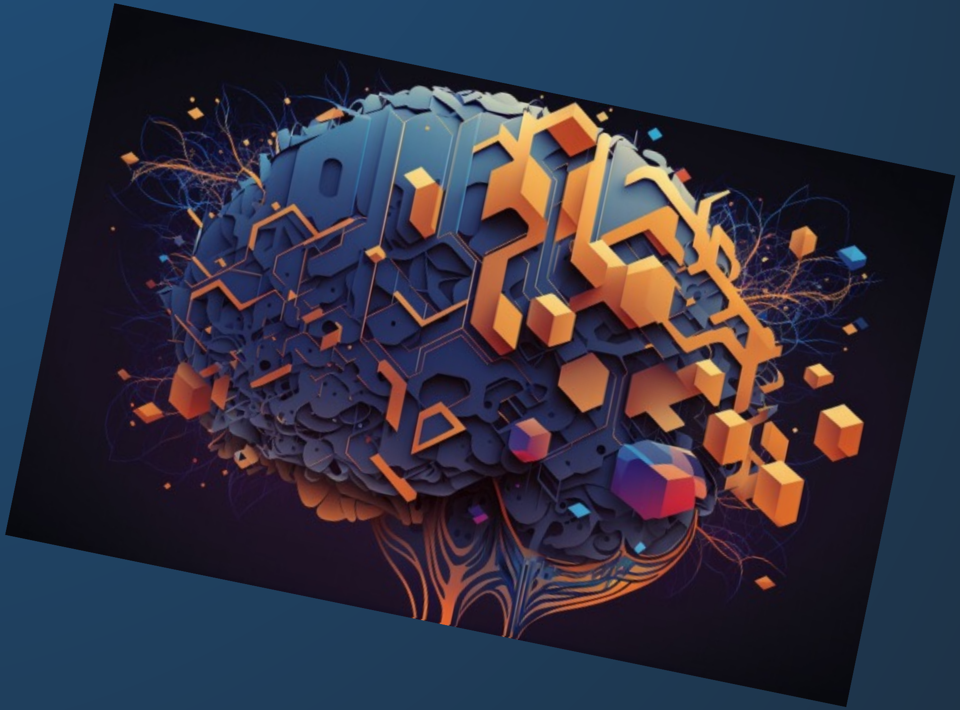
Go to menti.com and enter the code 7229 2576.

Segmentation overview

Segmentation methods

Assignment brief

Summary



Assignment brief
Overview

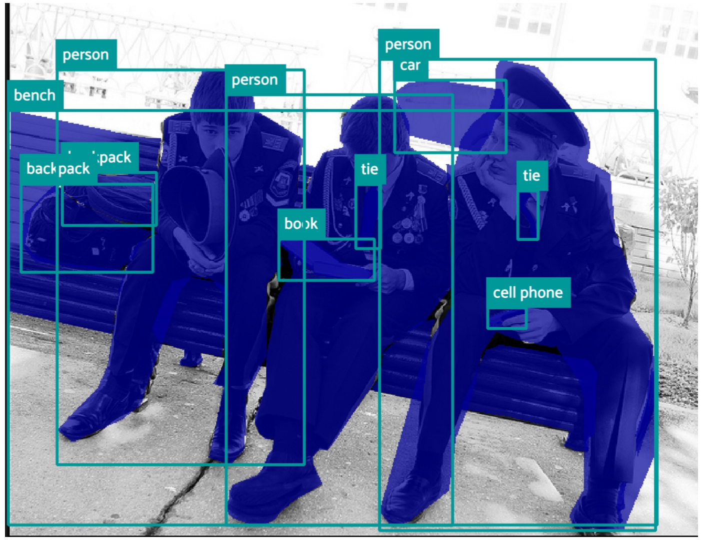
Deadline: 14th of June, 13:00

- Your first assignment will focus on **Image Segmentation** with Deep Learning using a subset of the **COCO-2017 dataset**, which contains images with labelled segmentation masks for various object classes (see Canvas).

- The assignment will build heavily on the first three tutorials. Tutorial 3 next week will introduce a tool (FiftyOne) to help with EDA on the COCO dataset.

- This is an **individual assignment** where each student will produce individual code & report on the data analysis they have performed.

- Learning to perform image segmentation will provide you with a **deeper understanding** and ability to **tackle computer vision tasks**.



Assignment brief

Task

- Your **practical work** will include:
  - Performing exploratory data analysis
  - Preprocessing of your data
  - Model selection and implementation
  - Evaluating your model performance

Assignment brief

Task

- Your **practical work** will include:
  - Performing exploratory data analysis
  - Preprocessing of your data
  - Model selection and implementation
  - Evaluating your model performance

You are advised to use Google Colab for your assignment.

I encourage you to put your code on GitHub for visibility.

Assignment brief
Task

- Your **practical work** will include:
  - Performing exploratory data analysis
  - Preprocessing of your data
  - Model selection and implementation
  - Evaluating your model performance
- Your **project report** will include:
  - Background and literature review
  - Dataset description and EDA
  - Methodology
  - Results and discussion
  - Outlook

You are advised to use Google Colab for your assignment.

I encourage you to put your code on GitHub for visibility.

Assignment brief
Task

- Your **practical work** will include:
  - Performing exploratory data analysis
  - Preprocessing of your data
  - Model selection and implementation
  - Evaluating your model performance
- Your **project report** will include:
  - Background and literature review
  - Dataset description and EDA
  - Methodology
  - Results and discussion
  - Outlook

You are advised to use Google Colab for your assignment.

I encourage you to put your code on GitHub for visibility.

Upload your report (including your code as an appendix) and a sharing (!) Google Colab link.

Report will have 700-1000 words (excluding refe-rences and code).

The key thing in your report is to critically analyse your results. Do not merely state your results.

Assignment brief

Rubric

- You will be evaluated on:
  - Implementation quality of the chosen model, data preprocesssing, model architecture, use of metrics and analytics.
  - Quality of analysis in report.
  - Critical evaluation of results and model as well as critical representation of the data.
  - Comparison to literature.
  - Writing quality including spelling, grammar and academic language.

Individual project report

Criteria	Ratings				Pts
Thoroughness of literature review Collect relevant and timely information about the topic of interest using peer-reviewed papers. Introduce the research field to the reader. Discuss the motivation and why research is being done in this field. Describe briefly what methods are used and the challenges involved in collecting and processing the relevant data before it its used in analysis. At least 3 peer-reviewed papers to be read and discussed.	**15 to >10.0 Pts** **Excellent** You have introduced the topic to the reader so that the rest of your report can be clearly understood by a non-expert. You briefly described innovative methods used in the field and elaborated on possible challenges involved in collecting and processing the data. You have clearly read, understood and used the findings of multiple papers in your literature review.	**10 to >5.0 Pts** **Good** You have used <3 papers as your source material. You introduced the topic but did not discuss the literature well.	**5 to >0.0 Pts** **Poor** You introduced the topic poorly and only used a single paper for the literature review.	**0 Pts** **No marks** You did not provide a literature review.	15 pts
Dataset and exploratory data analysis (EDA) Introduce the data and show a few plots that provide key information about the dataset. Describe any preprocessing you have done. Show that you have explored the dataset thoroughly and evaluated it by using visualisations and relevant metrics, plots.	**10 to >7.0 Pts** **Excellent** Your exploratory data analysis is insightful and well-justified. You made use of visualisation tools such as histograms, correlation matrices, chord diagrams, maps, etc. You managed to gain intuition from the visualisations in order to construct your data-driven narrative. You paid attention to your visualisation plots, graphs and explained the reasoning behind these visuals well.	**7 to >3.0 Pts** **Good** You made use of basic visualisations but failed to provide a strong justification for your choices. You did not take into consideration colour schemes, symbols, etc., and how these may affect your narrative.	**3 to >0.0 Pts** **Poor** You only used basic out-of-the-box plot functions.	**0 Pts** **No marks** No data introduction and exploratory data analysis were provided.	10 pts
Method implementation, discussion and elaboration of the results Describe the method used in the analysis, why it is used, how the method is implemented and any challenges encountered in this process. Discuss whether the method is unsupervised, supervised, etc. Evaluation of the model and the results obtained. Discuss what your findings mean. This should be achieved using a narrative instead of just reporting your results, findings. Compare your findings and the evaluation of the method(s) with the results in the literature using peer-reviewed papers.	**15 to >10.0 Pts** **Excellent** You have given a detailed explanation of the method used. You have discussed what metrics are used to assess the models. The model is efficient and clearly tailored toward the task. You explored several approaches to reach the final model. You explained the selection and implementation processes well. You discussed your findings and what they mean. You presented your findings with plots and visualisations (where relevant). You compared your findings and the success of your method (or methods) with those published in the literature.	**10 to >5.0 Pts** **Good** Your model performs well on the data, but you did not provide any further insights in terms of model choice and data-preprocessing, augmentation where applicable. You discussed your findings briefly providing relevant key results only.	**5 to >0.0 Pts** **Poor** You did not demonstrate that you understood and implemented your model independently and thoroughly. The model did not perform the task and poor insights were provided. You reported your findings as numbers in tables only and did not elaborate on what they mean. You did not demonstrate how these findings compare to the literature.	**0 Pts** **No marks** The model does not perform the task outlined in the assignment and a reasoning behind this is not provided. You did not elaborate on your results at all.	15 pts
Quality of report writing The quality, style and flow of the produced report, including spelling, grammar, structure.	**10 to >7.0 Pts** **Excellent** The report is nicely set out and easy to follow. You used figures and references appropriately. Excellent use of English and the correct technical language.	**7 to >3.0 Pts** **Good** The report is readable, good English is used and appropriate references have been used. You have shown some figures to illustrate the arguments of the report.	**3 to >0.0 Pts** **Poor** The report is poorly written. The structure and logic of your report are not easy to follow. You have not used any figures.	**0 Pts** **No marks** The report is not complete and very poorly constructed. The use of generative AI tools for writing this report will result in zero marks for the whole assignment.	10 pts
					Total points: 50

Assignment brief

Rubric

- You will be eval
  - Implementation the chosen model, preprocesssing, model architecture, use of metrics and analytics.
  - Quality of analysis in report.
  - Critical evaluation of results and model as well as critical representation of the data.
  - Comparison to literature.
  - Writing quality including spelling, grammar and academic language.

You will lose marks according to this rubric if a specific criterion has not been appropriately addressed.

Individual project report					
Criteria	**Ratings**				**Pts**
Thoroughness of literature review Collect relevant and timely information about the topic of interest using peer-reviewed papers. …duce the research field to the reader. Discuss …tion and why research is being done in … briefly what methods are used …lved in collecting and …efore it its used in	**15 to >10.0 Pts** **Excellent** You have introduced the topic to the reader so that the rest of your report can be clearly understood by a non-expert. You briefly described innovative methods used in the field and elaborated on possible challenges involved in collecting and processing the data. You have clearly read, understood and used the findings of multiple papers in your literature review.	**10 to >5.0 Pts** **Good** You have used <3 papers as your source material. You introduced the topic but did not discuss the literature well.	**5 to >0.0 Pts** **Poor** You introduced the topic poorly and only used a single paper for the literature review.	**0 Pts** **No marks** You did not provide a literature review.	15 pts
… evalu… metrics, plot…	…sightful and well-justified. You made use of …rrelation matrices, chord diagrams, maps, etc. …tions in order to construct your data- …n plots, graphs and explained the	**7 to >3.0 Pts** **Good** You made use of basic visualisations but failed to provide a strong justification for your choices. You did not take into consideration colour schemes, symbols, etc., and how these may affect your narrative.	**3 to >0.0 Pts** **Poor** You only used basic out-of-the-box plot functions.	**0 Pts** **No marks** No data introduction and exploratory data analysis were provided.	10 pts
Method implementation, discuss… elaboration of the results Describe the method used in the analysis, why i… used, how the method is implemented and any challenges encountered in this process. Discuss whether the method is unsupervised, supervised, etc. Evaluation of the model and the results obtained. Discuss what your findings mean. This should be achieved using a narrative instead of just reporting your results, findings. Compare your findings and the evaluation of the method(s) with the results in the literature using peer-reviewed papers.	… efficien… approaches t… implementation pro… they mean. You presented … (where relevant). You compared y… method (or methods) with those publish…	… >5.0 Pts …s well on the …ide any	**5 to >0.0 Pts** **Poor** You did not demonstrate that you understood and implemented your model independently and thoroughly. The model did not perform the task and poor insights were provided. You reported …findings as numbers in tables only and did …te on what they mean. You did not …these findings compare to the	**0 Pts** **No marks** The model does not perform the task outlined in the assignment and a reasoning behind this is not provided. You did not elaborate on your results at all.	15 pts
Quality of report writing The quality, style and flow of the produced report, including spelling, grammar, structure.	**10 to >7.0 Pts** **Excellent** The report is nicely set out and easy to follow. You used figures and references appropriately. Excellent use of English and the correct technical language.	**7 to >3.0 Pts** **Good** The report is readable, good English is used a… appropriate references have been used. You have shown some figures to illustrate the arguments of the report.	…ritten. The …f your report are not e… You have not used any figures.	**0 Pts** **No marks** The report is not complete and very poorly constructed. The use of generative AI tools for writing this report will result in zero marks for the whole assignment.	10 pts
					Total points: 50

Assignment brief

Do’s and don’ts

DO’s

- Carefully read through the assignment brief and the grading rubric on Canvas.
- This assignment will take ~40 hrs (assuming you have worked through and understood lectures and tutorials) so plan ahead.
- Talking to each other about your assignment is encouraged.
- Check your writing using spell checkers such as Grammarly.
- Take advantage of Library SkillUp.

DON’TS

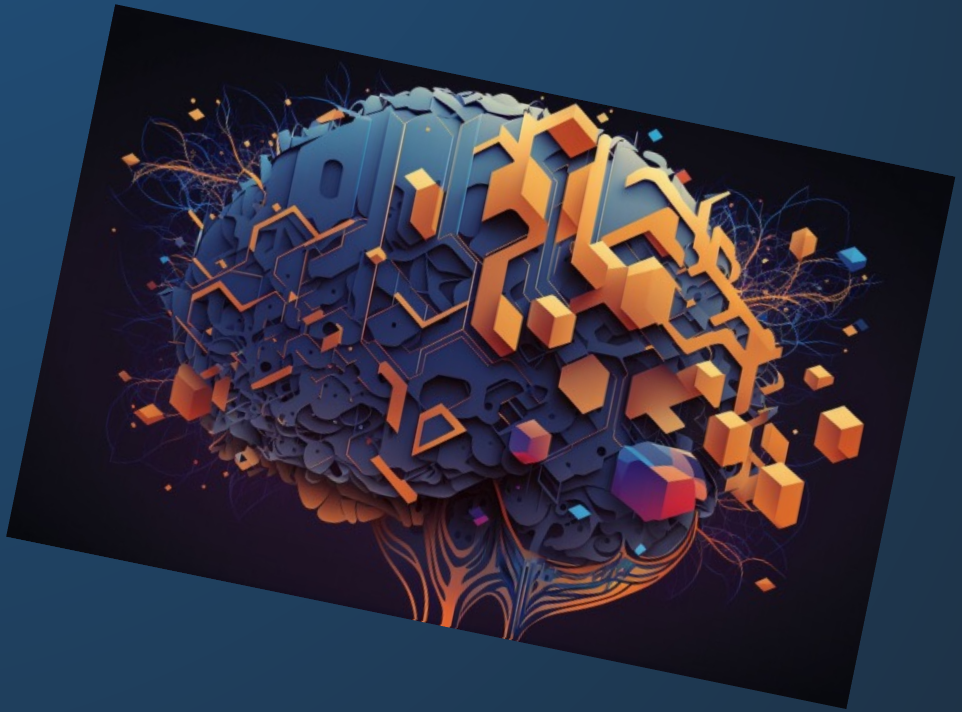
- For this individual assignment, your code and report will be unique.
- Do not plagiarise your code or report. Everything will be checked through Turnitin.
- Do not use generative AI tools like ChatGPT to write your assignment.
- You will get zeros for plagiarism collusion, and AI use, and repor-ted for academic misconduct.
- Don’t forget to cite images / tables.

Segmentation overview

Segmentation methods

Assignment brief

Summary



Further reading

Summary of several references discussed today

- Wang et al., *A Comprehensive Review of Modern Object Segmentation Approaches*, Foundations and Trends in Computer Graphics and Vision, 13, 111 (2022)
- Long et al., *Fully Convolutional Networks for Semantic Segmentation*, arXiv e-prints, arXiv: arXiv:1411.4038 (2014)
- Rupprecht et al., *Image Segmentation in 20 Questions*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3314 (2015)
- Ronneberger et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Medical Image Computing and Computer-Assisted Intervention, 234 (2015)
- Badrinarayanan et al., *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 2481 (2015)
- Chen et al., *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 99 (2016)

Summary

- Image segmentation is the field of computer vision that aims to break up images into pixel-by-pixel regions based on the objects' character.
- We want to answer ”Which pixels belong to which objects?” by **assigning specific segmentation masks**. This is the most complex aspect of the image recognition concepts discussed in our lectures.
- In general, segmentation approaches can be grouped into **semantic segmentation, instance segmentation and panoptic segmentation**.
- Traditional approaches for segmentation (threshold, edge, region or clustering based) have been outperformed by methods based on DL.
- DL methods specifically focus on **fully convolutional networks** (no dense layers) and employ an **encoder-decoder architecture** to up-sample feature maps that then capture global and local information.