METODE INTELIGENTE DE REZOLVARE A PROBLEMELOR REALE

Laura Dioşan

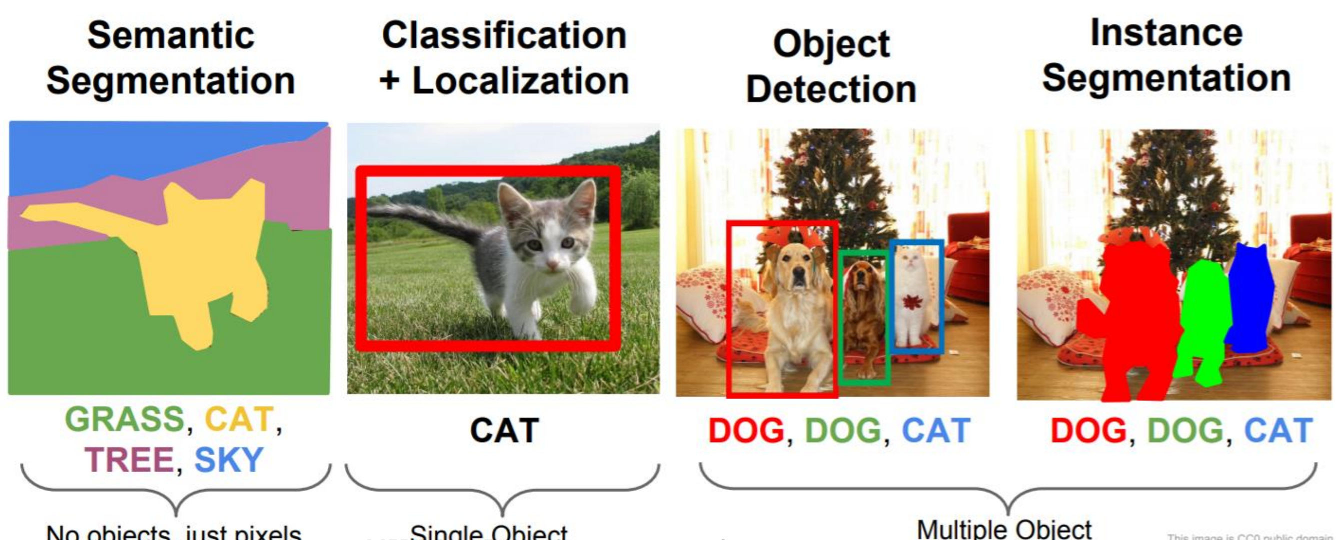
Image segmentation

Automatic image processing

- Image classification

CAT or DOG or OTHER?

- Other tasks



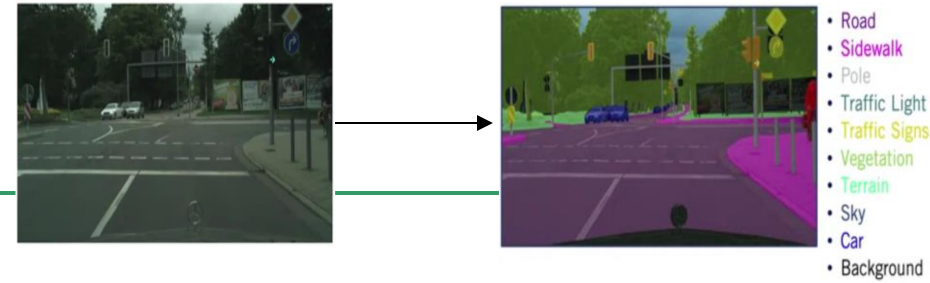
Automatic image processing

- Image classification
  - Does an image contain object X? [yes/no]

- Image detection and segmentation
  - Does an image contain object X? [yes/no]
  - Where is the object X? ➔ Location of the object
    - Pixel-based granularity ➔ semantic/instance segmentation
    - Object-based granularity ➔ object detection
  - Which object does this image contain? [where?]

  - Aprox. localisation (Bounding box)
  - Accurate localisation (contour) → Segmentation

Image segmentation

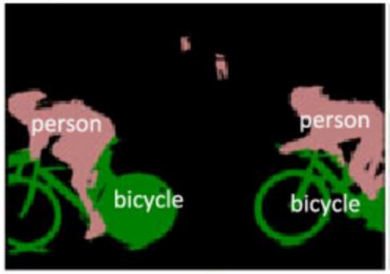


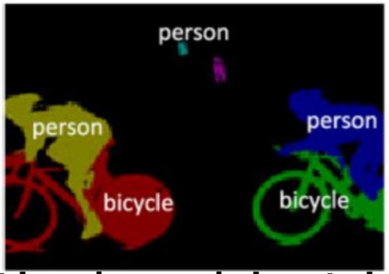
- Problem
  - Aim
    - Classify each pixel
  - Tasks
    - How many segments?
    - How many objects in an image?

Image segmentation

- Problem → Tasks

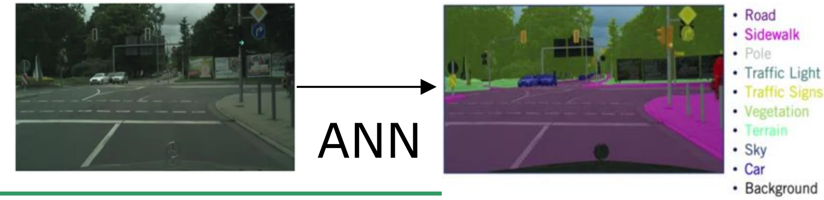
  - Semantic segmentation
    - Labels for every pixel
    - No differences across different instances of the same object





  - Instance segmentation
    - Labels for every pixel
    - unique label to every instance of a particular object in the image
    - Special topic: Panoptic segmentation
      - Instance segmentation for background

Image segmentation



- Problem
  - Challenges
    - Occlusion, Truncation, Scale, Illumination
    - Smooth boundaries
  - Evaluation
    - TP - #pixels correctly classified as belonging to class X
    - FP - #pixels classified as belonging to class X, but they belong to other classes
    - FN - #pixels that belong to class X, but are not classified as belonging to class X
    - IOU_{class} = TP / (TP + FP + FN) – over all images

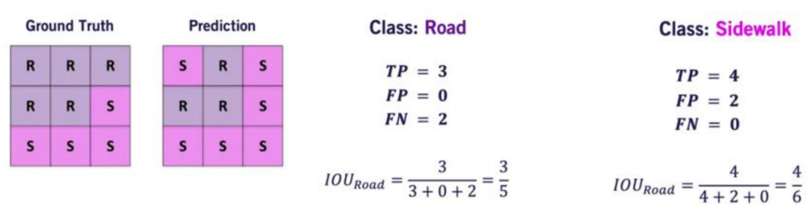
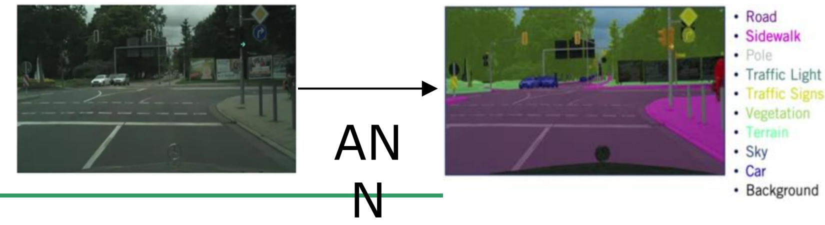


Image segmentation

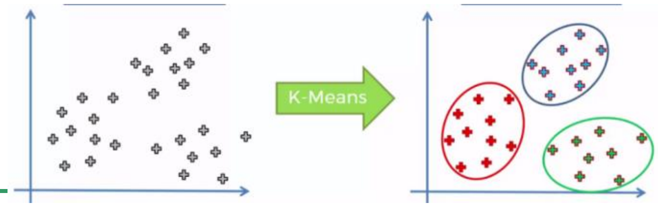


- Problem
  - Datasets
    - 2001 Berkeley
      - https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/
      - Good for edge detection problem (also)
    - 2005 – Pascal VOC
      - 20 classes
    - 2015 – COCO dataset (detection and segmentation)
      - https://cocodataset.org/#detection-2015
      - 91 classes
    - 2015 – CityScapes
      - https://www.cityscapes-dataset.com/
      - 30 classes grouped in 8 categories
    - CamVid
      - http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/

Image segmentation

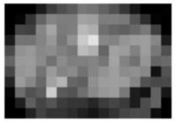
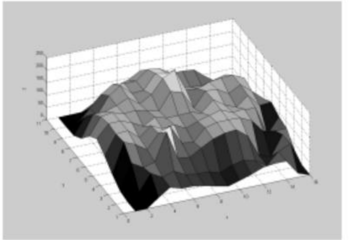
- How?
  - Before Computer Vision
    - Gestalt: whole or group
      - Whole is greater than sum of its parts
      - Relationships among parts can yield new properties/features
    - Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)
      - “I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have "327"? No. I have sky, house, and trees.” Max Wertheimer (1880-1943)
  - Computer Vision’s era
    - Segmentation as clustering (K-means, GAMMs and EM, Mean Shift, …)
    - Segmentation as grouping by boundaries
    - Graph-based segmentation
    - Segmentation as energy minimization
    - Region-based segmentation (->Thresholding, Region growing)
    - Edge detection segmentation
    - Deep learning algorithms

Image segmentation



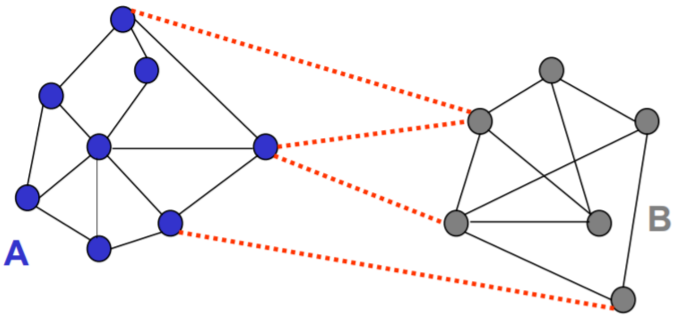
- How? ->Computer Vision’s era
  - Segmentation as clustering
    - Main idea
      - Group the “similar” pixels into clusters
    - Algorithms:
      - K-means, GAMMs and EM, Mean Shift, …
      - https://scikit-learn.org/stable/modules/clustering.html
    - See
      - Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(5), 603-619. https://courses.csail.mit.edu/6.869/handouts/PAMIMeanshift.pdf
      - http://cs229.stanford.edu/notes2020spring/cs229-notes8.pdf
      - http://cs229.stanford.edu/notes2020spring/cs229-notes7b.pdf
    - + works well on a small dataset with convex clusters
    - - large computational time, shape of clusters

Image segmentation

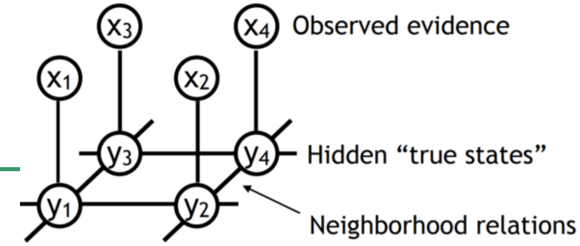
- How? ->Computer Vision's era
  - Segmentation as grouping by boundaries
    - Main idea
      - Edge-based methods
    - Algorithms:
      - Watershed – good for hierarchical segmentation
        - the image is regarded as a topographic landscape with ridges and valleys
      - Level-sets
    - See
      - https://members.accu.org/index.php/journals/1469
      - https://hub.gke2.mybinder.org/user/scikit-image-scikit-image-lpeqi3jb/notebooks/notebooks/auto_examples/segmentation/plot_watershed.ipynb
    - + Fast (apply filters)
    - - if there are too many edges or less contrast objects

Image segmentation



- How? ->Computer Vision's era
  - Graph-based segmentation
    - Main idea
      - Images as graphs (nodes – pixels, weights (affinity matrix) – location/intensity/color/textureFilters) and break graph in segments
    - Algorithms
      - Graph-Cut – eigen values of affinity matrix
      - Min-cut
    - See
      - http://cs.brown.edu/people/pfelzens/segment/
      - Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(8), 888-905. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=868688&casa_token=b23BGFlY2CwAAAAA:wbUB6ZhAc3vHP11Ii6cl2NyjfpI0vAHGefdvKegPJLacEiB332Xn0EnIF94R1qKk4MUdXgcFALPA&tag=1
    - + Flexible to choice of affinity matrix
    - + Generally works better than other methods
    - - Can be expensive, especially with many cuts.
    - - Bias toward balanced partitions
    - - Constrained by affinity matrix model

Image segmentation



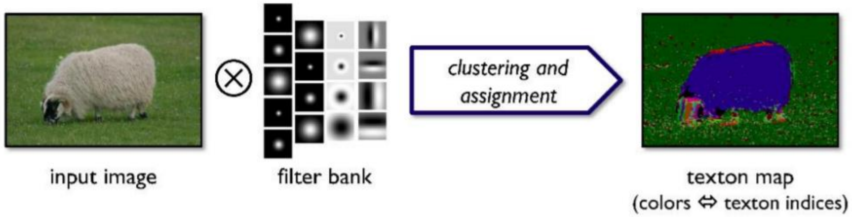
- How? ->Computer Vision's era
  - Segmentation as energy minimization
    - Main idea
      - Markov Random Fields (MRFs) and Conditional Random Fields (CRFs)
        - Rich probabilistic model for images
        - Built in local, modular way - Get global effects from only learning/modeling local ones
        - After conditioning, get a Markov Random Field (MRF)
    - Algorithms
      - Grab-Cut (2004)
    - See
      - Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, *23*(11), 1222-1239. http://luthuli.cs.uiuc.edu/~daf/courses/Opt-2017/Combinatorialpapers/00969114.pdf
    - \+ Very powerful, get global results by defining local interactions
    - \+ Very general
    - \+ Rather efficient
    - \- Only works for sub modular energy functions (binary)
    - \- Only approximate algorithms work for multi-label case

Image segmentation

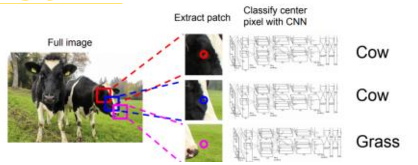
- How? ->Computer Vision's era
  - Region-based segmentation (low-level methods)
    - Main idea
      - rely mainly on the assumption that the neighboring pixels within one region have similar values.
    - Algorithms
      - Thresholding (Otsu's algorithm), Region growing (GrowCut)
    - See
      - https://im.snibgo.com/growcut.htm
    - + simple, fast,
    - - doesn't work if there are overlapped gray levels in image
  - Machine learning algorithms (high-level methods)
    - + simple, general
    - - high training time

Image segmentation

- How? ->Computer Vision's era
  - Machine learning algorithms
    - Before deep learning
      - CRF + pixels/superpixels



        - Jamie Shotton https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=03a26b7066269523698278314ebf1143a175072f
        - CRFs https://pub.ist.ac.at/~chl/papers/r
      - Sliding window



        - http://yann.lecun.com/exdb/publis/pdf/farabet-pami-13.pdf
        - https://ronan.collobert.com/pub/matos/2014_scene_icml.pdf
    - Deep learning era
      - Unet, Unet++, U2net &co
        - see https://causlayer.o
      - SegNet
      - DeepLab
      - FCN
      - DenseNet
      - …
      - Please check
        - https://github.com/mrgloom/awesome-semantic-segmentation

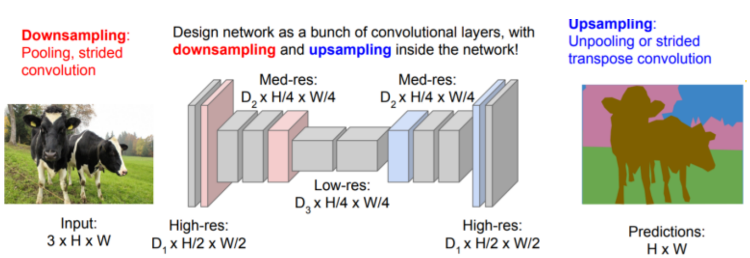
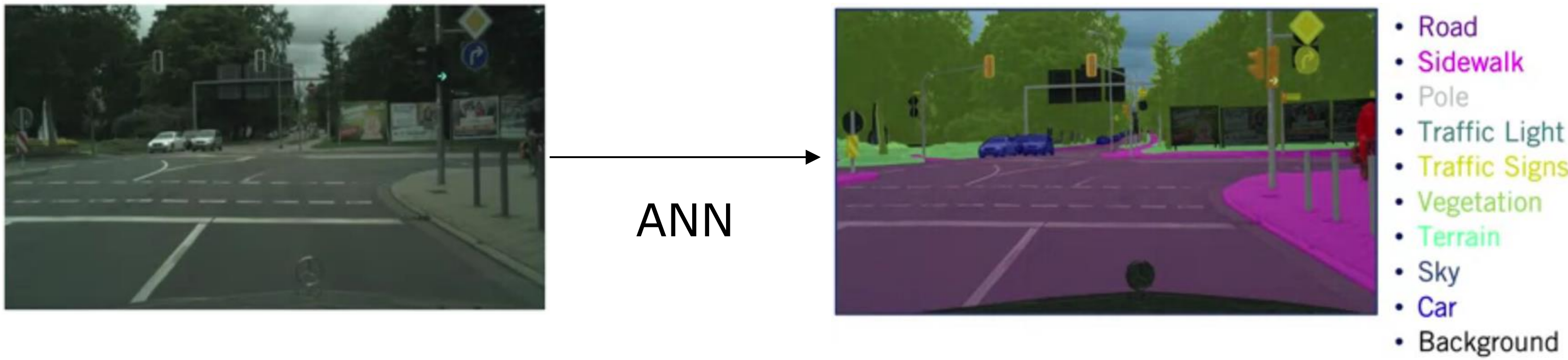


Image segmentation

- How? ->Computer Vision's era
  - Deep learning algorithms
    - Main idea
      - Extract features -> encoding
      - Decoding and classify pixels
    - Algorithms
      - Fully Convolutional Networks
      - Convolutional Networks with Graphical Models (CRFs and MRFs)
      - Multi-scale and Pyramid Network based models
    - + simple, general
    - - high training time

Image segmentation

- Problem (modern formulation)



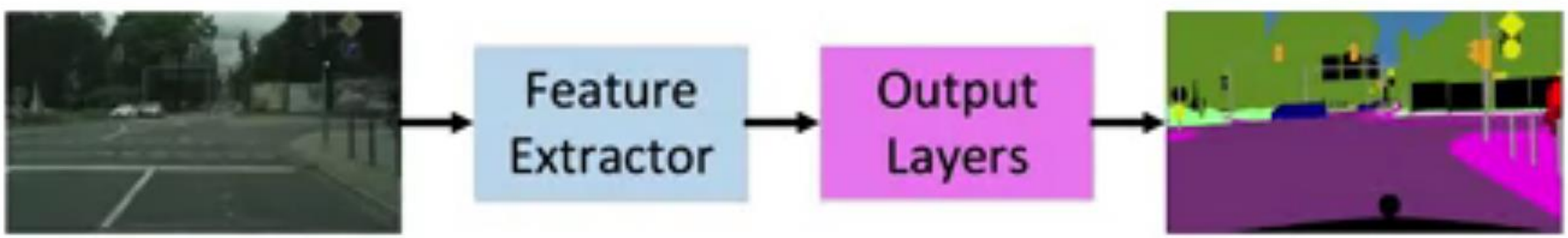
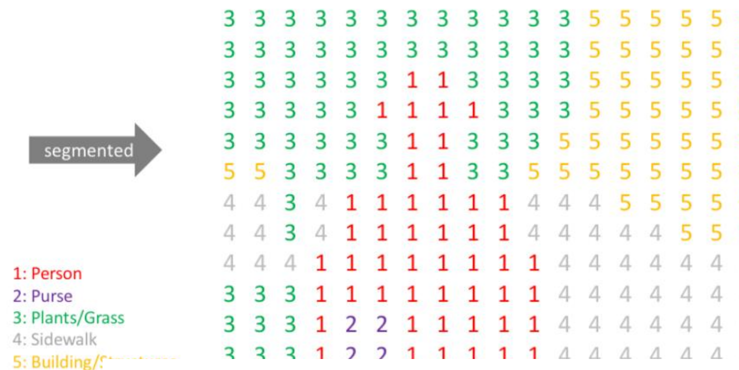


Image segmentation

- Semantic segmentation

Input



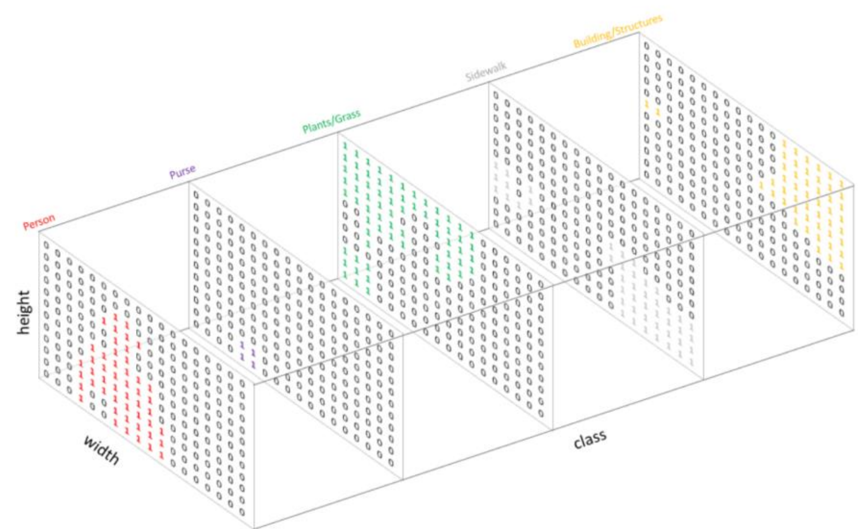


Image segmentation

- Problem -> feature extraction

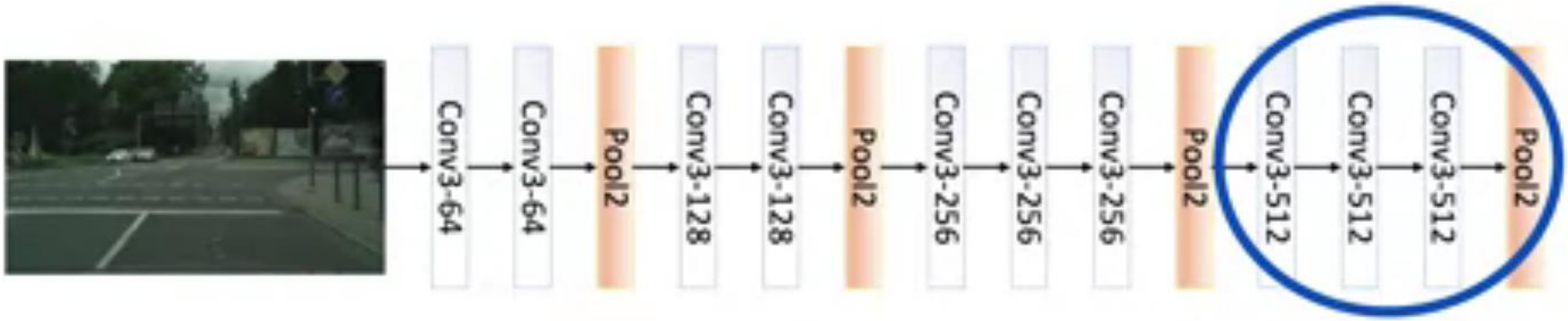
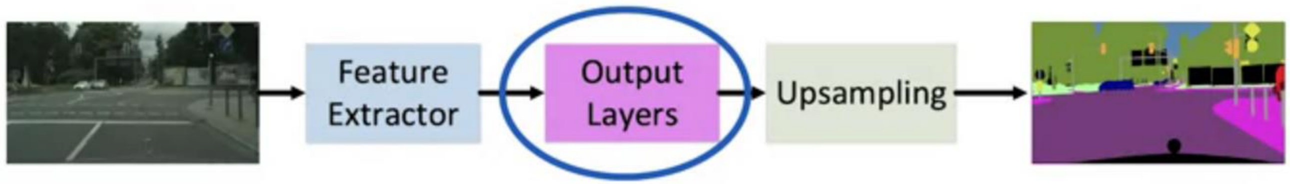


	Image	Conv1	Conv2	Conv3	Conv4
Width	M	M/2	M/4	M/8	M/16
Height	N	N/2	N/4	N/8	N/16
Depth	3	64	128	256	512

Image segmentation

- Up-sampling



- Learning same resolution feature maps

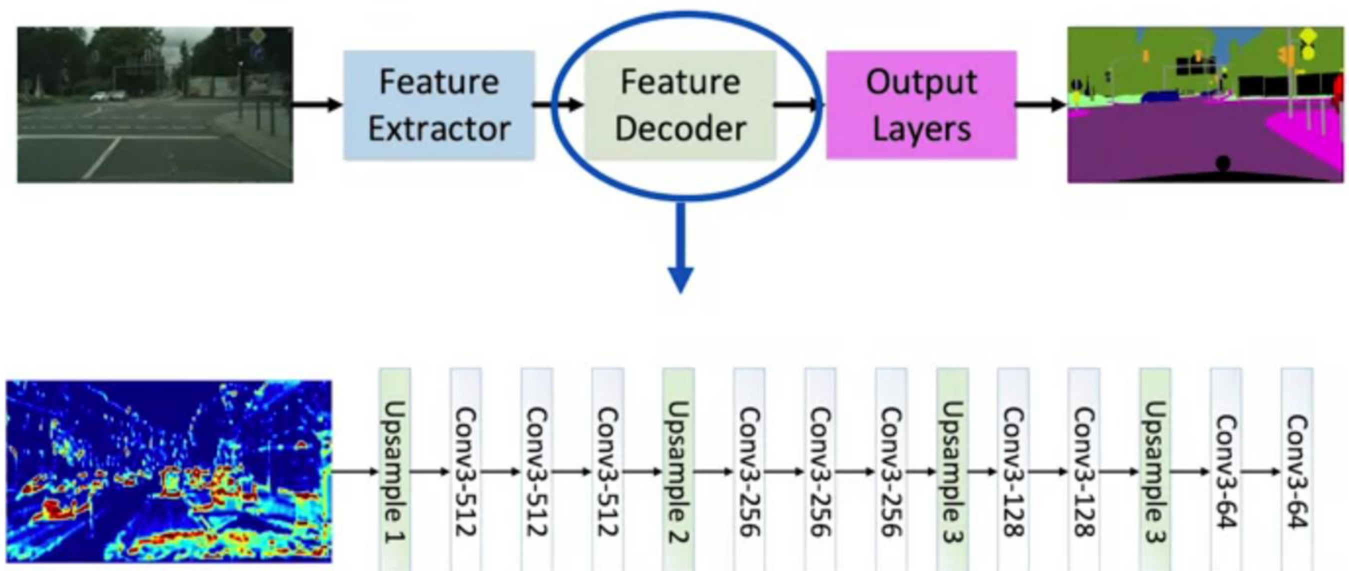
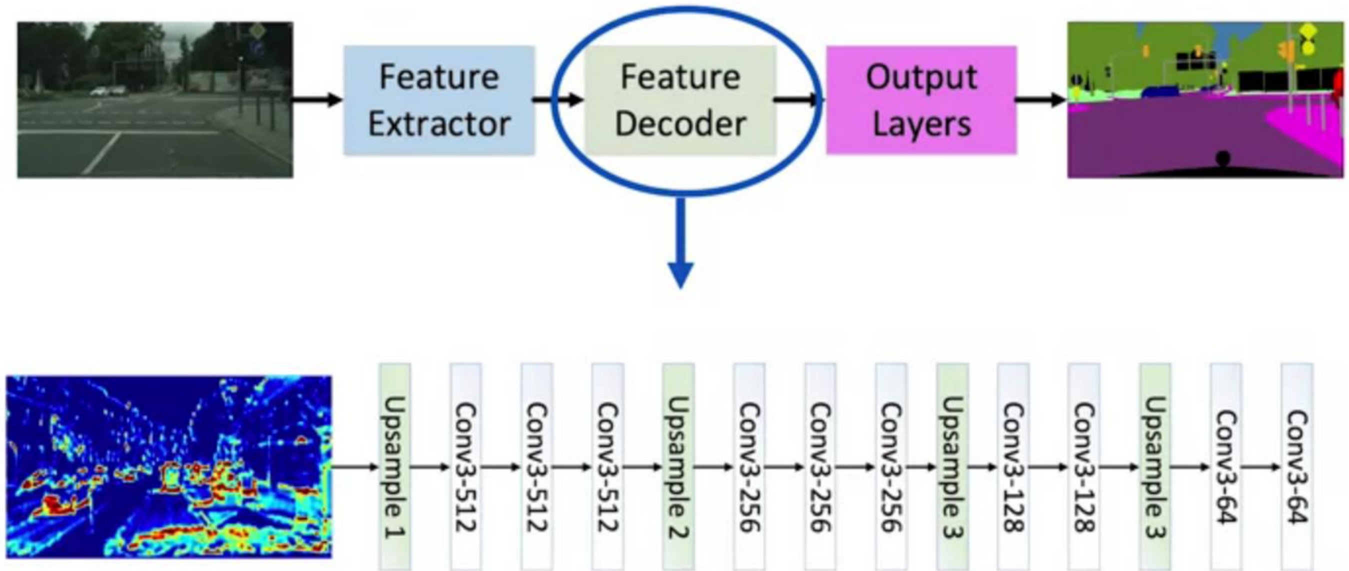


Image segmentation

- Learning same resolution feature maps



	Feature Map	Deconv1	Deconv2	Deconv3	Deconv4
Width	M/16	M/8	M/4	M/2	M
Height	N/16	N/8	N/4	N/2	N
Depth	512	512	256	128	64

Image segmentation

- Output computation

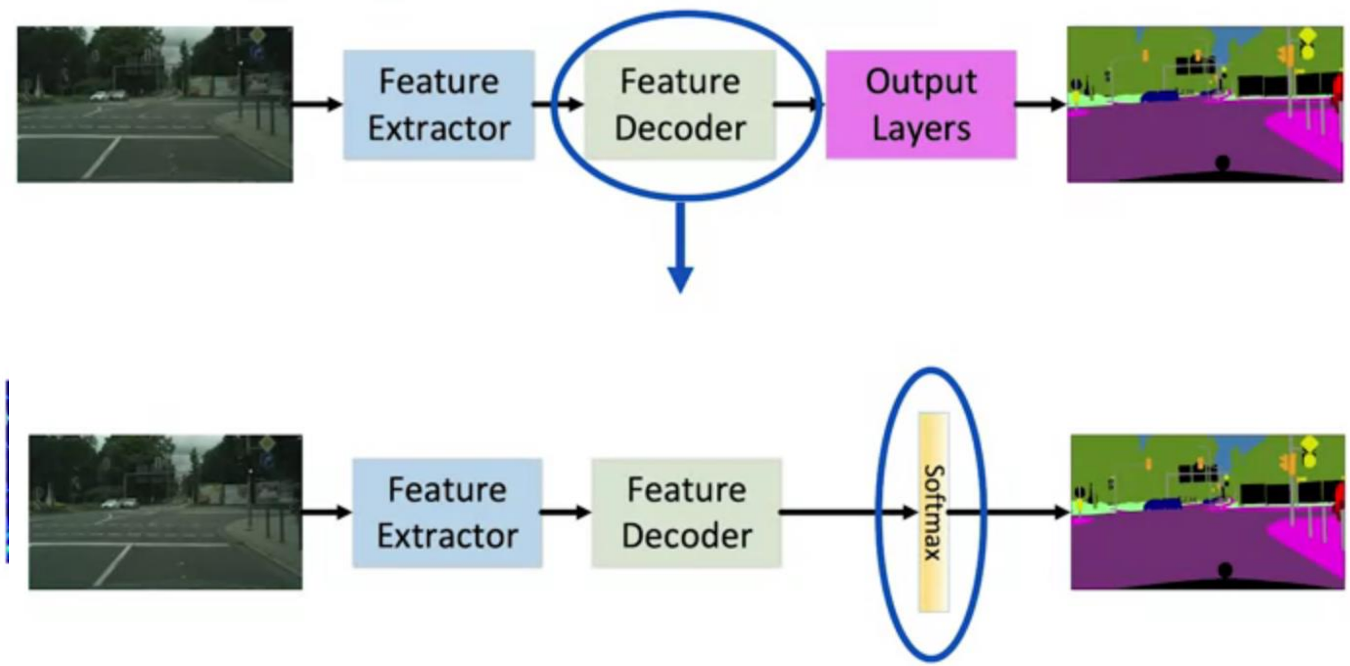


Image segmentation

- Fully convolutional networks
  - Feature extraction by convolutions (down-sampling / encoder path)
    - Extract and interpret the context (what?)
  - Segmentation map by recovering spatial information by convolutions (up-sampling / decoder part)
    - Enable precise location (where?)
    - Transform FC layers from a classification architecture into 1/more convolutions (deconvolutions or transposed convolutions) -> up-sampling
  - Skip connections
    - Recover the fine-grained spatial information lost in pooling or down-sampling layers
    - Merge (concatenate or sum) more feature maps from the down-sampling path with feature maps from the up-sampling path
      - Helps combining context information with spatial information

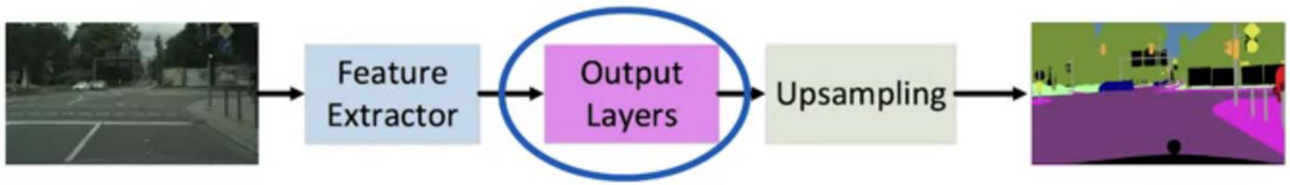


Image segmentation

- Fully convolutional networks
  - Most common architectures
    - FCN
      - https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf
      - Non-symmetric paths
        - #feature maps = #classes (down-sampling)
        - Up-samples only once (one layer for decoding + bilinear interpolation)
      - Skip connections by sum

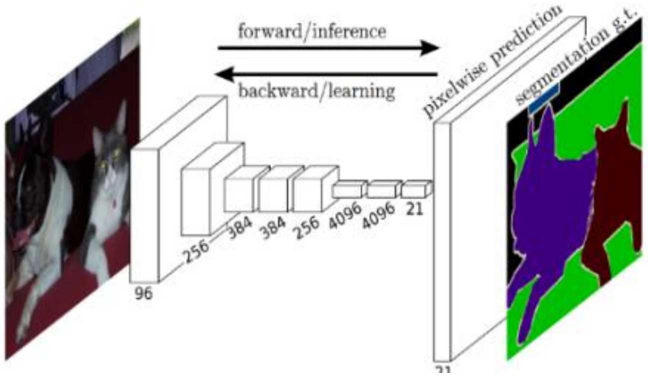
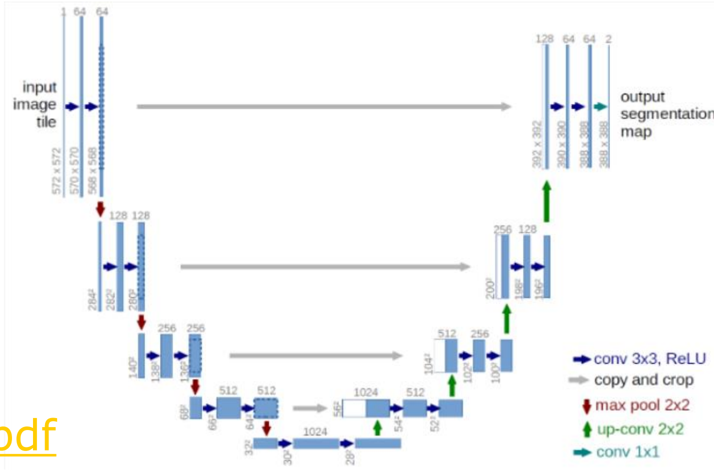


Image segmentation

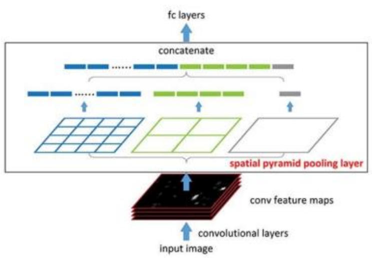
- Fully convolutional networks
  - Most common architectures
    - Unet
      - https://arxiv.org/pdf/1505.04597.pdf
      - a symmetric architecture
        - Larger #feature maps
        - Multiple up-sampling layers (=> learnable weight filters for interpolation)
      - Skip connections by concatenation

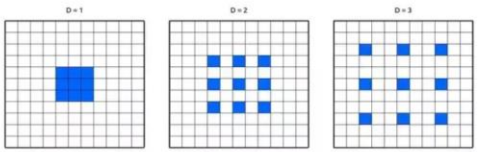


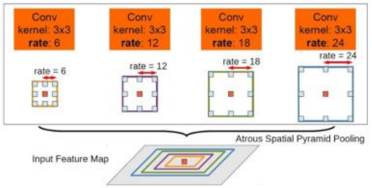
    - SegNet
      - https://arxiv.org/pdf/1511.00561.pdf
      - Similar to Unet

Image segmentation

- Fully convolutional networks
  - Most common architectures
    - DeepLab
      - https://github.com/tensorflow/models/tree/master/research/deeplab
      - New elements
        - Spatial pyramid pooling
        - Dilated (atrous) convolutions
        - Depthwise separable convolutions
        - Improving outputs with CRF







    - DenseNet and many others
      - https://github.com/mrgloom/awesome-semantic-segmentation

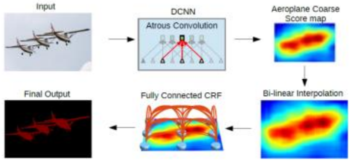


Image segmentation

- Unet / SegNet vs DeepLab

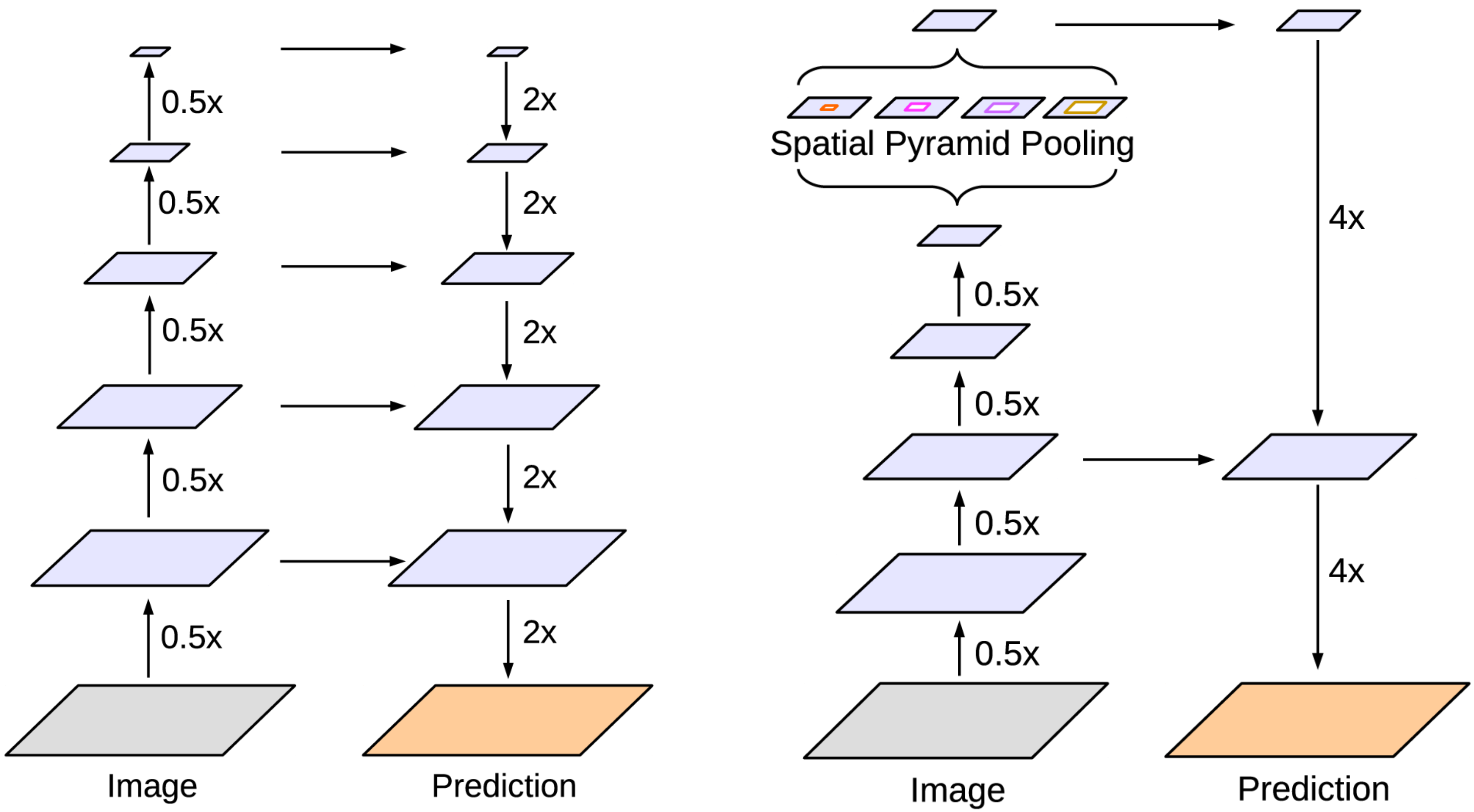
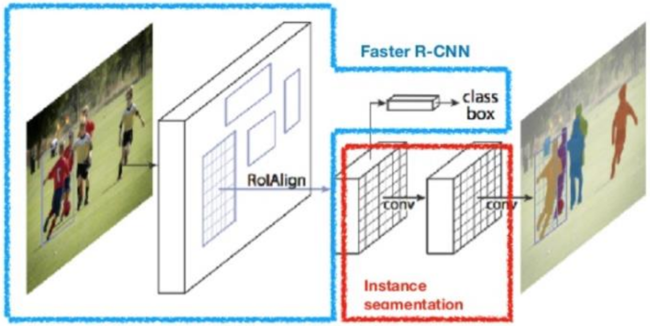
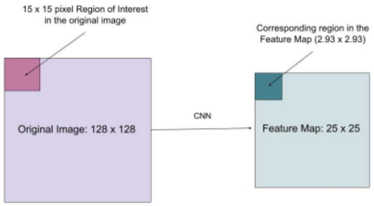


Image segmentation



- Mask R-CNN
  - https://arxiv.org/pdf/1703.06870.pdf
  - Similar to Faster R-CNN, but predict masks as well as BBs
    - a Fully CNN (on top of feature map) for determining a binary mask (object or not) for each RoI
    - RoI Alignment -> bilinear interpolation



    - Loss = Loss(classific) + Loss(bb) + Loss(mask)
      - Loss(mask) = cross-entropy

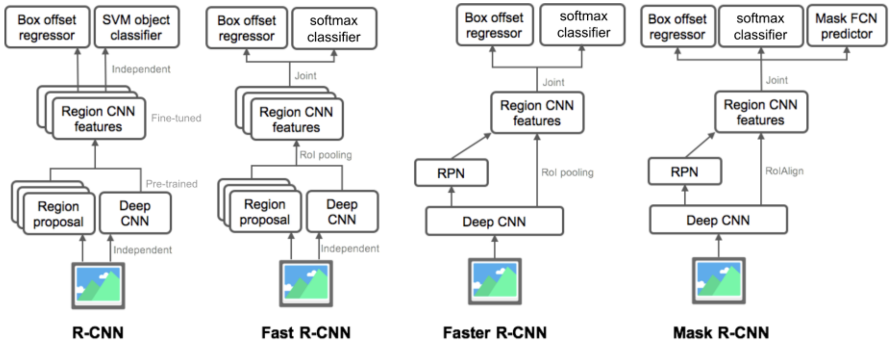


Image segmentation

- Detectron
  - Feature extraction
    - Feature pyramid network
    - Different backbones (ResNet)
  - Proposal generator
    - Region proposal network
  - Target tasks
    - BB prediction
    - BB classification
    - Pixel-level classification inside a BB (segmentation)
  - Loss = Loss(classific) + Loss(bb) + Loss(mask)
    - Loss(mask) = cross-entropy
    - Focal loss
  - Non-local NN https://arxiv.org/pdf/1711.07971.pdf
    - Long-range dependencies
      - Recurrent operations (repeated convolutions = local neighbourhood)
      - Non-local operations
        - Mean of all positions of an input =a very large receptive field
        - Self-attention (machine translation)
        - CRF (graphical models)

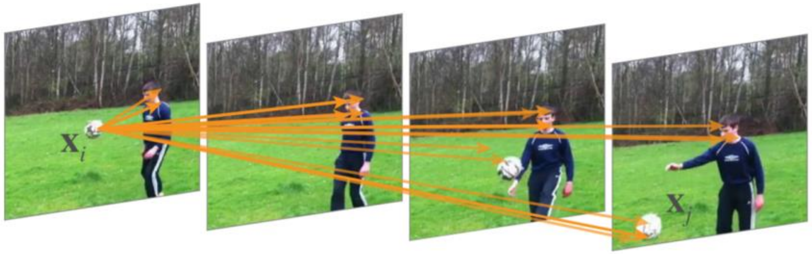


Image segmentation

- **YOLACT (You Only Look At CoefficienTs)**
  - https://github.com/dbolya/yolact
  - https://arxiv.org/pdf/1904.02689.pdf

- Vision Transformers (ViT)
  - reducing architecture complexity
  - exploring scalability and training efficiency
  - **An Image is Worth 16x16 Words**
    - https://arxiv.org/pdf/2010.11929.pdf
    - https://ai.facebook.com/research/publications/end-to-end-object-detection-with-transformers
    - **NLP transformers** http://jalammar.github.io/illustrated-transformer/
  - https://github.com/google-research/vision_transformer

Image segmentation

- More details
  - https://arxiv.org/pdf/2001.05566.pdf
  - https://heartbeat.fritz.ai/a-2019-guide-to-semantic-segmentation-ca8242f5a7fc
  - https://paperswithcode.com/sota/instance-segmentation-on-coco
  - https://link.springer.com/article/10.1007/s13735-020-00195-x

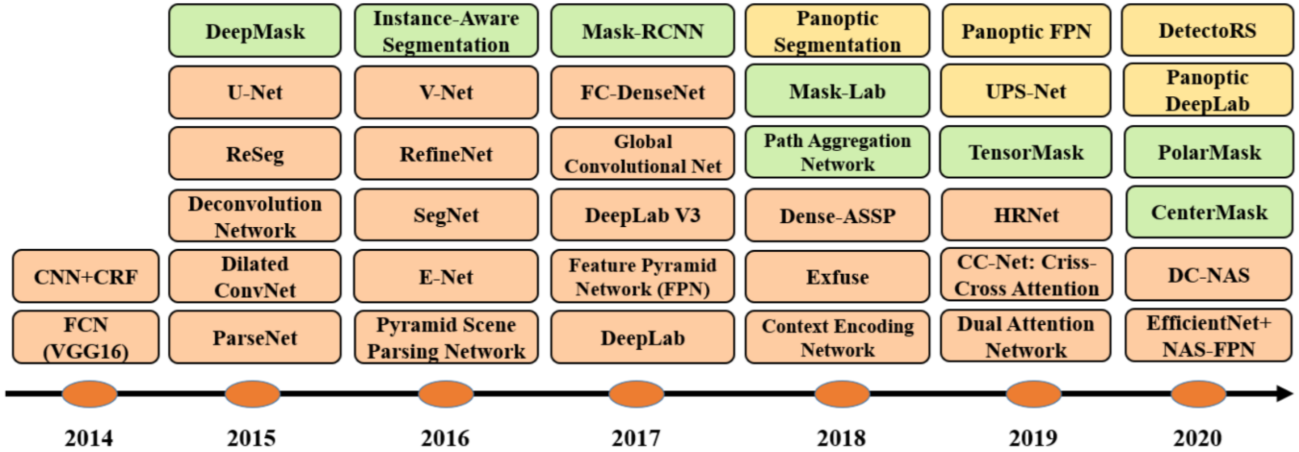


Fig. 32. The timeline of DL-based segmentation algorithms for 2D images, from 2014 to 2020. Orange, green, andn yellow blocks refer to semantic, instance, and panoptic segmentation algorithms respectively.

Image segmentation

- Segmentation
  - partitioning an image into meaningful segments, which share a common representation.
  - Dense pixel prediction -> it classifies each pixel into one of a few classes

- Semantic segmentation
  - Segment all interest objects (by different classes = semantic classes)

- Instance segmentation
  - Segment all interest objects (by different classes = semantic classes)
  - Predict an instance label for each object of interest

- Panoptic segmentation
  - Instance segmentation of all interest objects (by classes) and the background