



Vilniaus universitetas
Matematikos ir informatikos fakultetas
Informatikos katedra

Objektų vaizduose atpažinimas ir vaizdų segmentavimas taikant konvoliucinius neuroninius tinklus

prof. dr. Olga Kurasova
Olga.Kurasova@mii.vu.lt

2018

Kompiuterinės regos (*computer vision*) uždaviniai

Classification



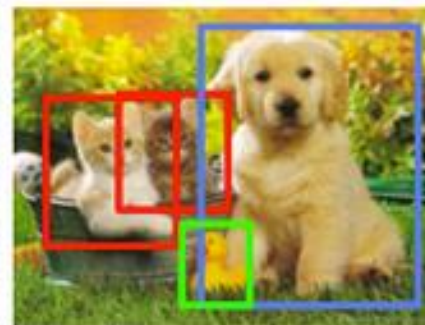
CAT

**Classification
+ Localization**



CAT

Object Detection



CAT, DOG, DUCK

**Instance
Segmentation**

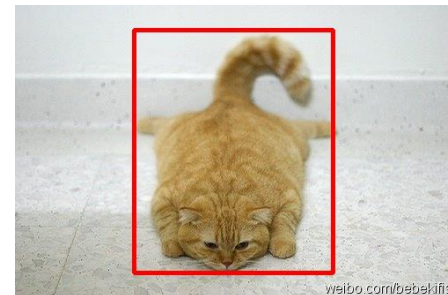


CAT, DOG, DUCK

<https://mathematica.stackexchange.com/questions/141598/object-detection-and-localization-using-neural-network>

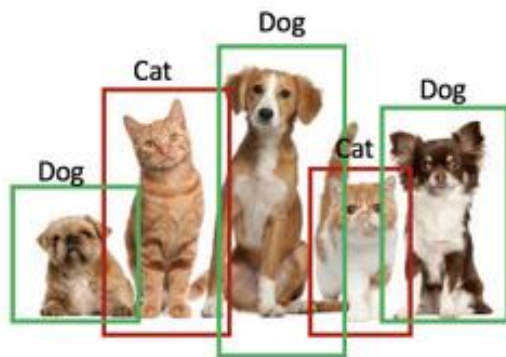
Objektų lokalizavimas

- Dažnai objektams lokalizuoti taikomo **iš anksto apmokyti tinklai (modeliai)** (*pre-trained*) (YOLO, R-CNN ir kiti).
- Todėl lokalizavimas **greitas** ne tik statiniams vaizdams, bet ir **video**.
- Tiek mokymo duomenyse, tiek gautuose tinklo išvestyse yra **ne klasė** (kaip yra klasifikavimo uždaviniuose), bet objektą apibrėžiantis **stačiakampis** (*bounding box*)



Objektų lokalizavimas ir klasifikavimas

- Dažnai reikia ne tik nustatyti objekto vietą, bet ir **identifikuoti**, koks tai objektas.
- Vaizde gali būti **keli** to paties tipo objektai.
- Rezultate gaunamas masyvas, kuriame nurodytos **stačiakampio parametrai** (dažnai centro koordinatės, ilgis ir plotis) ir **klasių tikimybės**.



<https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad>

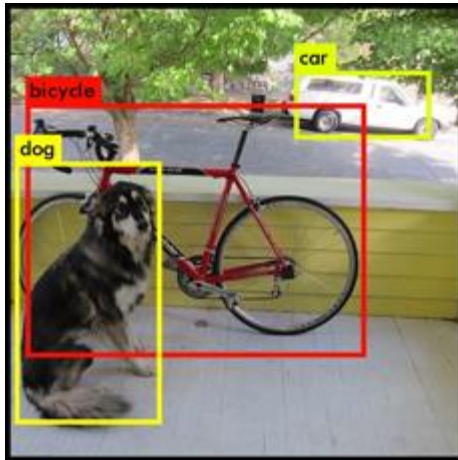
Objektų lokalizavimas ir klasifikavimas

- Naudojant iš **anksto apmokyta** tinklą (jis atpažįsta norimus objektus), per vaizdą **slenkamas langas** ir tikrinama, ar tinklas atpažįsta jame **kažkokį** objektą – *sliding window detection*
- Greitesnis yra **YOLO** algoritmas (You Only Look Once) – čia visas langas sudalinamas tinkleliu ir ieškomi objektai tinklelio langeliuose.

<https://www.coursera.org/lecture/deep-learning-in-computer-vision/sliding-windows-UlbVI>

<https://tryolabs.com/blog/2017/08/30/object-detection-an-overview-in-the-age-of-deep-learning/>

YOLO (You Only Look Once:)



Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

<https://pjreddie.com/darknet/yolo/>

<http://machinethink.net/blog/object-detection-with-yolo/>

<https://hackernoon.com/gentle-guide-on-how-yolo-object-localization-works-with-keras-part-1-aec99277f56f>

R-CNN

R-CNN ([Girshick et al., 2014](#)) is short for “Region-based Convolutional Neural Networks”. The main idea is composed of two steps. First, using [selective search](#), it identifies a manageable number of bounding-box object region candidates (“region of interest” or “RoI”). And then it extracts CNN features from each region independently for classification.

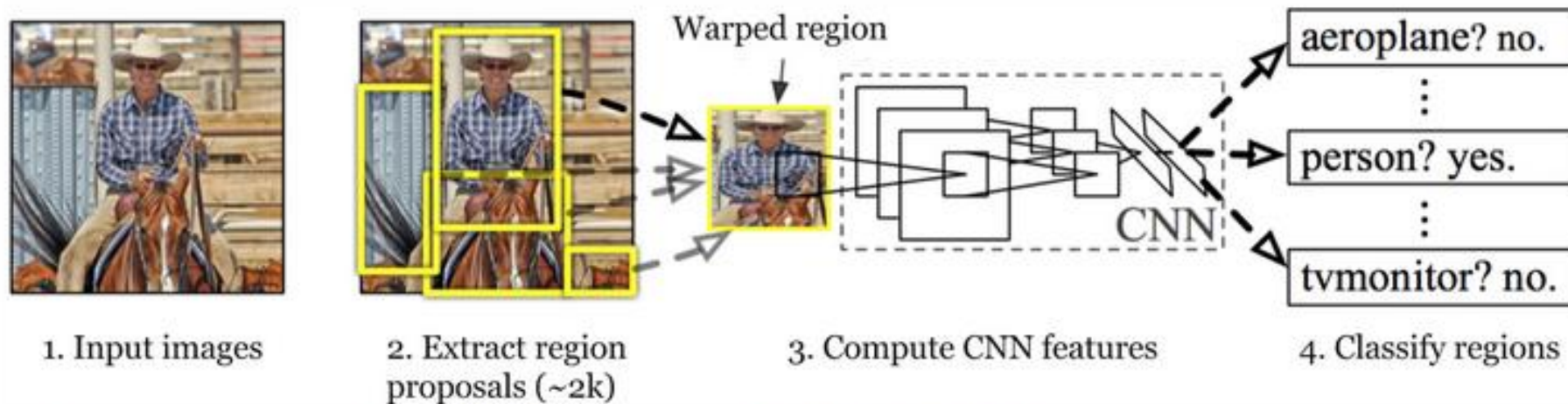


Fig. 1. The architecture of R-CNN. (Image source: [Girshick et al., 2014](#))

Objektų lokalizavimas ir klasifikavimas: taikymai

- **Veidų** lokalizavimas
- Žmogaus (automobilio ir kt.) **lokalizavimas** ir **sekimas** (pvz., norint suskaičiuoti šiuos objektus)
- Žmogaus atliekamų **veiksmų identifikavimas**
- **Medicininį** vaizdų analizė (nors čia svarbiau segmentavimas, žr. tolesnėse skaidrėse)
- **Satelitinių** (arba **dronų** darytų) vaizdų analizė
- Kiti

Objektų lokalizavimas

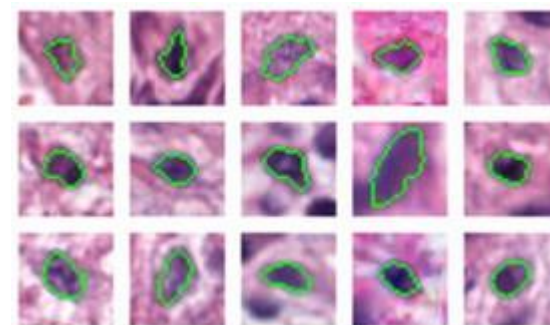
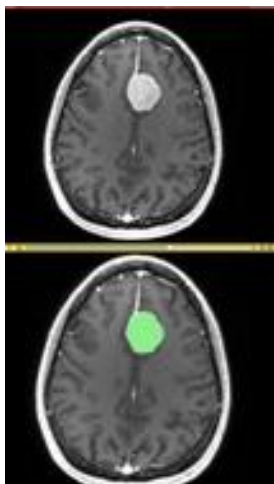
Įdomūs demo:

- <https://jeeliz.com/>
- <https://js.tensorflow.org/>
- https://trackingjs.com/examples/face_camera.html

Vaizdų segmentavimas

- **Vaizdų segmentavimas** (*image segmentation*) – tai procesas, kurio metu skaitmeninis vaizdas padalinamas į tam tikrus segmentus (dalis).
- Dažnai svarbu išskirti objektus ir jų ribas.
- Segmentavimas ypač svarbus **medicininiuose vaizduose**, kai siekiama nustatyti ligos pažeistus regionus.
- Vėliau (arba ir kartu) gali būti sprendžiamas **atpažinimo uždavinys**.

Vaizdų segmentavimo pavyzdžiai



Vaizdų segmentavimo taikymai

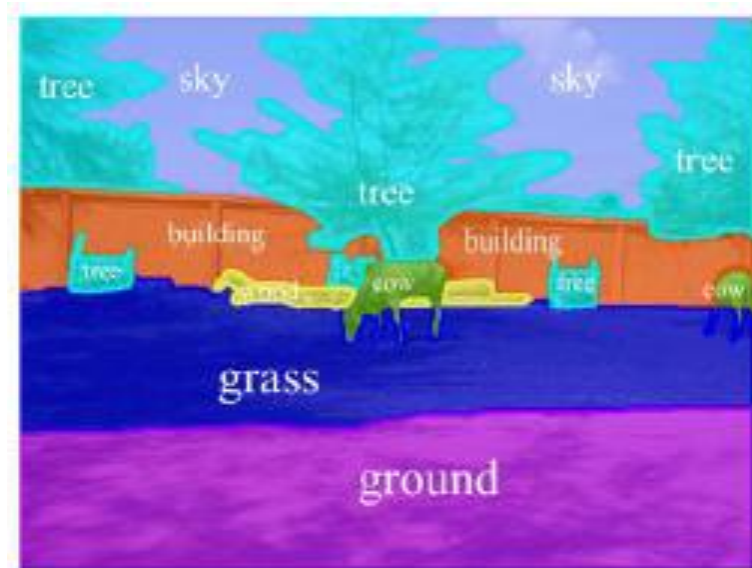
- **Medicininųjų vaizdų** analizė (kompiuterinė tomografija, magnetinis rezonansas)
- **Objektų** nustatymas
 - Pėsčiųjų nustatymas, veido nustatymas
- **Atpažinimo** uždaviniai
 - Veido atpažinimas, pirštų anspaudų atpažinimas, akies rainelės atpažinimas
- **Eismo kontrolės** sistemos (autonominiai automobiliai)
- **Mikroskopijos** uždaviniai

Vaizdų segmentavimo metodai

- Otsu metodas
- K-vidurkių metodas
- Histogramomis pagrįsti metodai
- Kraštinių (briaunų) nustatymo (*edge detection*) metodai
- SIFT metodas
- Įvairūs kiti vaizdų apdorojimo metodai
- **Dirbtiniai neuroniniai tinklai**

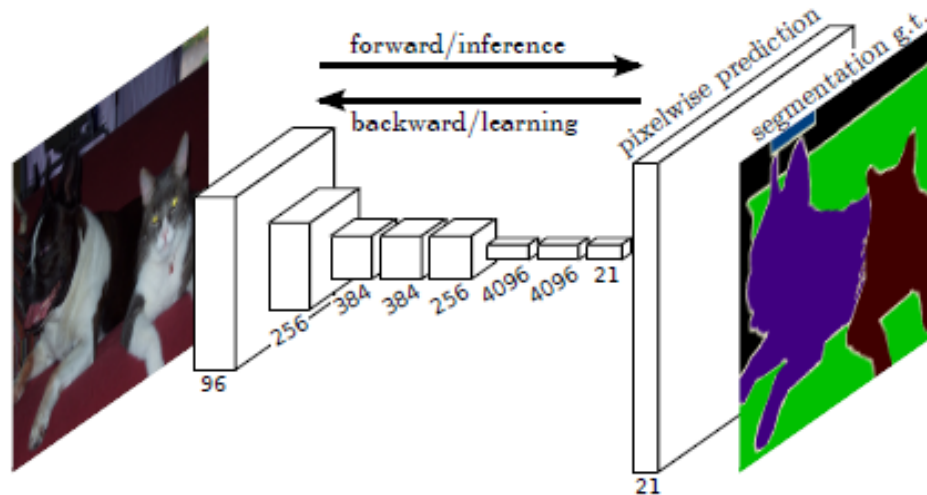
Semantinis segmentavimas

- Segmentavimo uždavinys, kai reikia ne tik nustatyti objektus, bet juos **priskirti ir klasės** (pvz., vanduo, automobilis, dangus, pastatas), vadinamas **semantinio segmentavimo uždaviniu**.



Pilnai konvoliucinis tinklas

- **Pilnai konvoliucinio tinklo** (*Fully Convolutional Networks, FCN*) tinklas – numatyti kiekvieno vaizdo pikselio reikšmę (*pixels-to-pixels*)



Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Pilnai konvoliucinis tinklas

- **Pilnai konvoliucinis tinklas** sudarytas iš:
 - Konvoliucinių sluoksnių
 - Sujungimo sluoksnių (*pooling*)
 - Atstatymo sluoksnių (*upsampling*)
- Čia nėra **pilnai sujungto sluoksnio** (*fully connected (dense) layers*).
- Išėjimo sluoksnio dydis **sutampa** su įvesties dydžiu.

Pilnai konvoliucinis tinklas

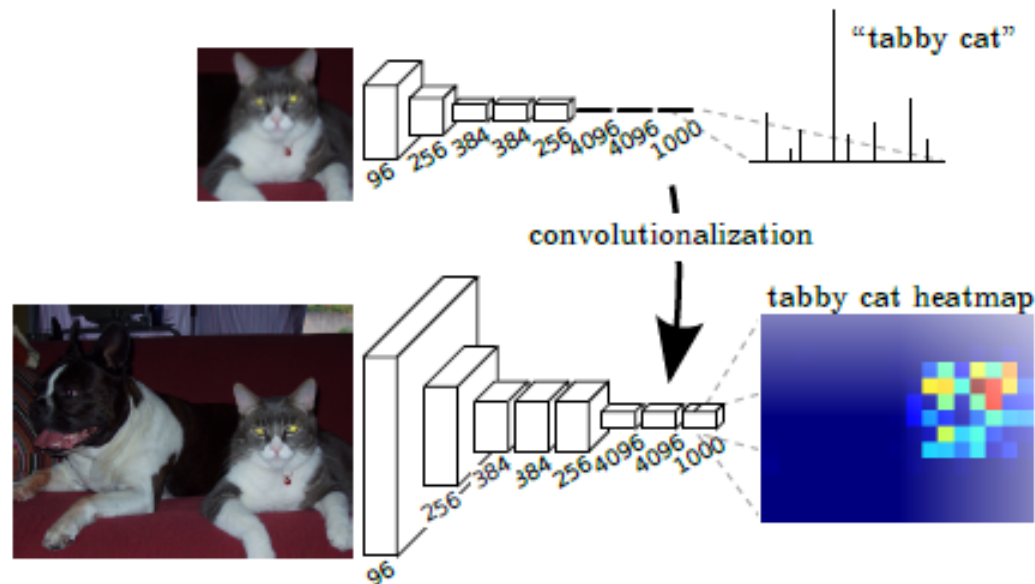


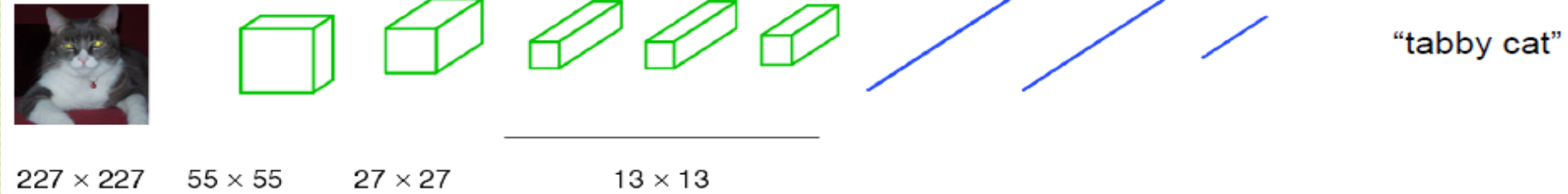
Figure 2. Transforming fully connected layers into convolution layers enables a classification net to output a heatmap. Adding layers and a spatial loss (as in Figure 1) produces an efficient machine for end-to-end dense learning.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

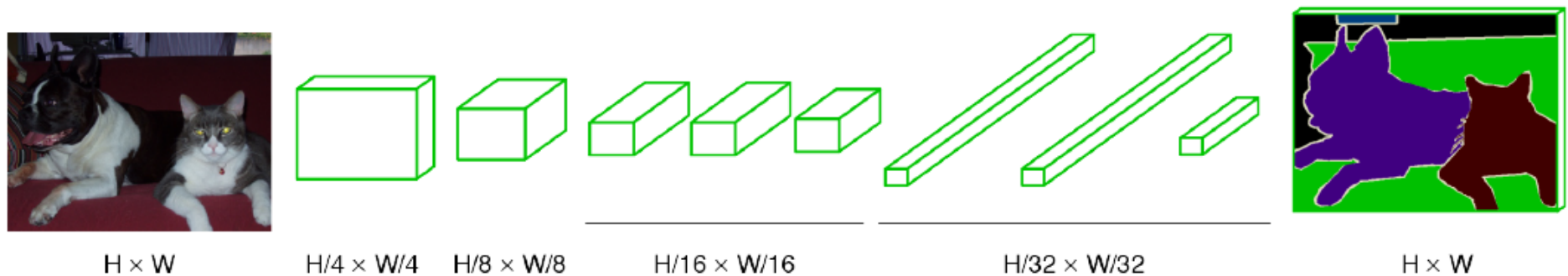
Pilnai konvoliucinis tinklas

convolution

fully connected



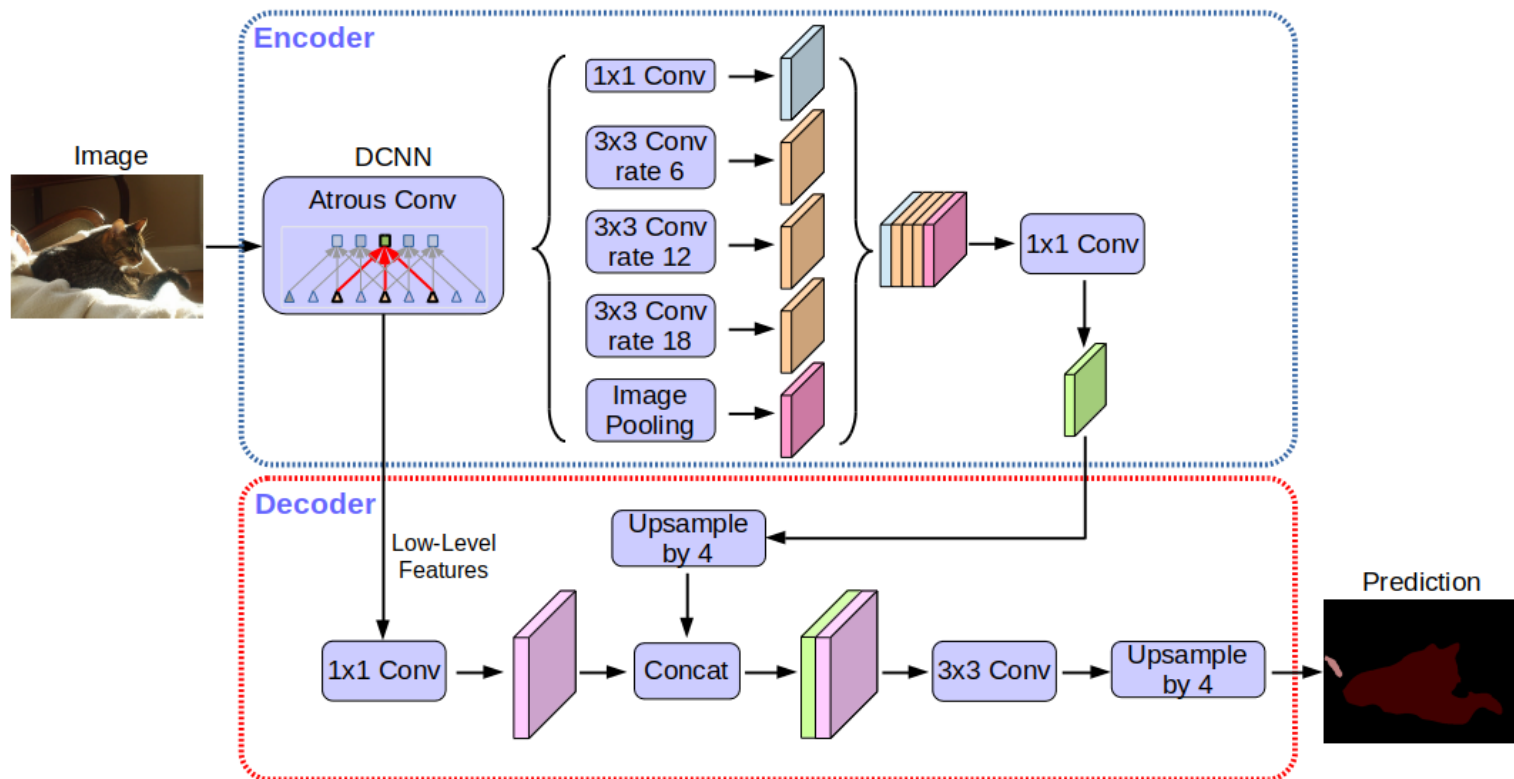
convolution



https://leonardoaraujosantos.gitbooks.io/artificial-intelligence/content/image_segmentation.html

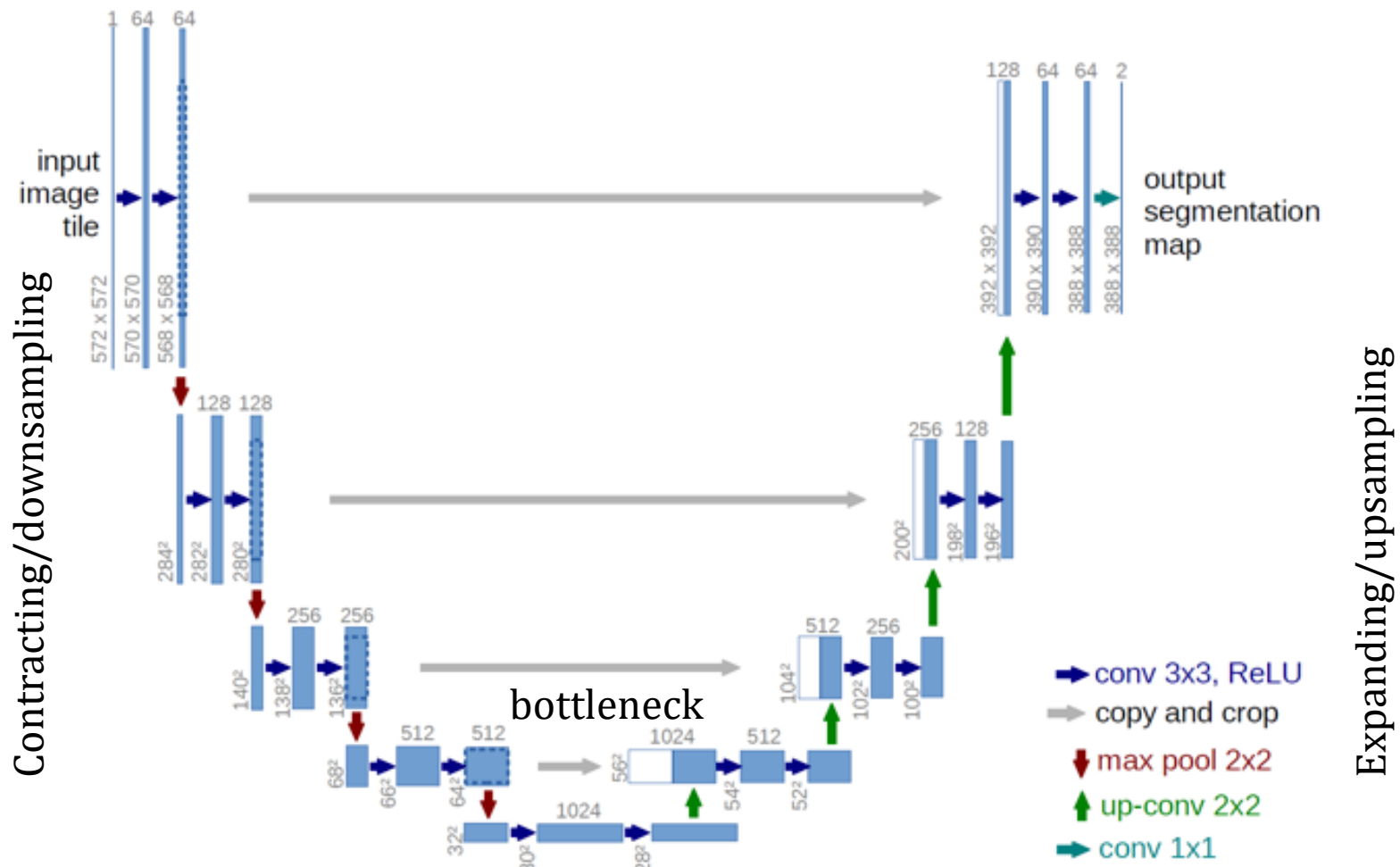
Enkoderis - dekoderis

- Kito tipo tinklai veikia enkoderis – dekoderis principu.



<https://ai.googleblog.com/2018/03/semantic-image-segmentation-with.html>

U-net



Kodėl U-net?

Dėl tinklo formos, primenančios U raidę

U-net

- Tai **patobulintas** pilnai konvoliucinis tinklas.
- U-net – **simetrinis tinklas**.
- **Tikslas** – klasės nustatymas kiekvienam pikseliui.
- Patobulinimas – DeepUNet

Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., & Li, W. (2018). DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

U-net

- Ronneberger O., Fischer P., Brox T. **(2015) U-Net: Convolutional Networks for Biomedical Image Segmentation**. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-24574-4_28
- The **full implementation** (based on Caffe) and the trained networks are available at <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>

U-net

Table 2. Segmentation results (IOU) on the ISBI cell tracking challenge 2015.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	0.9203	0.7756

Results from (Ronneberger et al., 2015)

Kokie duomenys reikalingi?

Vaizdai

Tikri poligonai

Nustatyti poligonai

Input



Ground truth



Prediction



Kokie duomenys reikalingi?

- Taigi, reikia turėti ne tik vaizdus, bet ir juos **anotuoti**, t. y. sudaryti objektų poligonus atitinkančias **kaukes** (*mask*). Tai atliekama naudojantis grafikos redagavimo programomis.
- Šios kaukės bus tinklo **norimos** (**trokšamos**) reikšmės.
- Jei tai **kelių klasių** uždavinys, dar reikia nurodyti kurią klasę atitinka konkreti kaukė.

Mokymo duomenys

- Jei turime **didelius vaizdus** (pvz., satelitinius, medicininius), kurių kiekviename daug ieškomų objektų, tinklo **mokymo duomenys** paruošiami taip:
 - Į tinklo įvestį (input) pateikiamas ne visas vaizdas, tačiau **iškerpami jo fragmentai** (*patch sampling*), tokiu būdu stipriai padidinama mokymo aibė.
 - Iškirptiems objektams dar gali būti panaudojamos įvairios afininės transformacijos (pasukimai ir kt.), tai vadinama **duomenų papildymu** (*augmentation*)

Tikslumo matai

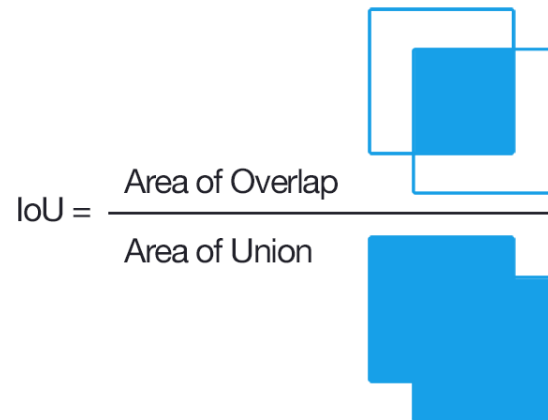
- **Tikslumas** pikselių lygmenyje

$$\text{acc}(P, GT) = \frac{\text{tiksliai nustatytų pikselių skaičius}}{\text{visų pikselių skaičius}}$$

- **Jaccard** metrika (*Intersection over Union*, IoU)

$$\text{jacc}(P(\text{class}), GT(\text{class})) = \frac{|P(\text{class}) \cap GT(\text{class})|}{|P(\text{class}) \cup GT(\text{class})|}$$

- čia P – nustatyta segmentavimo dalis,
GT – norima kaukė.



Dar apie U-net

Nors U-net sukurtas **medicininių** vaizdų segmentavimui, jis sėkmingai taikomas ir kitiems uždaviniams, pvz., **satelitinių** vaizdų segmentavimui



Apibendrinimas

- Liu, X., Deng, Z., & Yang, Y. (2018). Recent progress in semantic image segmentation. *Artificial Intelligence Review*, 1-18.
<https://link.springer.com/article/10.1007/s10462-018-9641-3>
- <https://www.jeremyjordan.me/semantic-segmentation/>