HELMHOLTZ AI Artificial Intelligence Cooperation Unit

TEACHING CONVOLUTIONAL NEURAL NETWORKS

Till Korten

With Material from

Johannes Müller, Robert Haase: PoL

Alex Krull, Uwe Schmidt: MPI CBG

Martin Weigert: EPFL Lausanne

Ignacio Arganda-Carreras: Universidad del Pais Vasco

December 2022

WHAT TO EXPECT

- 30 min presentation:
  - Teaching methods
  - Fundamentals of CNNs
- 60 min flipped classroom session
- 30 min break
- 60 min pair programming session
- 30 min feedback and discussion

CLASSICAL LECTURE

- Lecture given by teacher (this presentation)
- Lerners prepare exercises on their own
- Teaching assistants grade exercises
- Pro:
    - Familiar for teachers and learners
- Con:
    - One speed for all (too slow for some, too fast for others)
    - Once learners lose track, they remain lost
    - Wasted opportunity of being in a group

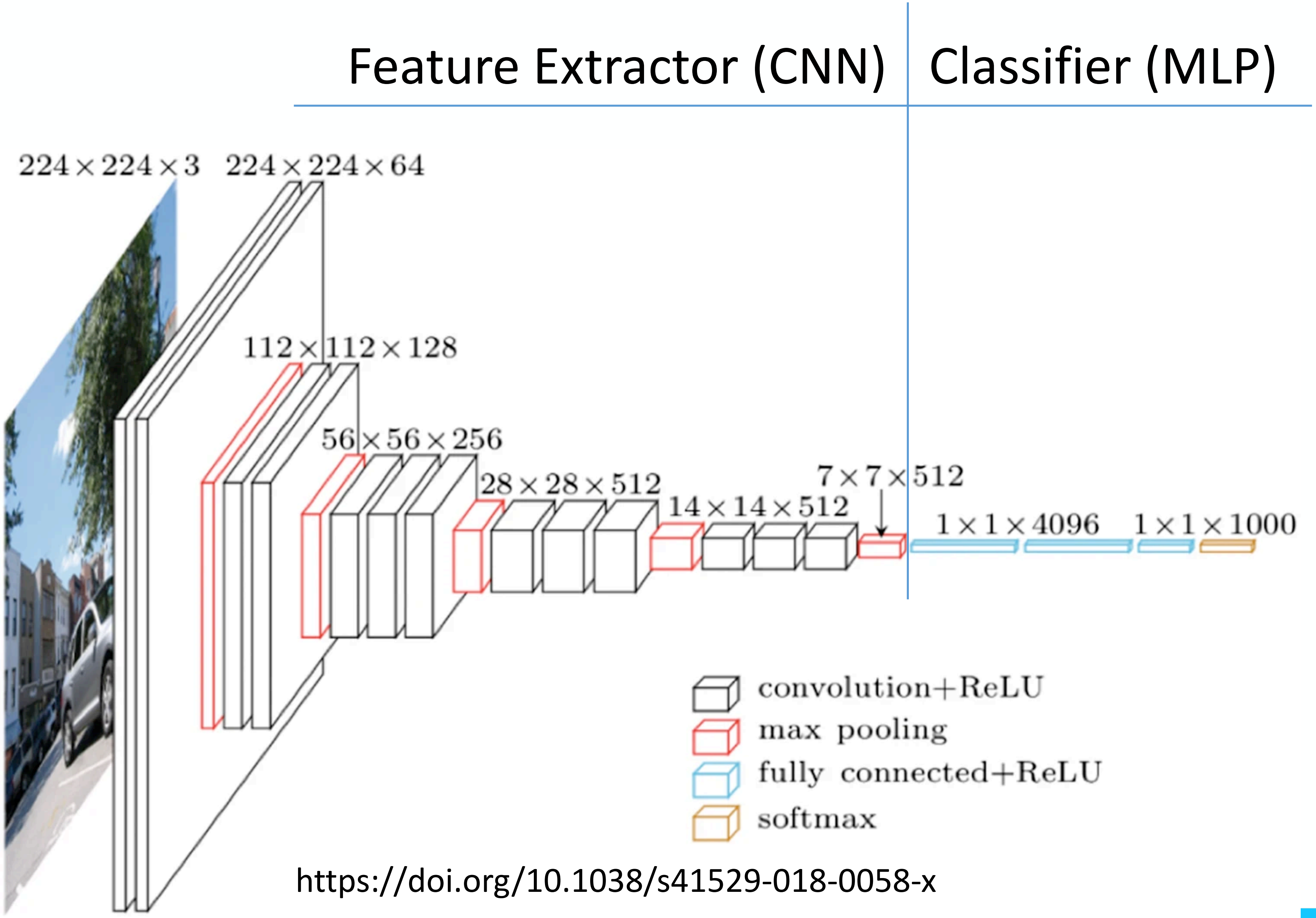
FLIPPED CLASSROOM

- Learners individually study materials (before course)
- Exercises solved by group in class
- Materials must be suited for self-study
- Pro:
    - Learners study at their own pace
    - Learner explain exercises to each other: they still remember what was the key information for them to understand
    - Time efficient
- Con:
    - Puts more responsibility on the learners: Learners that do not study will be completely lost

PAIR PROGRAMMING

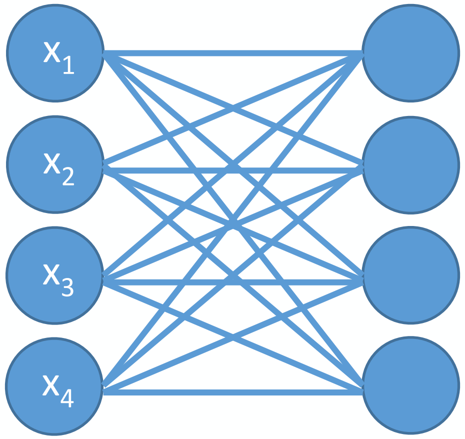
- Learners study materials and prepare exercises in small groups
- Materials and exercises are discussed in class
- Ideal for audiences with differing experience levels
- Works very well online (breakout rooms)
- For coding:
    - One person writes code (=driver, attention to detail)
    - The others advise and consult (=navigators, high level attention)
    - IMHO hands down the best method for teaching to code
- Pro:
    - Learning is more fun and more efficient in small groups
- Con:
    - Logistics in presence: Noise in large classrooms / Needs a lot of space

CNN EXTRACT FEATURES FROM IMAGES

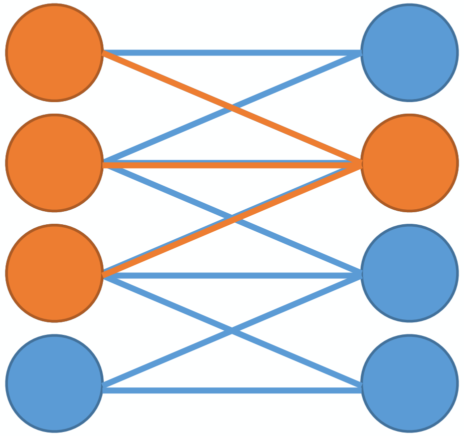


https://doi.org/10.1038/s41529-018-0058-x

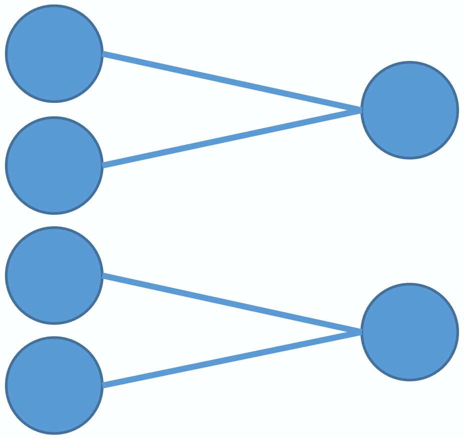
NETWORK LAYER ARCHITECTURES



Fully connected layer

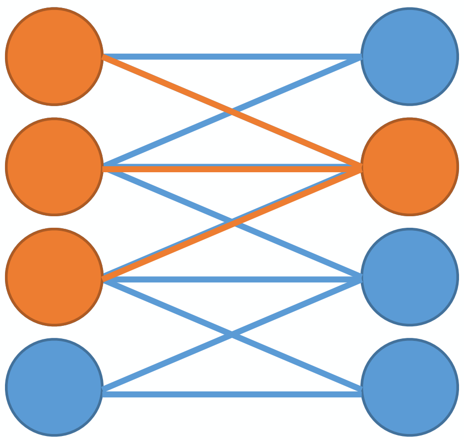


Convolutional layer

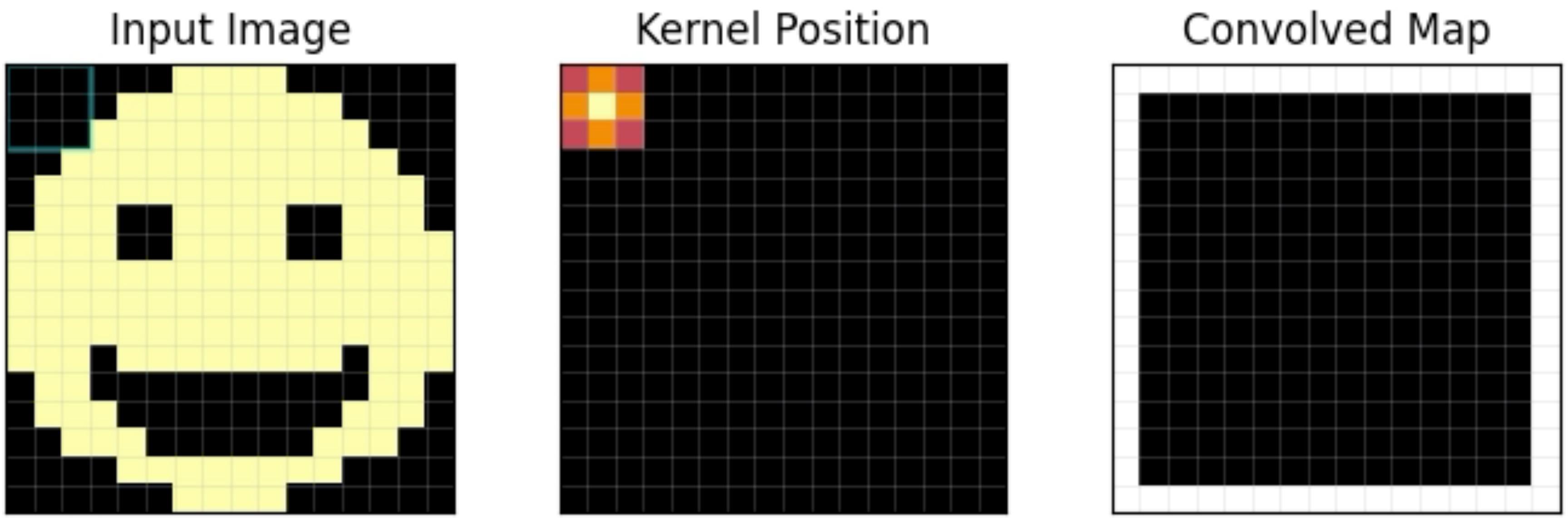


Pooling layer

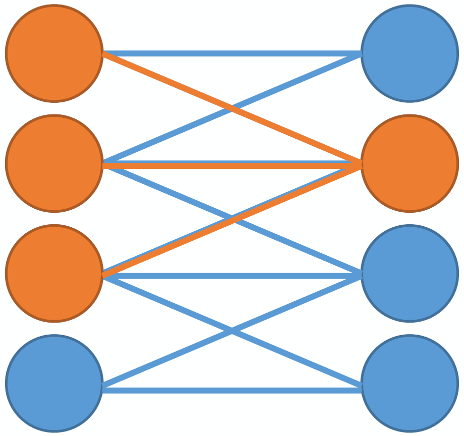
CNN – CONVOLUTION WITH LEARNED KERNELS



Convolutional layer



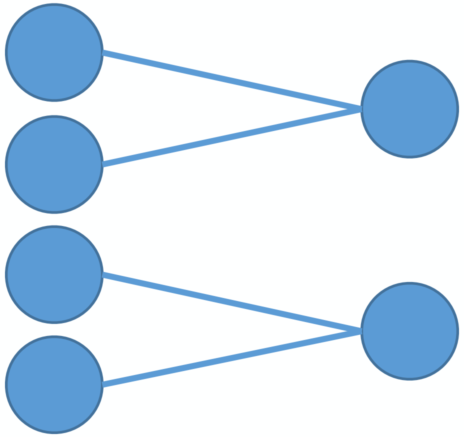
CNN – CONVOLUTION WITH LEARNED KERNELS



Convolutional layer

- Convolution operation enforces spatial invariance in the network architecture
- Reduces number of parameters
- Makes networks much more computationally and data efficient

POOLING LAYERS REDUCE THE LAYER SIZE



Pooling layer

3	15	1	13
9	7	0	10
11	5	5	3
1	8	9	6

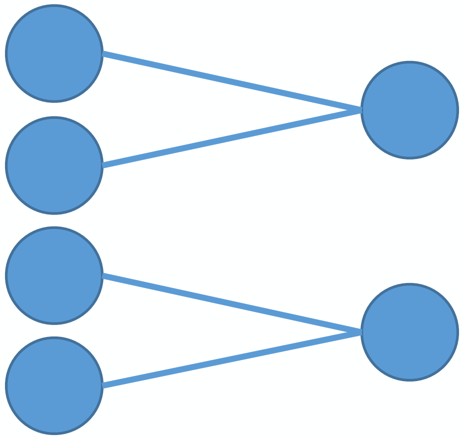
Maximal values

15	13
11	9

Average values

8.5	6.0
6.3	5.8

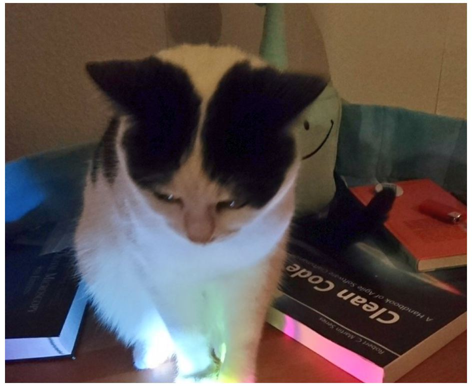
POOLING LAYERS REDUCE THE LAYER SIZE

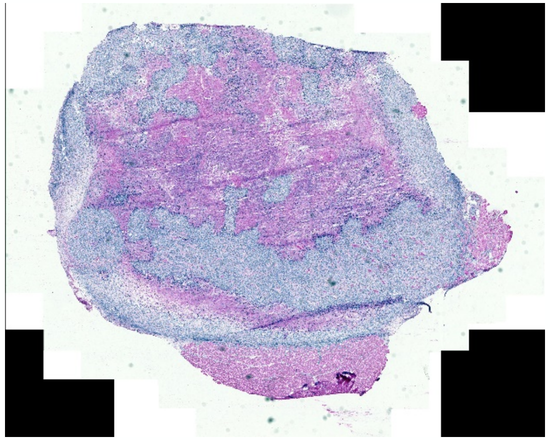


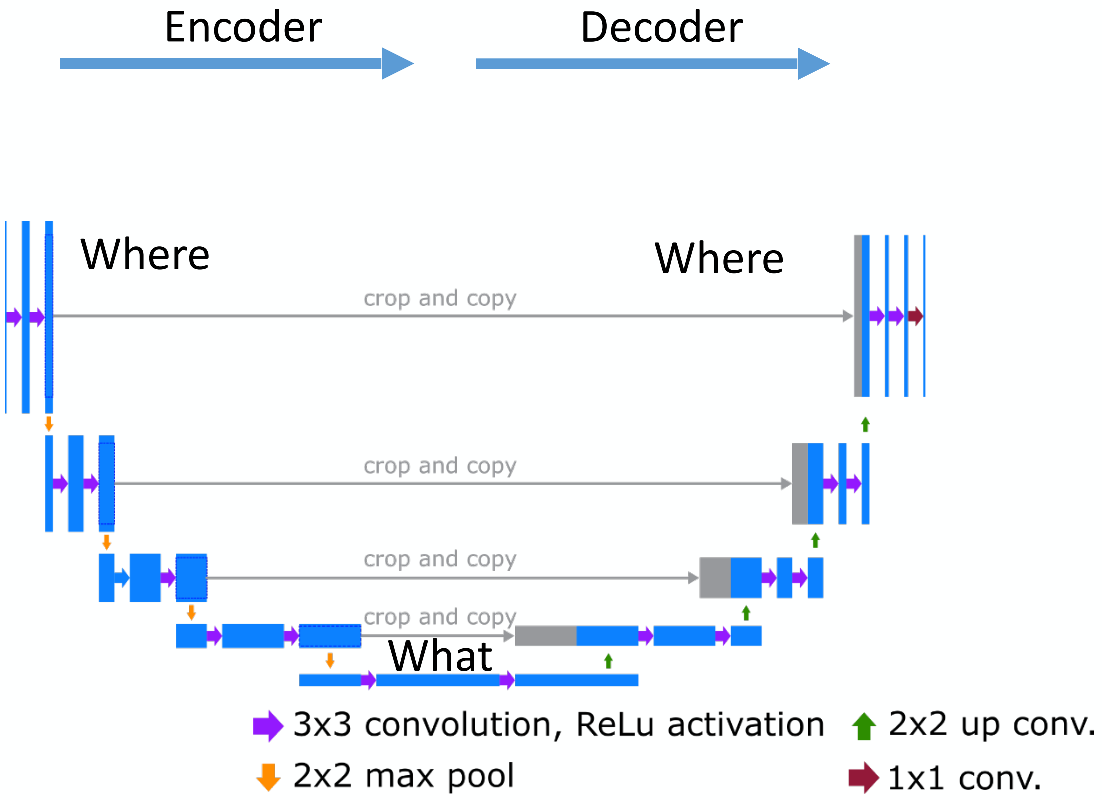
Pooling layer

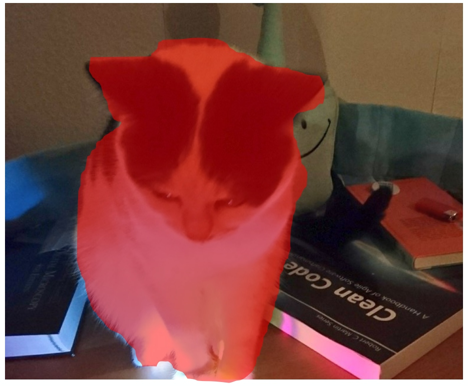
- Halves input image width/height
- Usually combined with increased size of feature dimension (depth)
- Often followed by batch normalisation to reduce risk of exploding weights
- Helps network to learn more complex features
- Doubles perceptive field

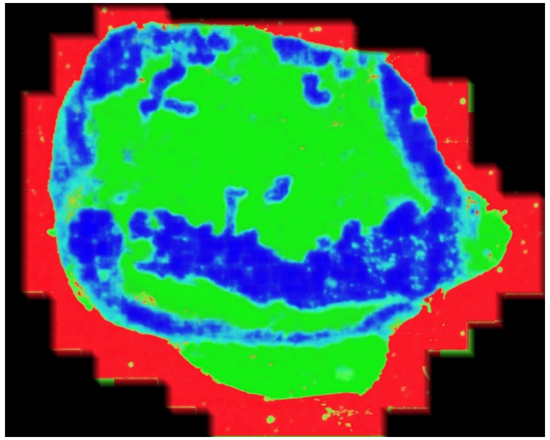
U-NET: IMAGE SEGMENTATION







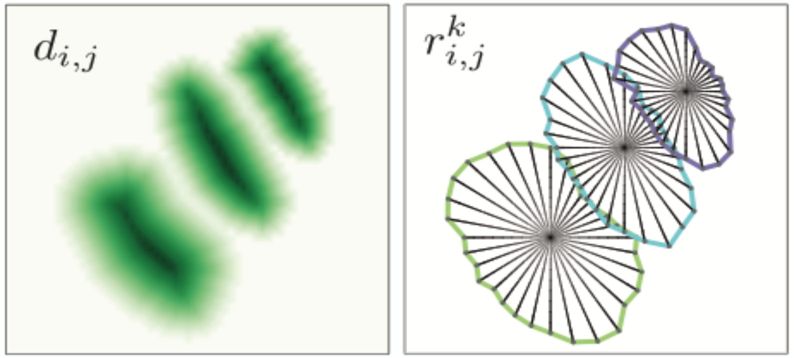


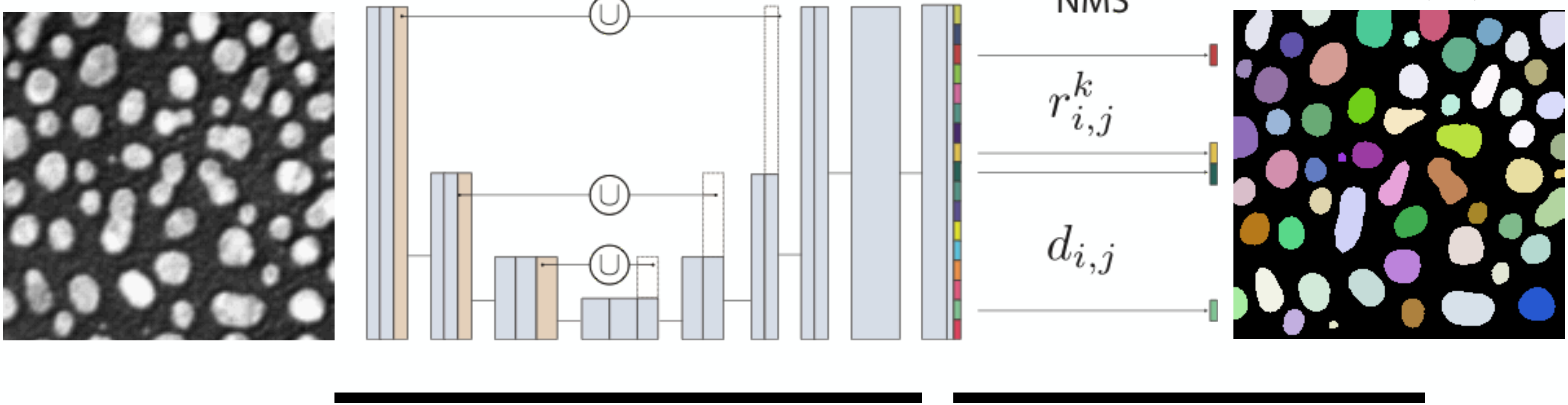


- The **U-net** is the most used network architecture in biological image processing using CNNs.
  - Encoder: Increase the “What”, decrease the “Where”
  - Decoder: Use the “What”, to identify the “Where”

STARDIST: NUCLEUS SEGMENTATION

Strategy:
→Add additional information to prediction
→Member pixels of objects (nuclei) can be reached via a straight line from the center





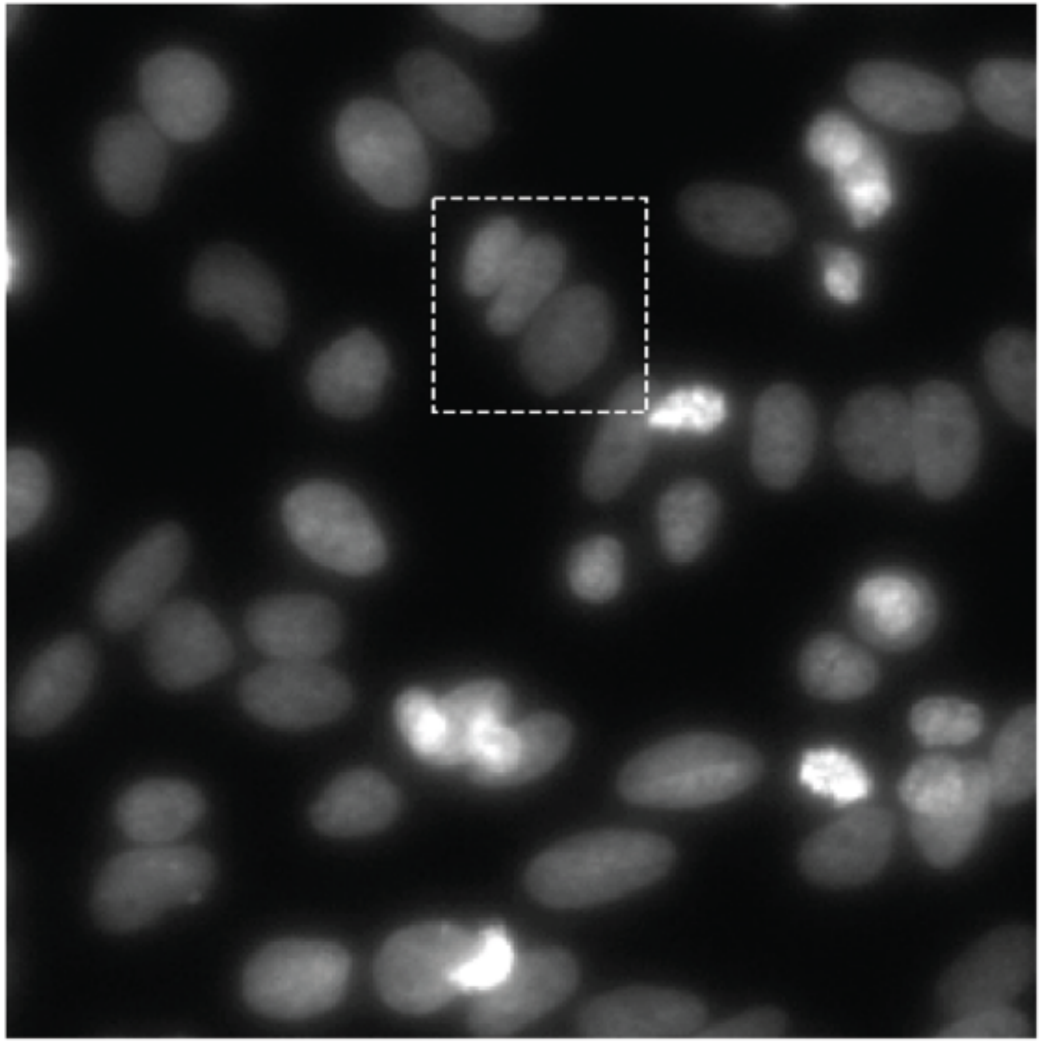
Dense Polygon Prediction
(e.g. U-Net, ResNet)

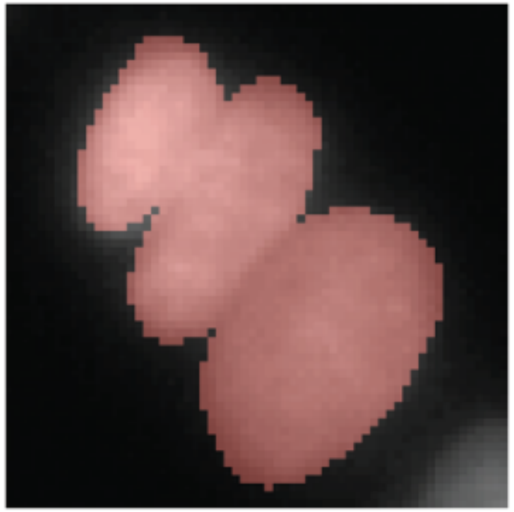
Polygon Selection
(Non-Maximum Suppression NMS)

Schmid et al., MICCAI (2018), https://github.com/stardist/stardist

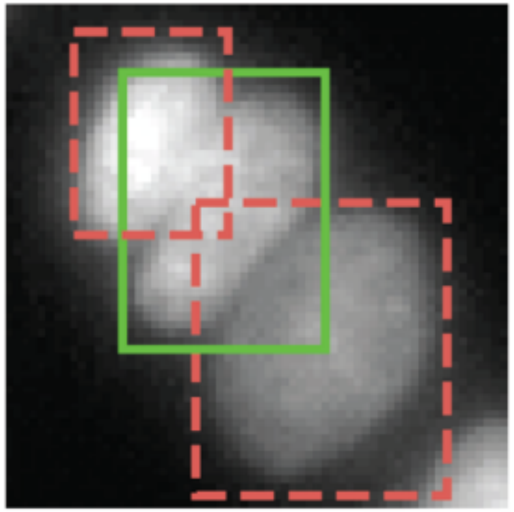
OTHER ALGORITHMS HAVE PROBLEMS WITH OVERLAPPING NUCLEI

Noisy images + Crowded cells = Common source of segmentation errors





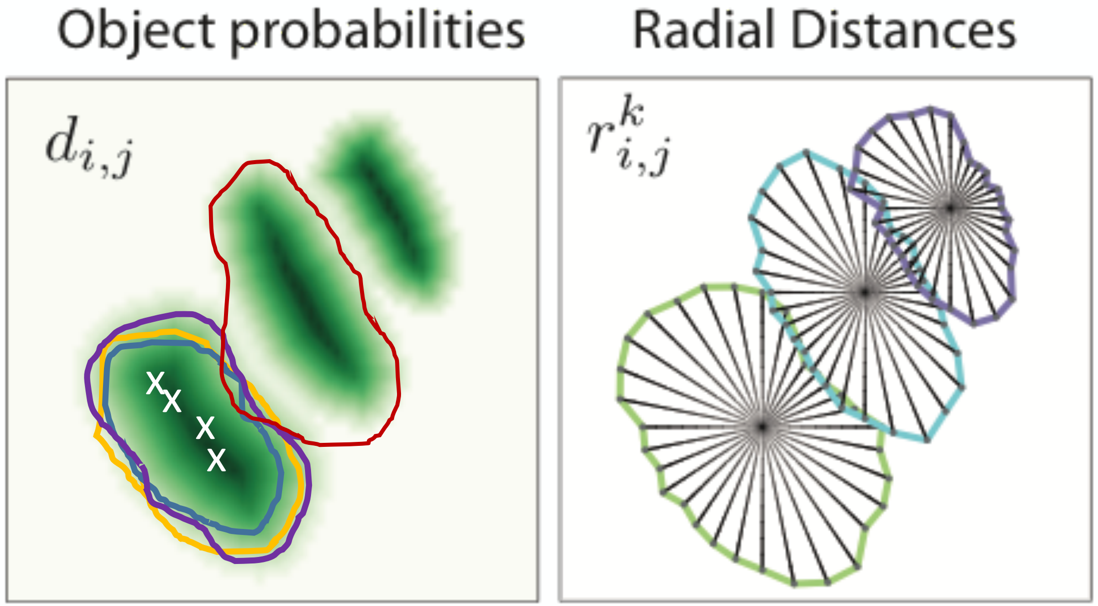
Dense Segmentation
(e.g. U-Net)



Bounding box based methods
(e.g. Mask-RCNN)

Schmid et al., MICCAI (2018), https://github.com/stardist/stardist

STARDIST: NON-MAXIMUM-SUPPRESSION RESOLVES OVERLAPS



© Schmid et al., MICCAI (2018)

Non-maximum-suppression (NMS):

→ Object probabilities: Probability that pixel belongs to class “nucleus”

→ Multiple maxima lead to multiple possible polygons for the same nucleus

Algorithm:

→ Select polygon with highest object probability inside:

→ Look at other polygons: Is the overlap of with larger than threshold τ?

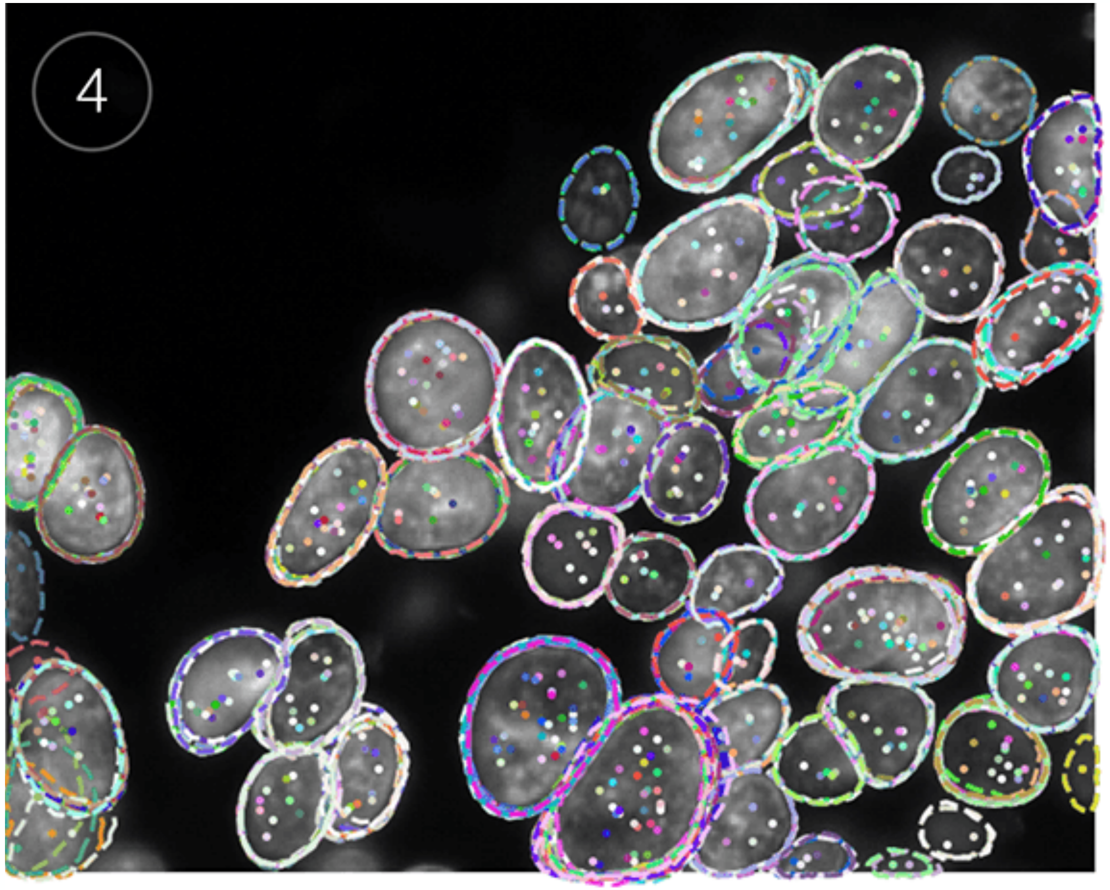
→ Yes: and are actually the same object, drop

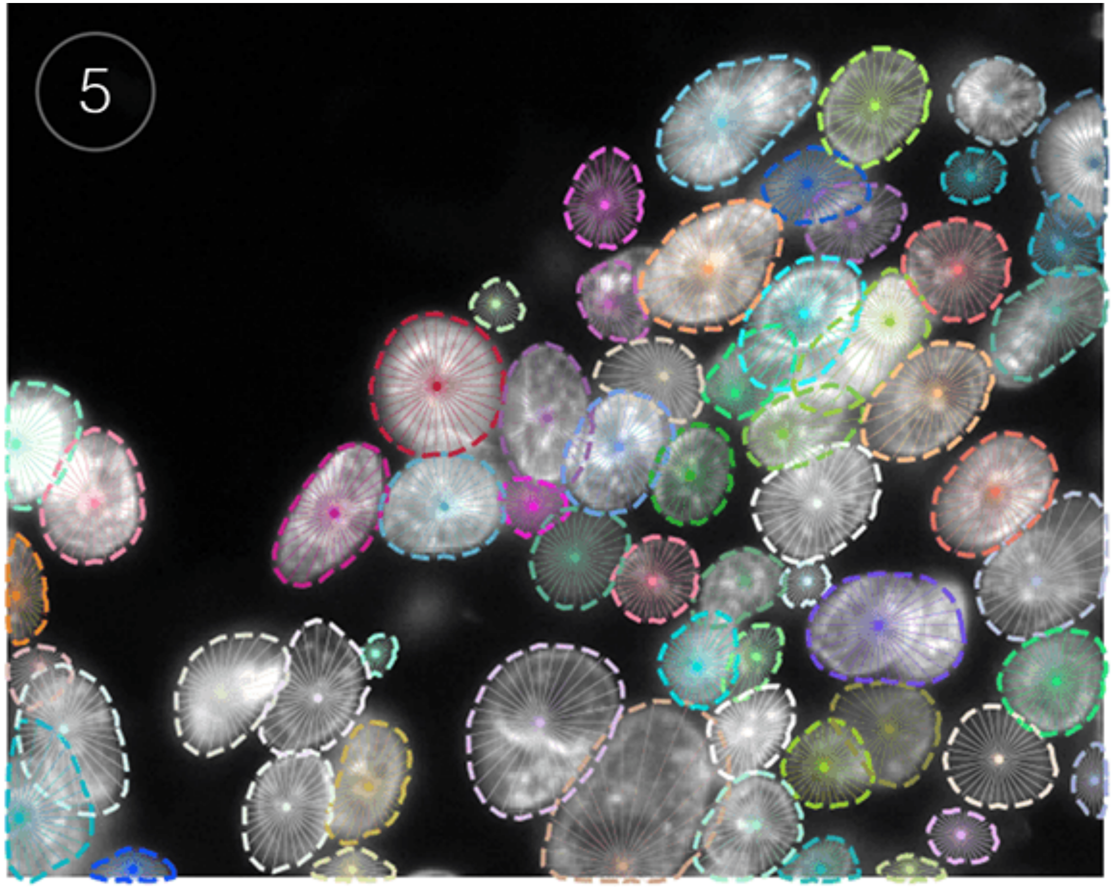
→ No: and are separate nuclei

→ Setting τ very high leads to many false positives!

STARDIST: NON-MAXIMUM-SUPPRESSION WORKS WELL

Non-maximum suppression

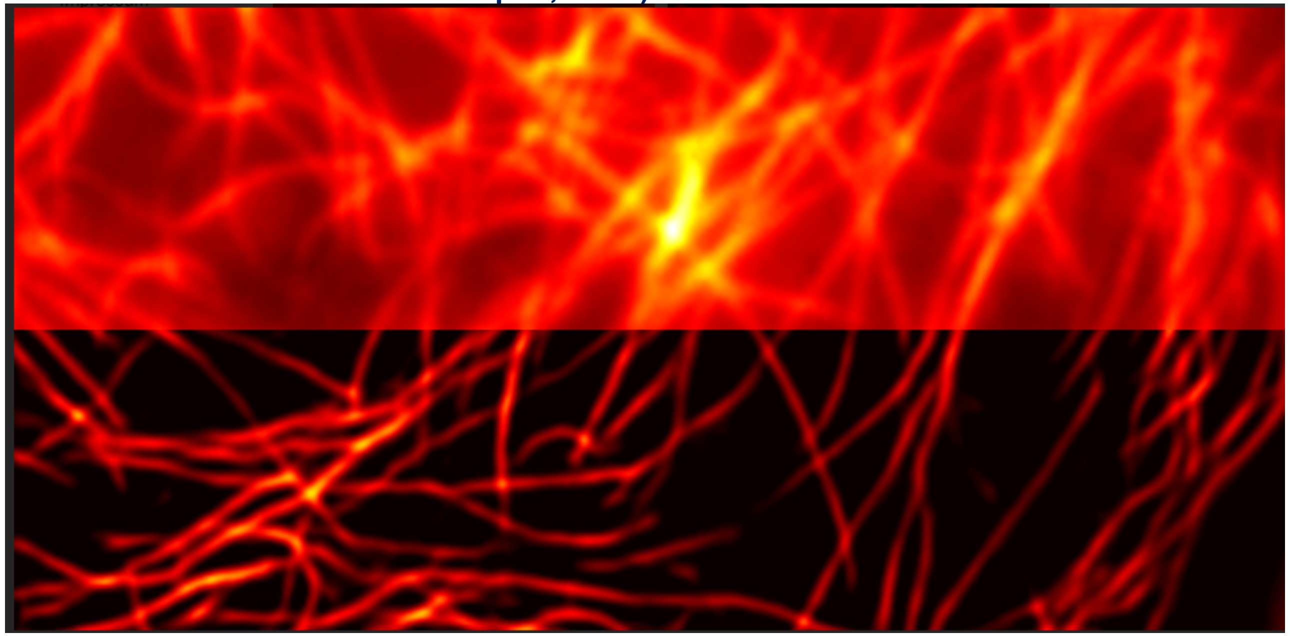


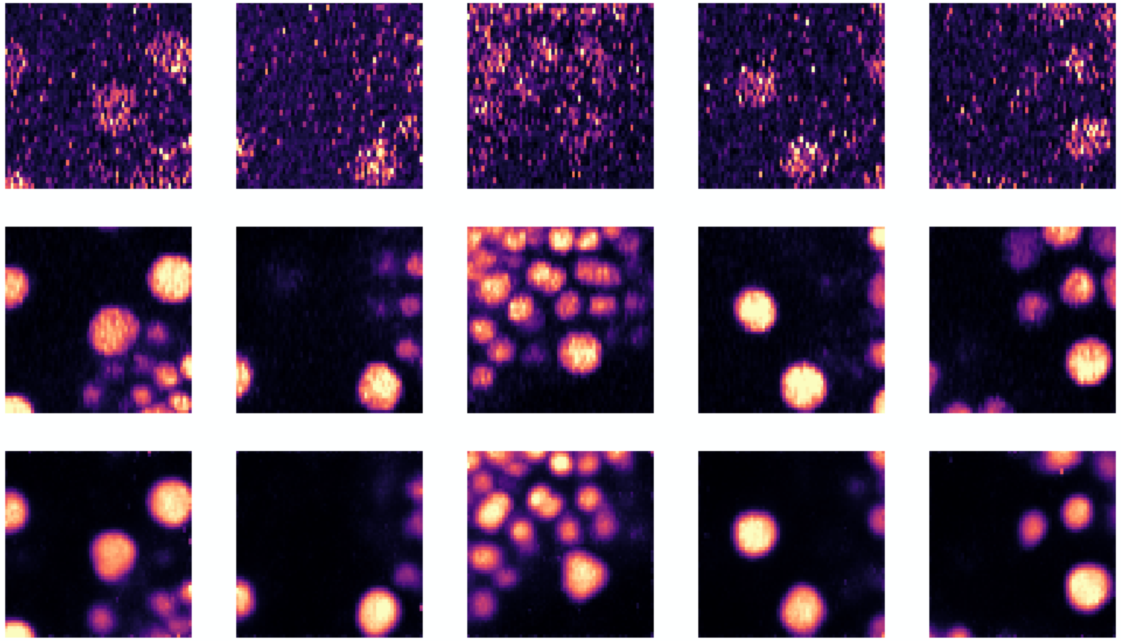


Schmid et al., MICCAI (2018), https://github.com/stardist/stardist

CARE: IMPROVING RESOLUTION AND DENOISING

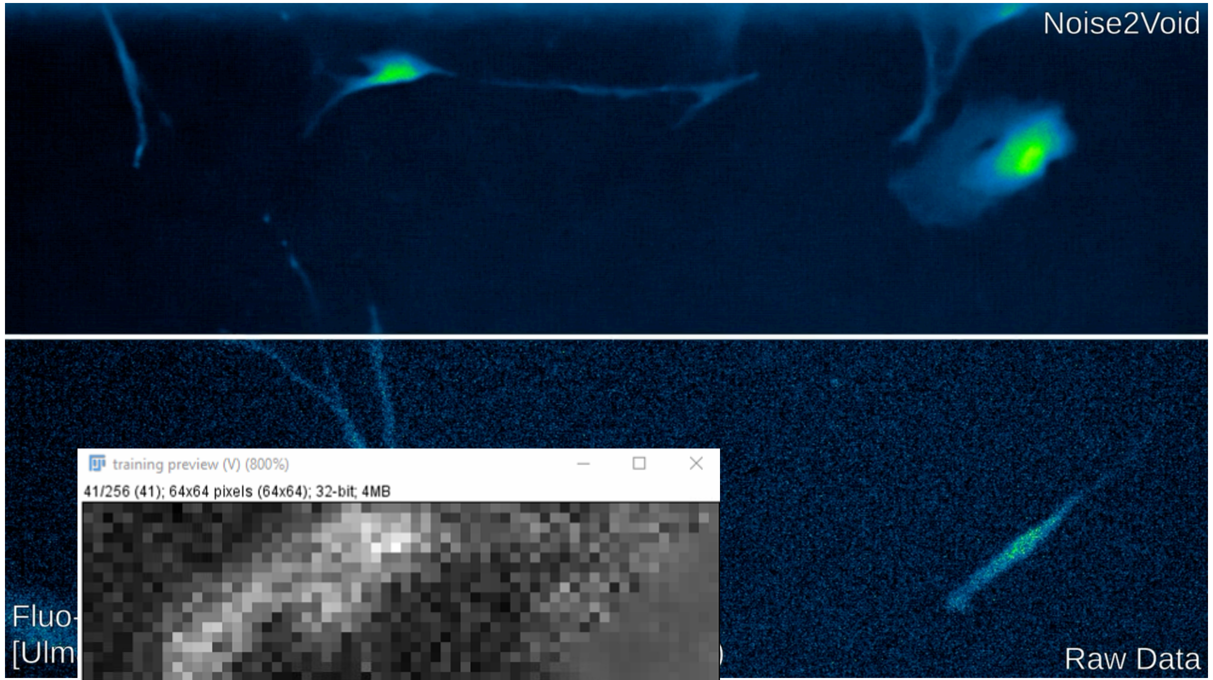
- **CARE**: content-aware restoration
- Image acquisition of pairs of images: A high-quality and a low-quality image.
- Caveats:
    - Reconstructs shot noise present in high quality training images
    - Trained model only applicable to image data of the same conditions (biological sample, microscope, etc)

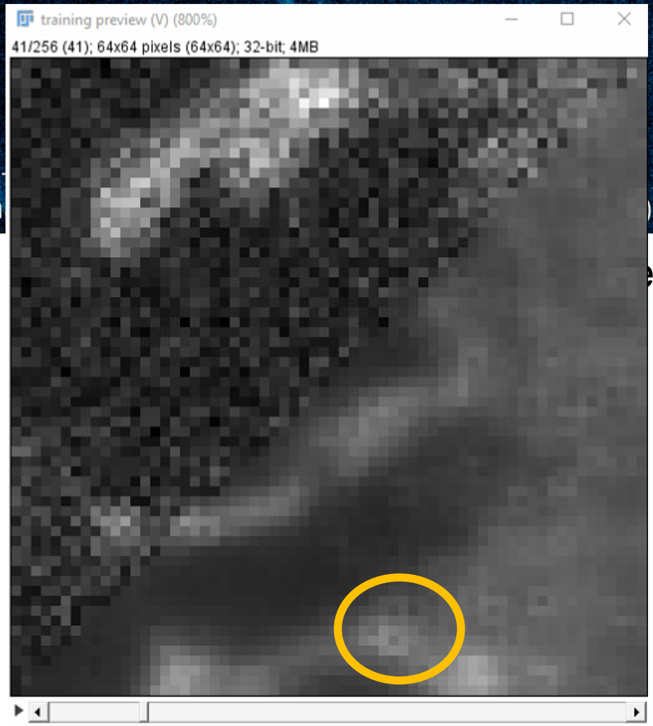




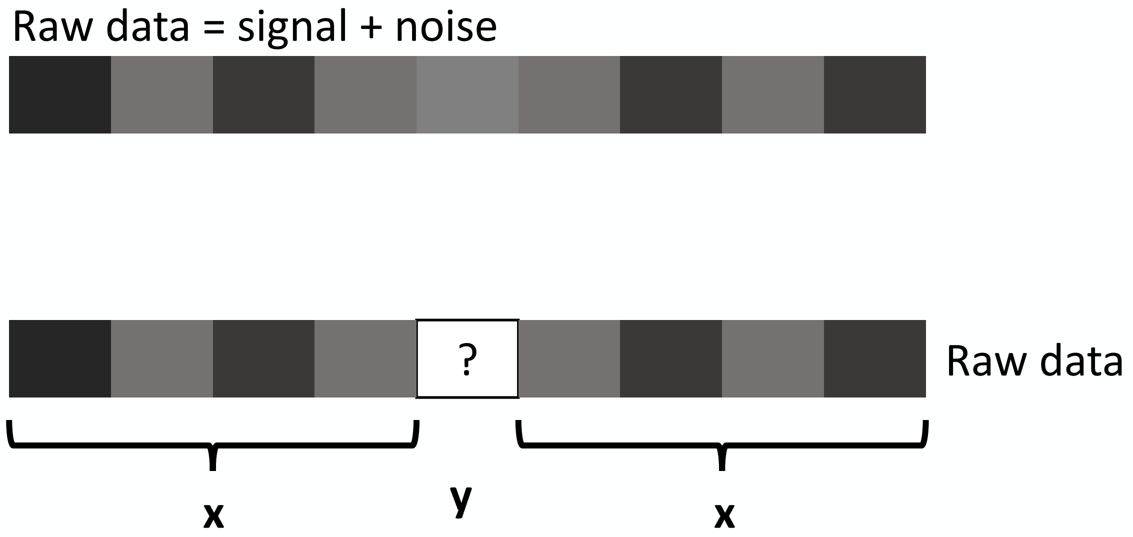
https://csbdeep.bioimagecomputing.com/

NOISE TO VOID: IMAGE DENOISING





© Cameron Nowell



Strategy:

→ Try to predict intensity of pixel y from surrounding pixels x

→ CNN fails to predict noise component → N2V can only reproduce signal from the surroundings of **y**

Beware:

→ Only **random** noise can be removed, otherwise artifacts occur

https://github.com/juglab/n2v

https://forum.image.sc/t/n2v-artefacts-in-training-data/70686

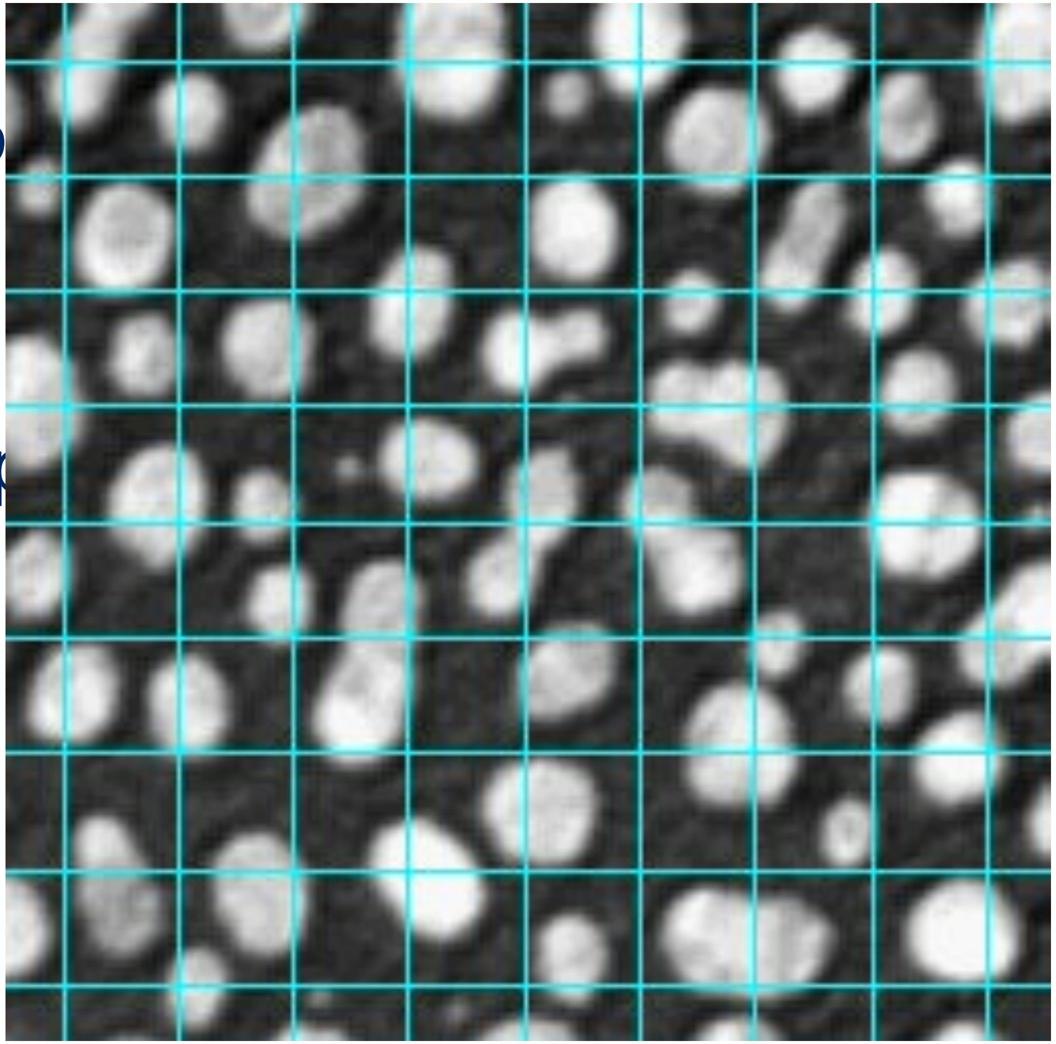
CAVEATS: GPU MEMORY LIMITS IMAGE SIZE

→Images are tiled

→limited "receptive field of the netwo

Receptive field:

→Objects must be smaller than recep



UNBALANCED TRAINING DATA LEADS TO BIASED RESULTS

Unbalanced training data:

→Some labels appear more often in training data than others

→Rare events will not be learned because missing them doesn't harm accuracy much

→Weighted data sampling

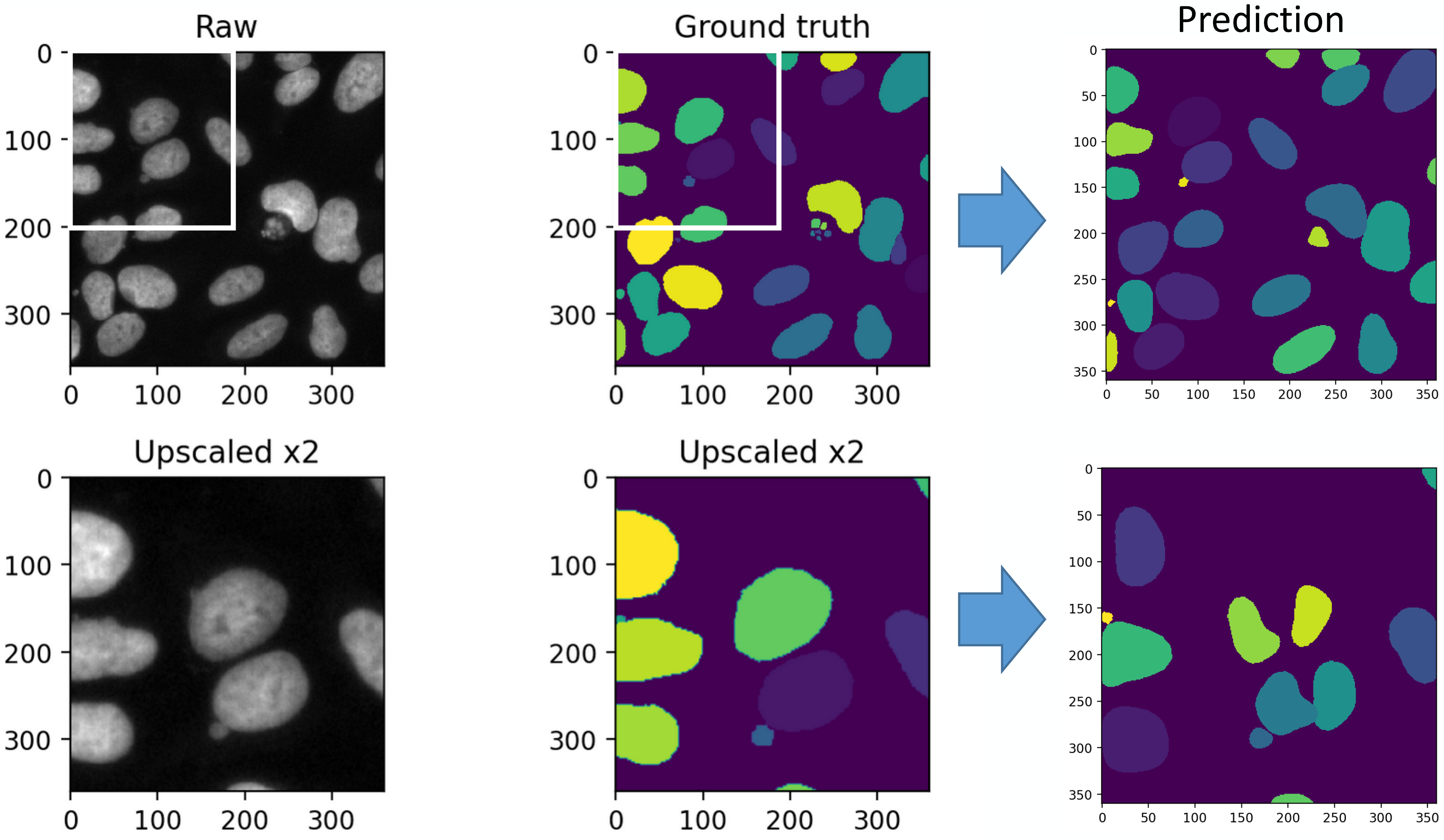
→**Biased results!!**

Is the iPhone racist? Chinese users claim iPhoneX face recognition can't tell them apart

APPLE has come under fire following numerous complaints from Chinese users who claim the iPhone X face recognition can't tell them apart.

https://www.news.com.au/technology/gadgets/mobile-phones/is-the-iphone-racist-chinese-users-claim-iphonex-face-recognition-cant-tell-them-apart/news-story/13814540e8c82ad466aca687e12af64c

WHEN THE INPUT DATA DOES NOT FIT TO THE TRAINING DATA



What happened here?

Receptive field too small

I used a different resolution than during training

Overfitting

TAKEAWAYS

- With great power comes great responsibility: **Validate your models well!**
- Better data more important than better model
- Often performs fantastic – *but you don’t know why*
- Generative neural networks (like CARE) can dream up data – to a hammer everything looks like a nail!

THANK YOU

- For your attention
- The Helmholtz AI Team
- Everyone who contributed to this material
    - Johannes Müller, Robert Haase: PoL
    - Alex Krull, Uwe Schmidt: MPI CBG
    - Martin Weigert: EPFL Lausanne
    - Ignacio Arganda-Carreras: Universidad del Pais Vasco