ISOO$^{V2}_{DL}$

SEMANTIC INSTANCE SEGMENTATION OF TOUCHING AND OVERLAPPING OBJECTS

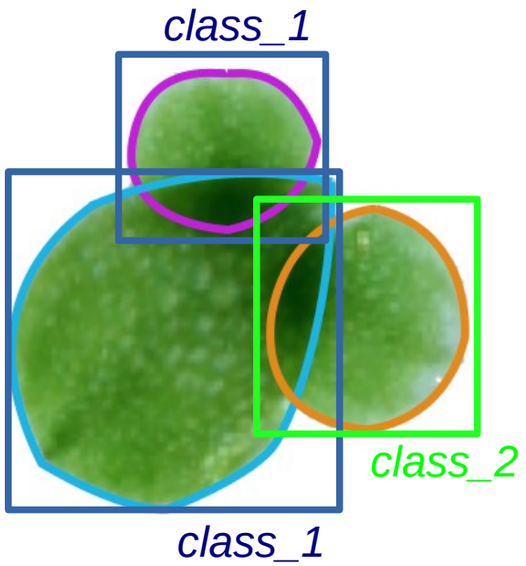
Anton Böhm [1] *Maxim Tatarchenko*[1]
Thorsten Falk [1,2]

[1] Department of Computer Science,
[2] BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

We thank the German Federal Ministry for Economic Affairs and Energy (FKz. ZF4184101CR5) and the DFG (EXC 294) for funding our research.

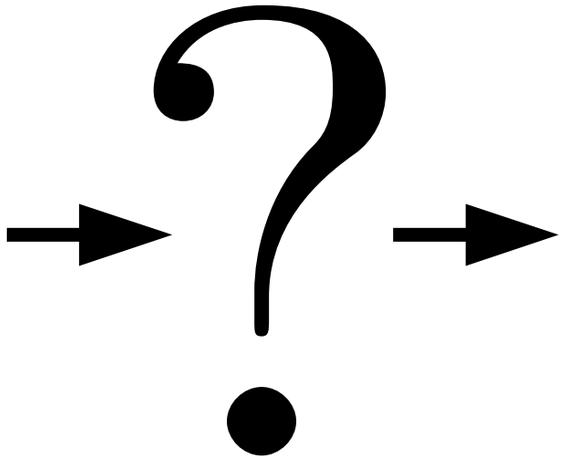
What is it about?

Given:



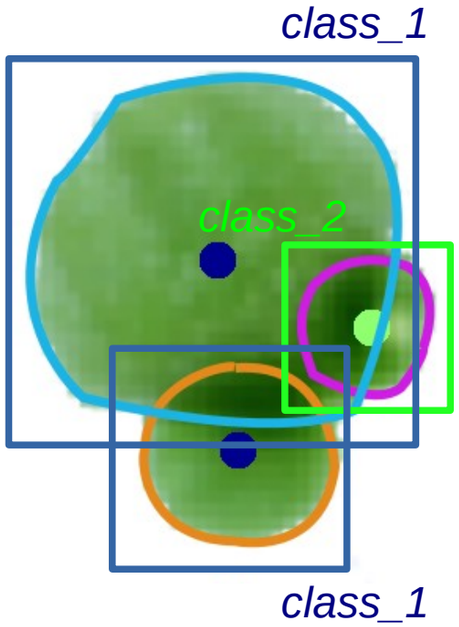
2D annotated images with overlapping objects

Method:



Deep learning approach

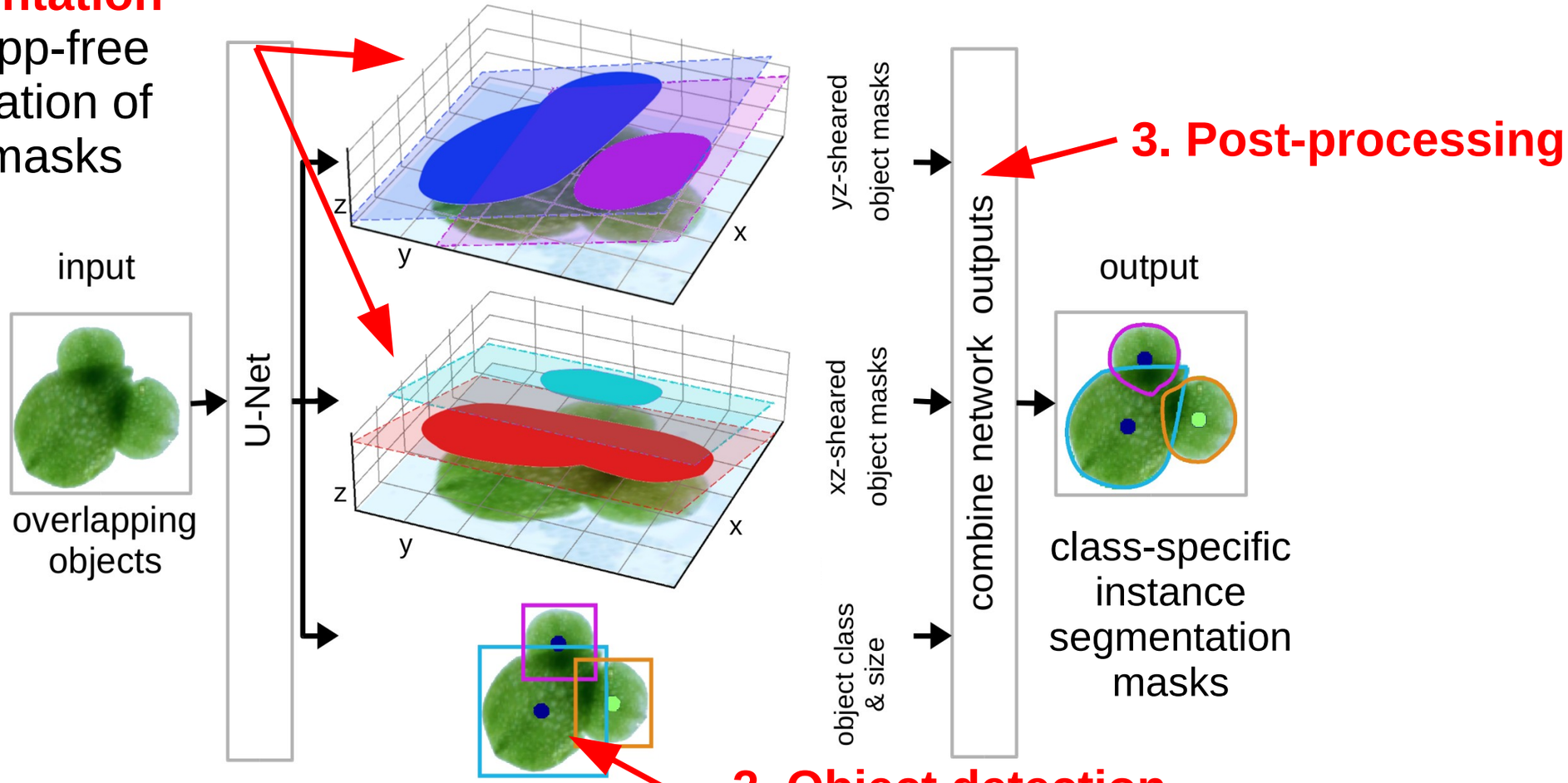
Wanted:



class-specific semantic instance segmentation masks for **unseen images**

ISOO-V1

1. Segmentation
3D overlapp-free representation of instance masks



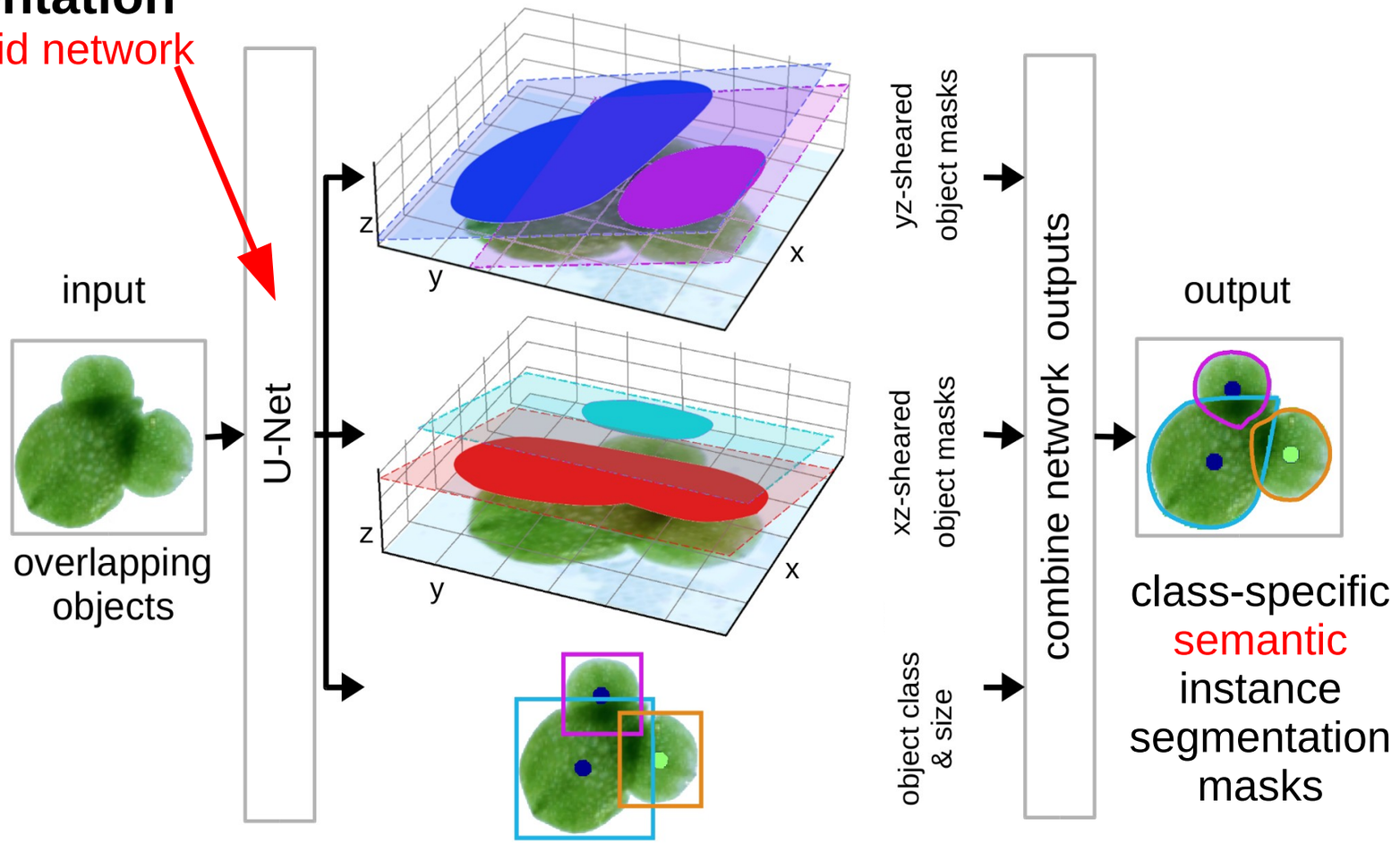
3. Post-processing

2. Object detection
reference points with object class and bounding box

ISOO-V2: Outline

1. Segmentation

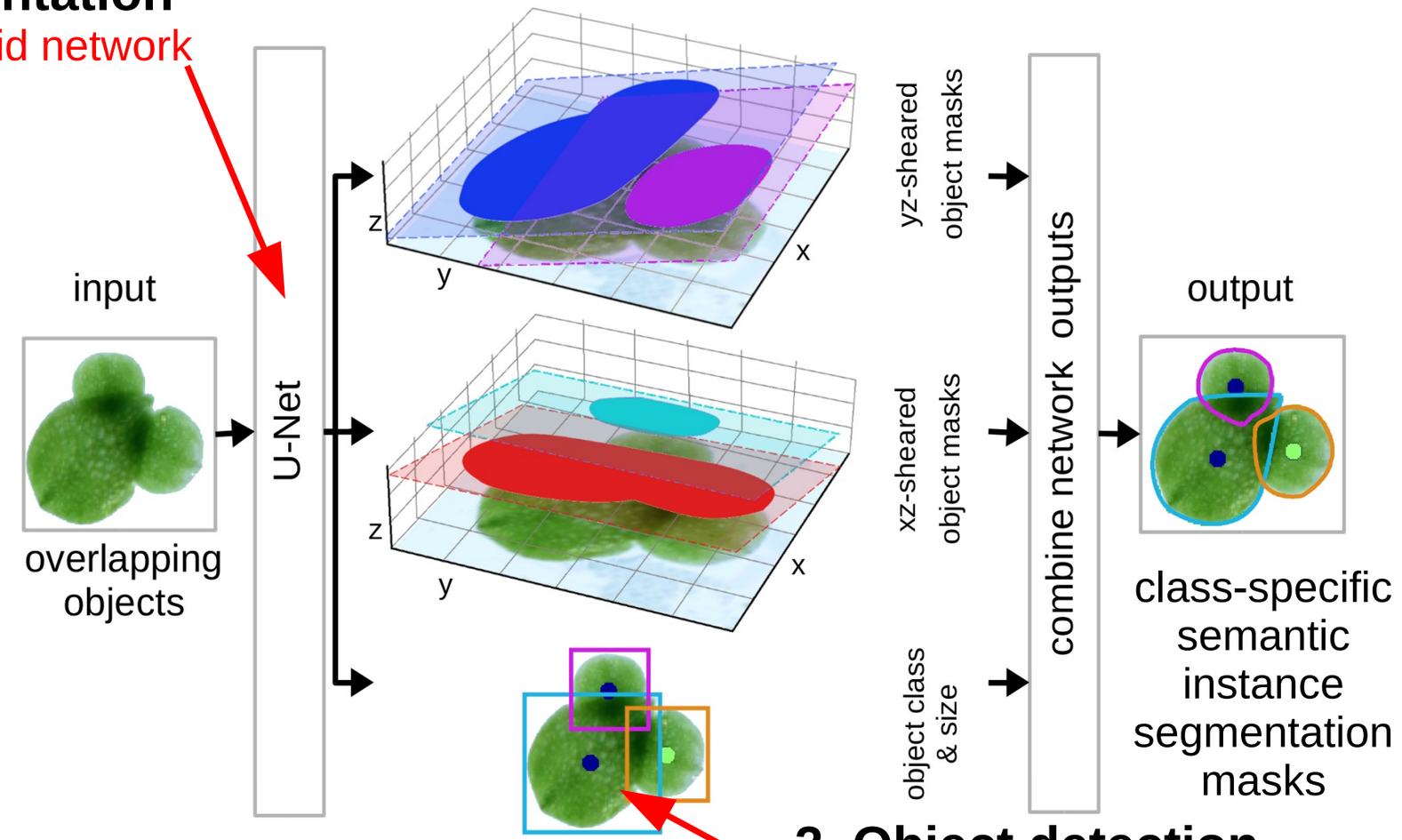
2D-3D hybrid network



ISOO-V2: Outline

1. Segmentation

2D-3D hybrid network



2. Object detection

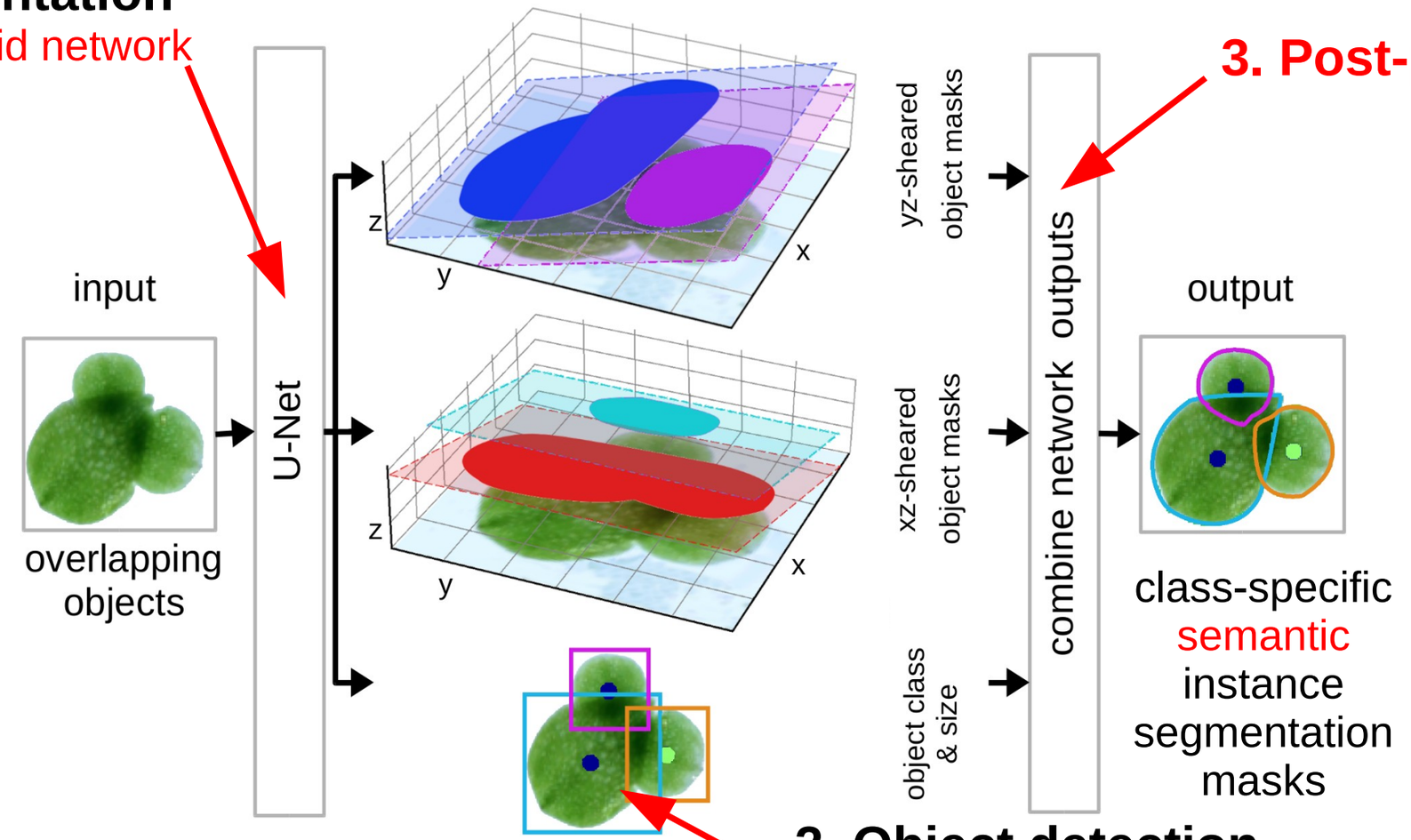
reference points with dynamic size adaptation

bounding box reparametrization

ISOO-V2: Outline

1. Segmentation

2D-3D hybrid network



3. Post-processing

2. Object detection

reference points with dynamic size adaptation

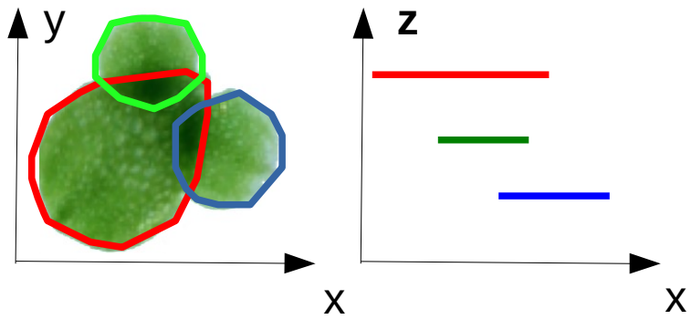
bounding box reparametrization

1. Segmentation

Segmentation: 3D representation

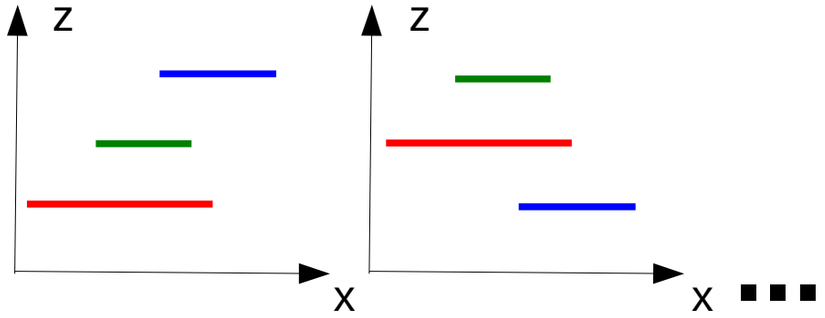
Basic Idea:

Extend label-space to **third dimension**



Problem:

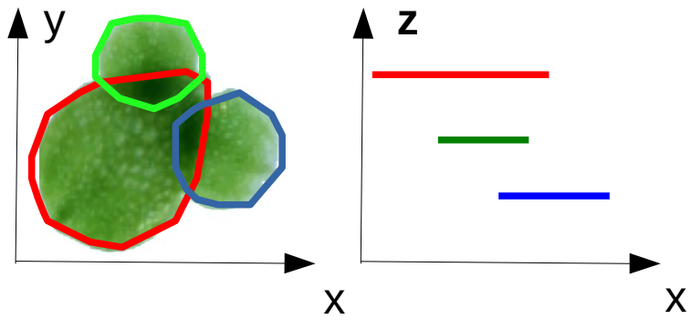
Ambiguous order of instances



Segmentation: 3D representation

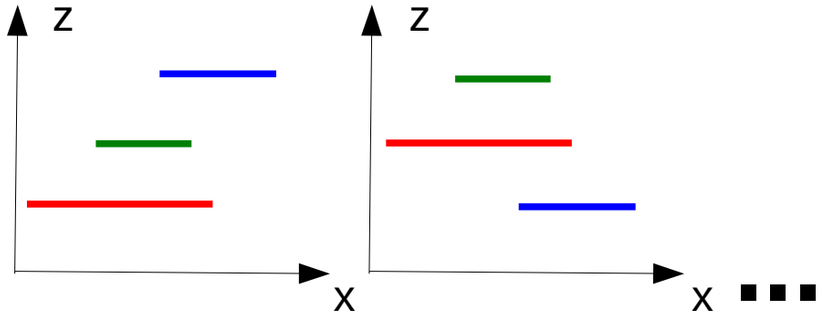
Basic Idea:

Extend label-space to **third dimension**



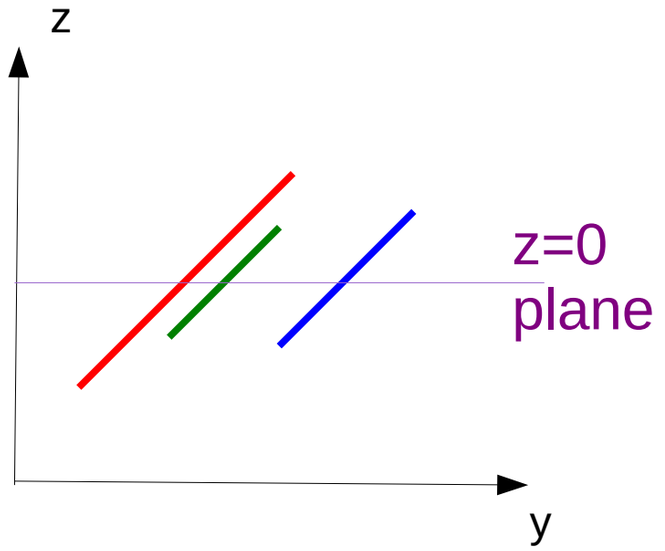
Problem:

Ambiguous order of instances



Solution:

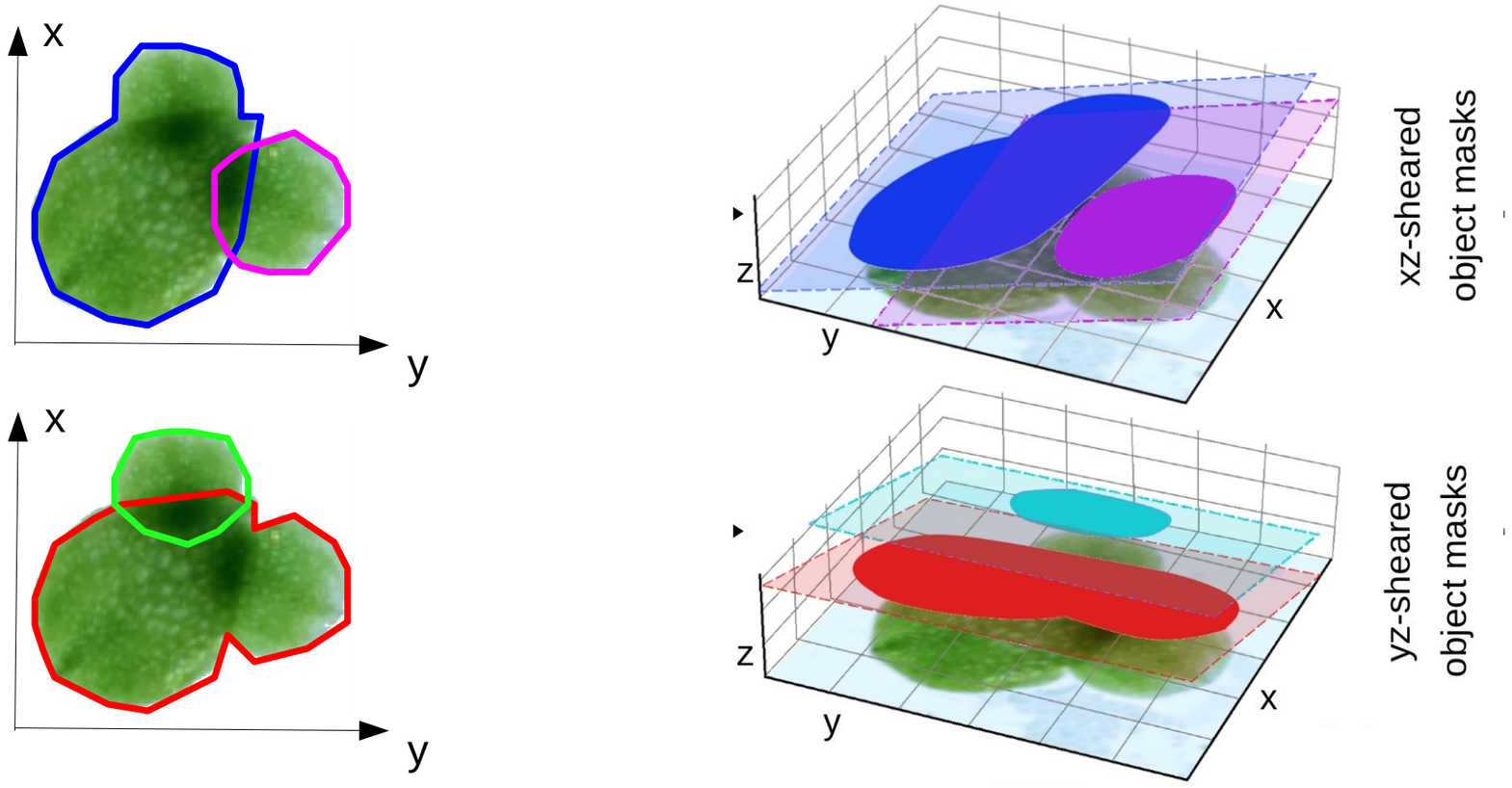
3D overlap-free order-independent masks encoding



Shearing w.r.t object bounding box centers which lie on **z=0 plane** (magenta)

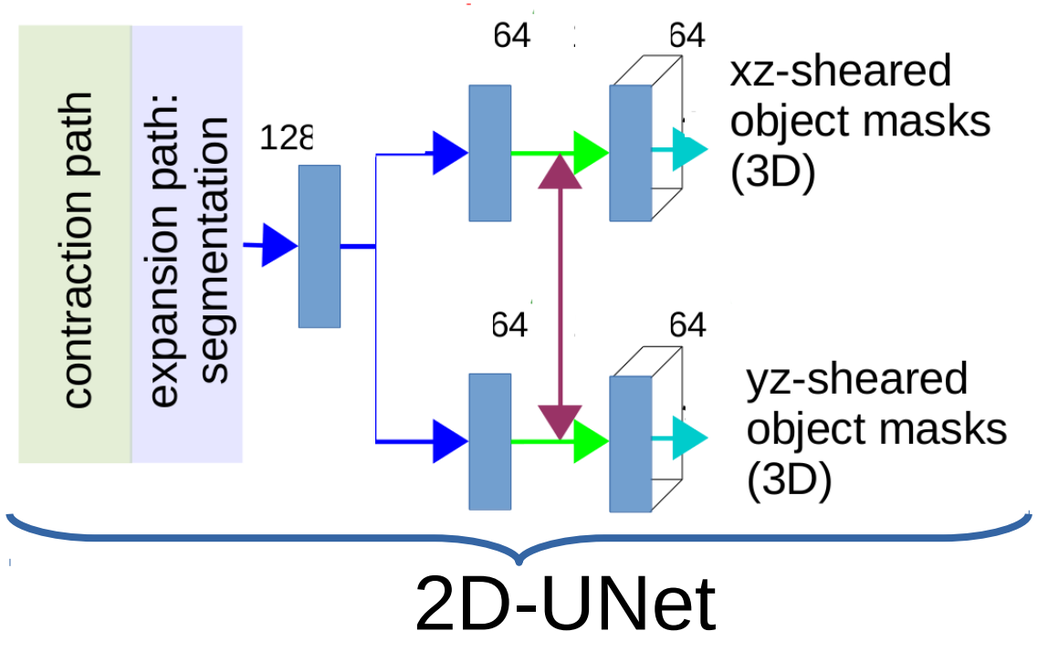
Segmentation: 3D representation

- **Shear** object masks in xz- **and** yz-direction

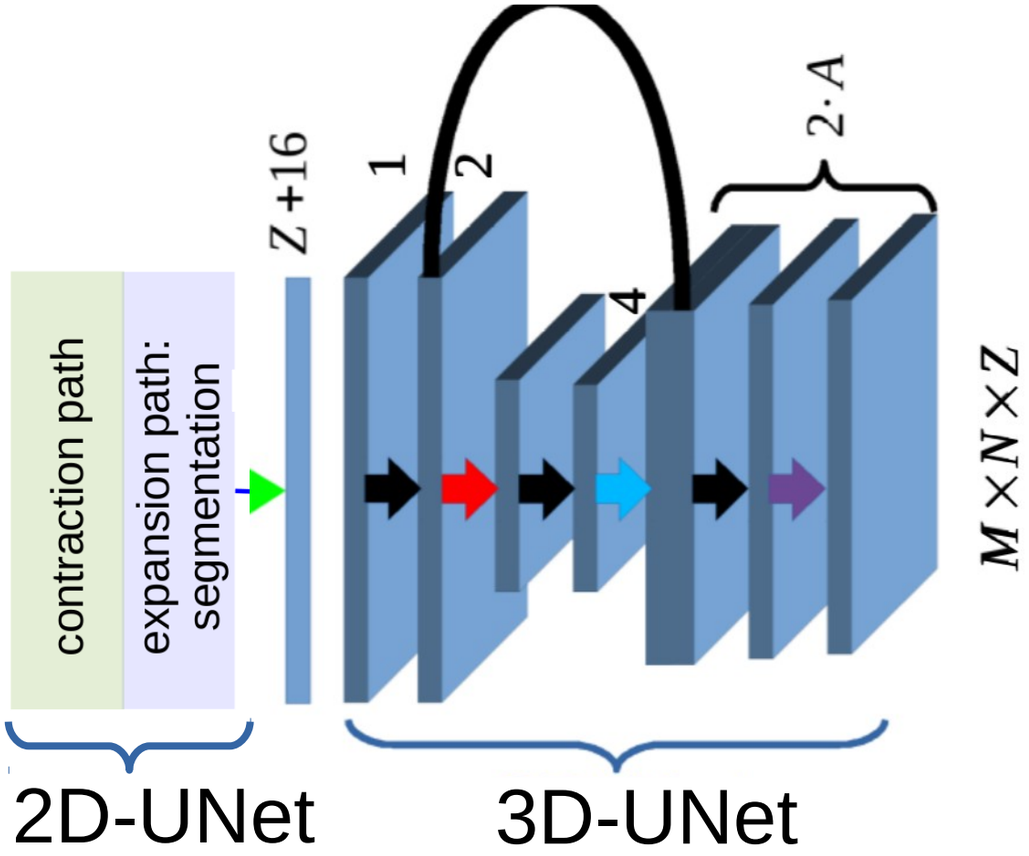


(binary representation, clustered for visibility purposes)

Segmentation: Network architecture

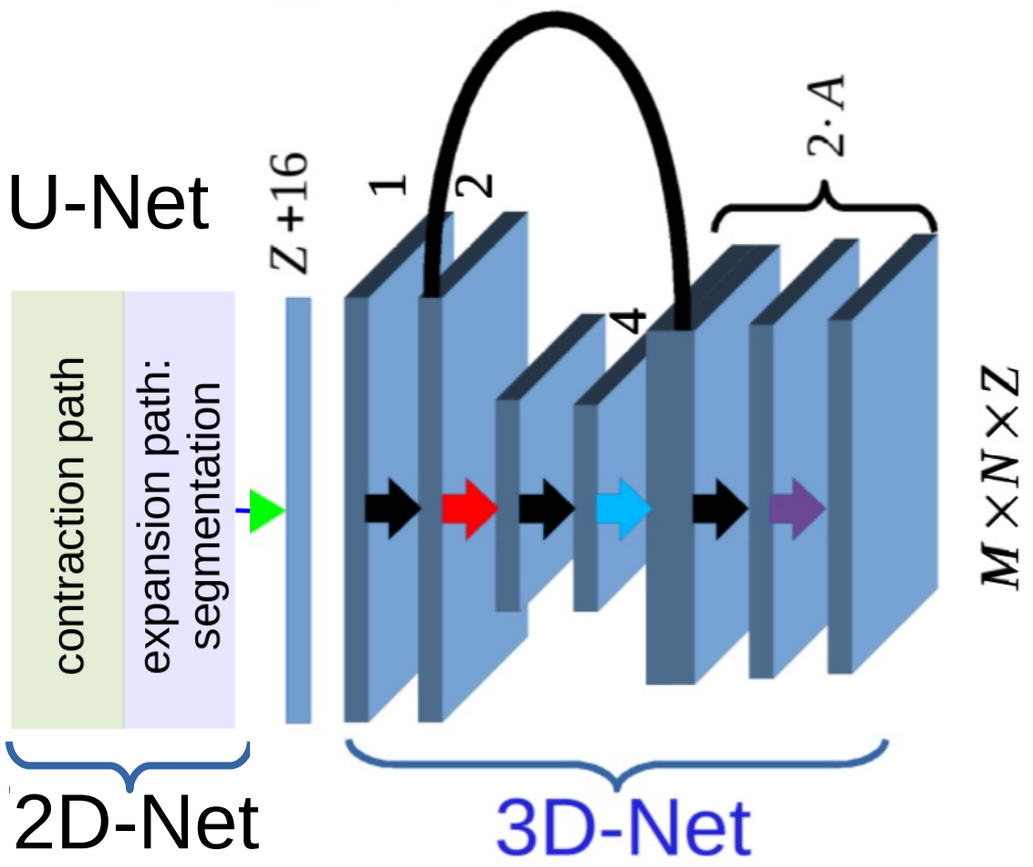


ISOO-V1: 2D Network

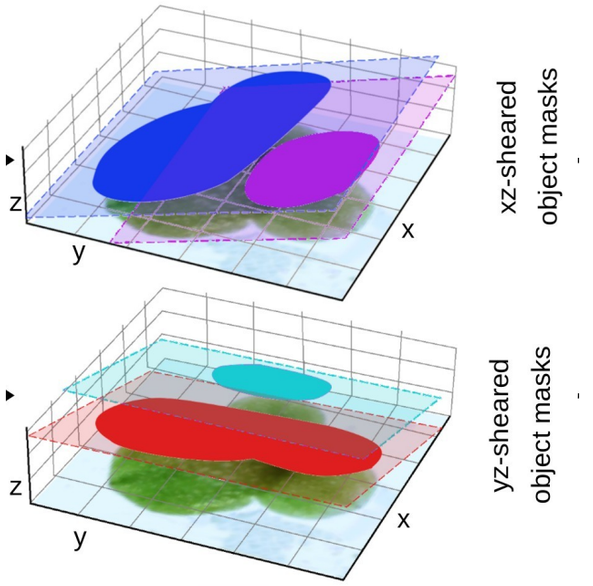


ISOO-V2: 2D-3D hybrid network

2D-3D network: Advantages

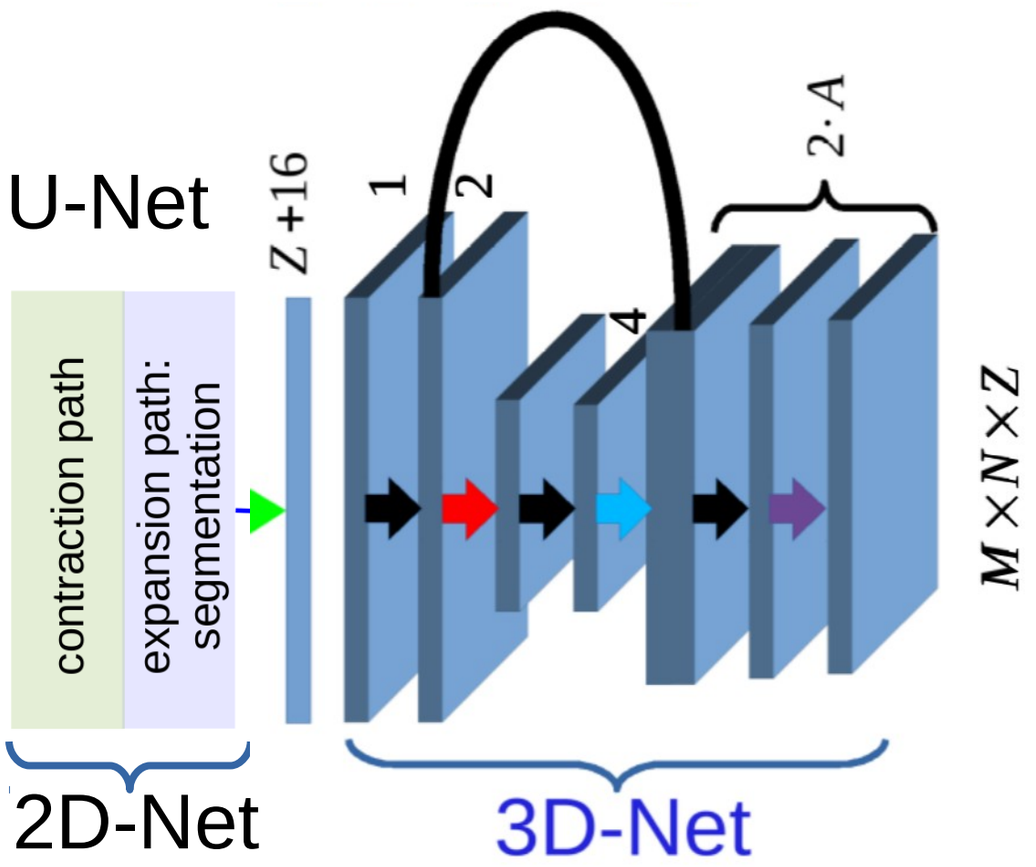


1) 2D network does not optimally exploit **correlations in the z-direction** of the label space. 2D-3D network does.



2D-3D hybrid network

2D-3D network: Advantages



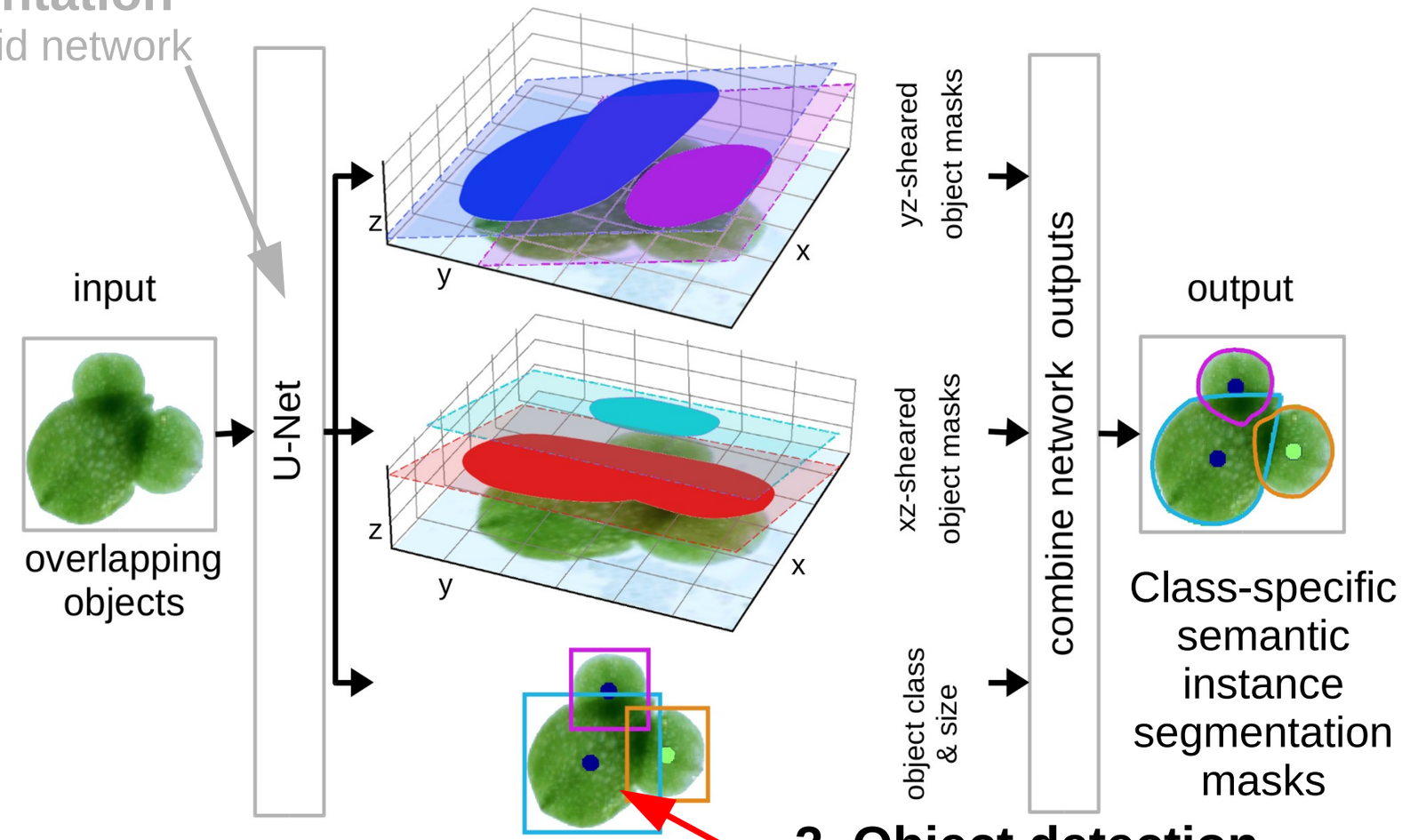
2) additional dimension allows for object's **sub-parts** prediction / **semantic segmentation**.

"A" is the number of semantical classes within an objects
e.g.
A=1 for cell body
A=2 for cell body and nucleus

2D-3D hybrid network

ISOO-V2: Outline

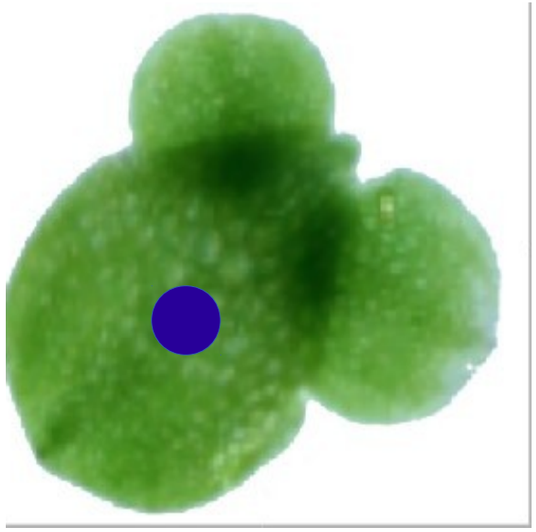
1. Segmentation
2D-3D hybrid network



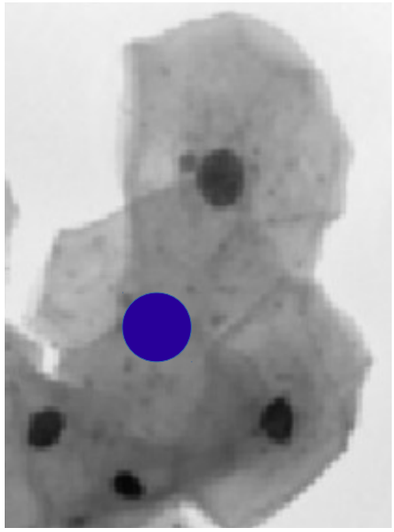
2. Object detection
reference point with size adaptation
bounding box reparametrization

Detection: Reference points

- Reference point is a unique position in object
- Reference points are grouped to disks
- Every disk represents an object



case: reference point is the bounding box center

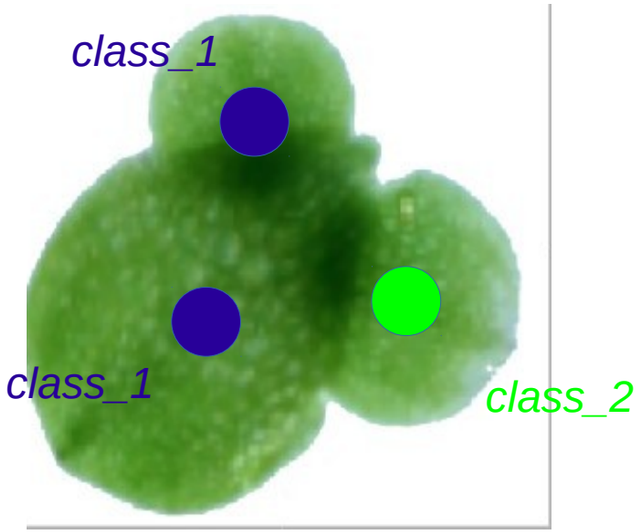


case: reference point is object's characteristic points (e.g. cell nuclei)

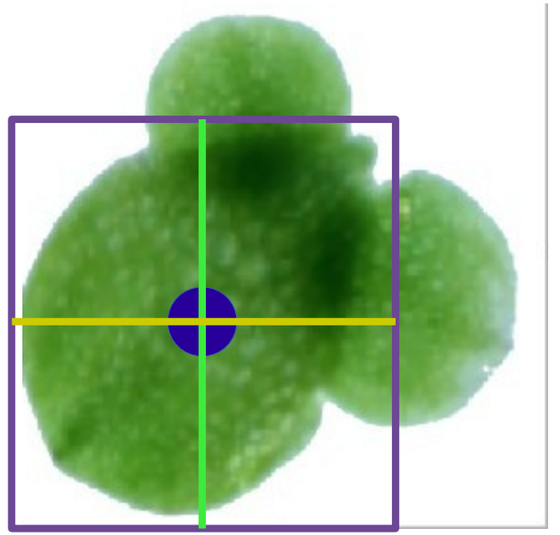
Only one disk is depicted per case

Detection: Encoding

Reference point encodes object-specific information



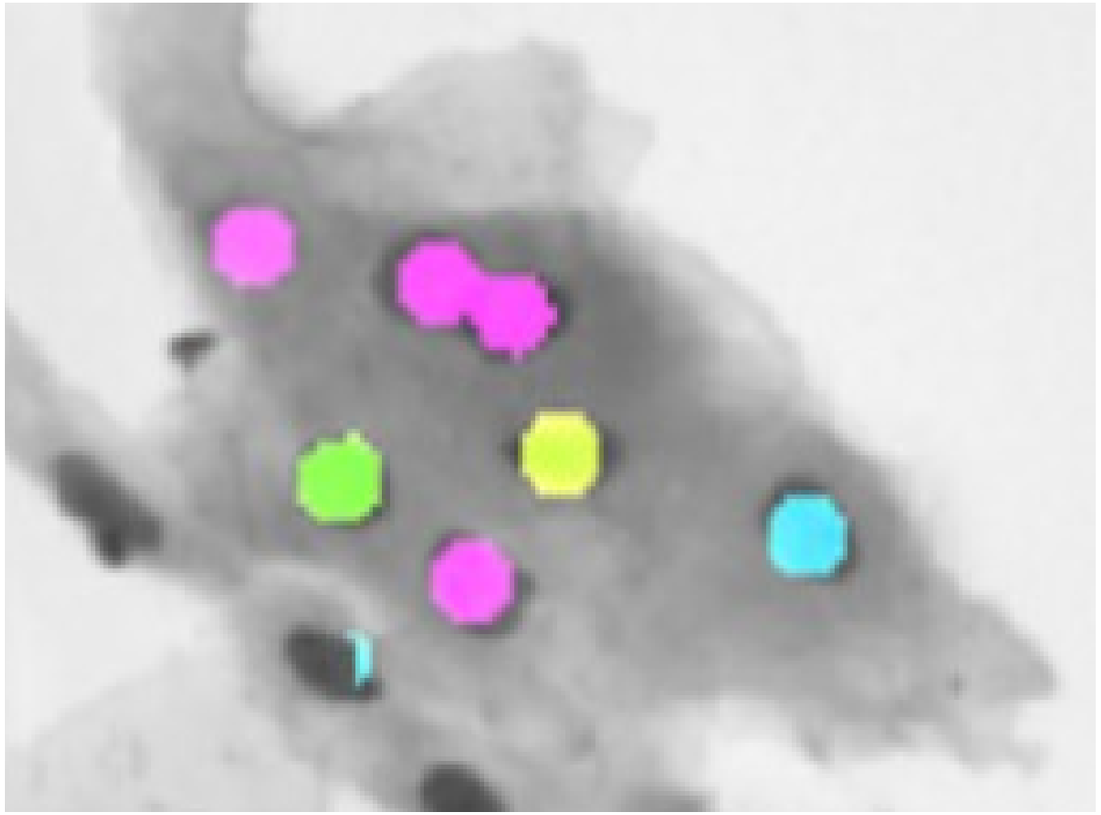
object class encoding



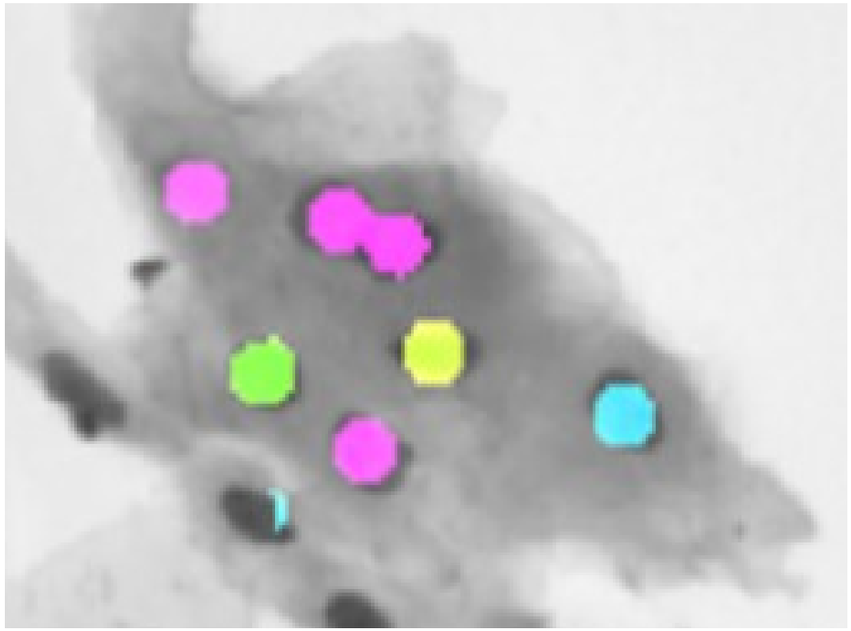
bounding box encoding
w.r.t. reference point.

Detection: Decoding

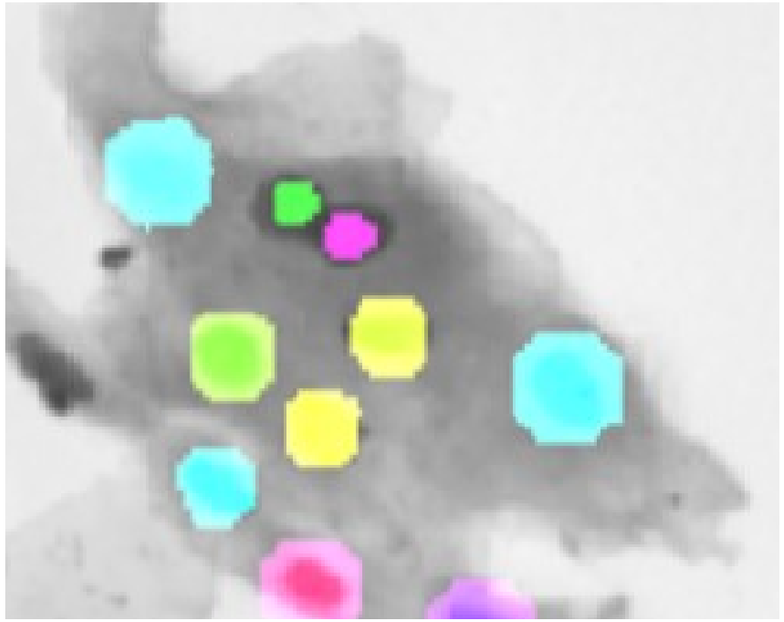
- For decoding, the disks must be spatially separated
- Disks mearging / disappearing leads to false negatives



Detection:
Dynamic disk size adaptation

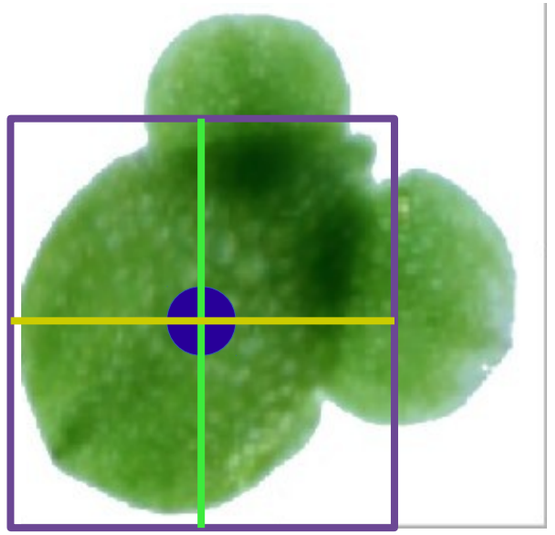


ISOO-V1: disks of fixed size

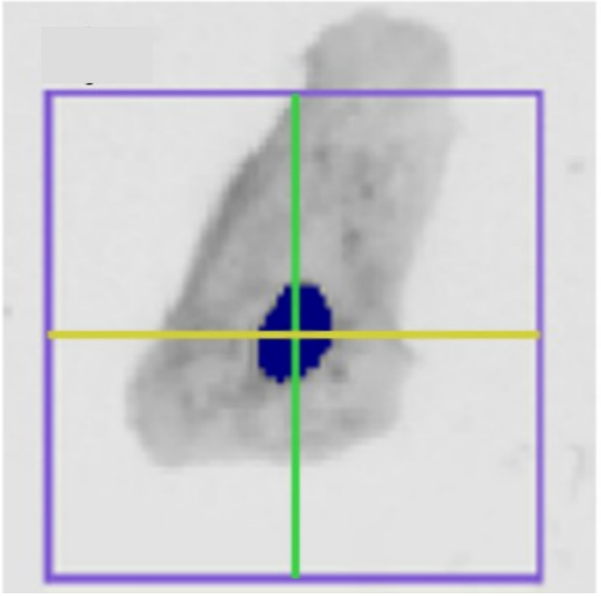


ISOO-V2: disks of adaptive size

Detection:
Reference point location

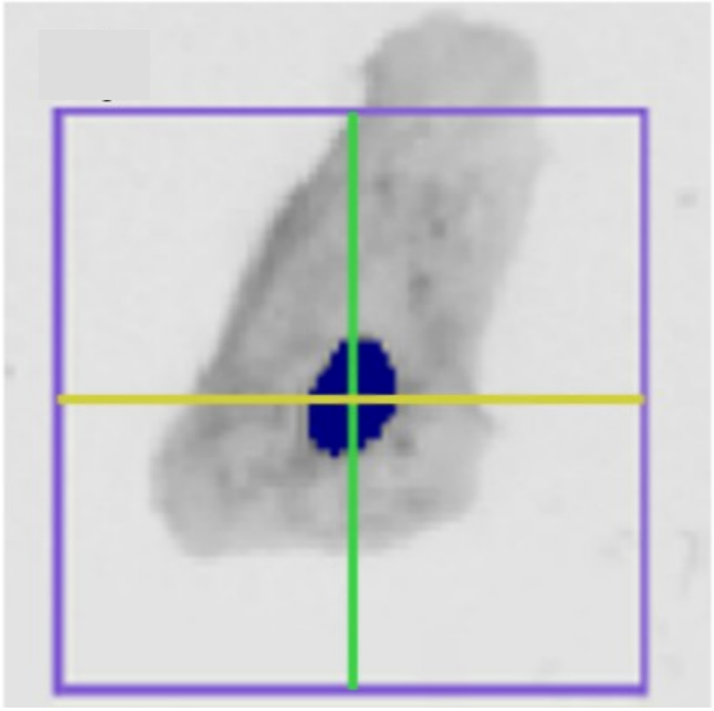


case: reference point is the bounding box center

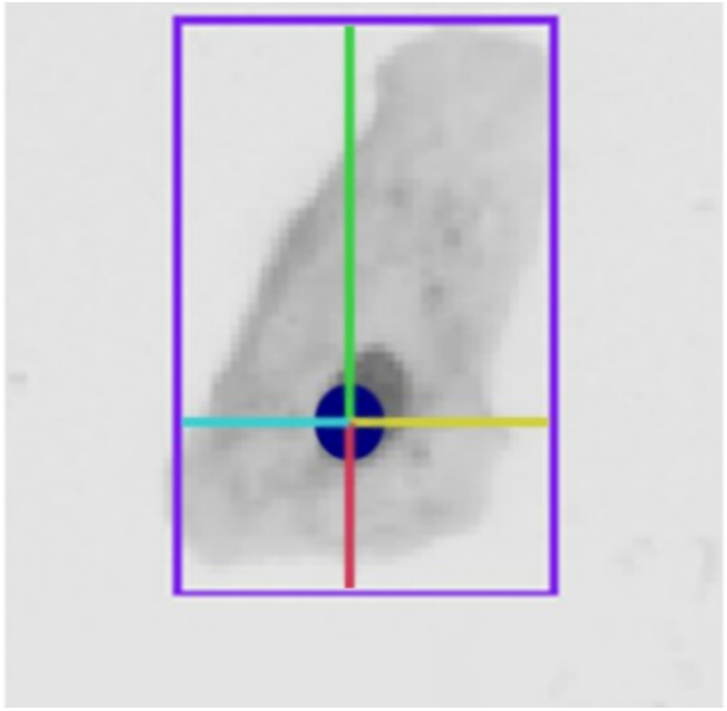


case: reference point is object's characteristic points (e.g. cell nuclei)

Detection:
Bounding box parametrisation



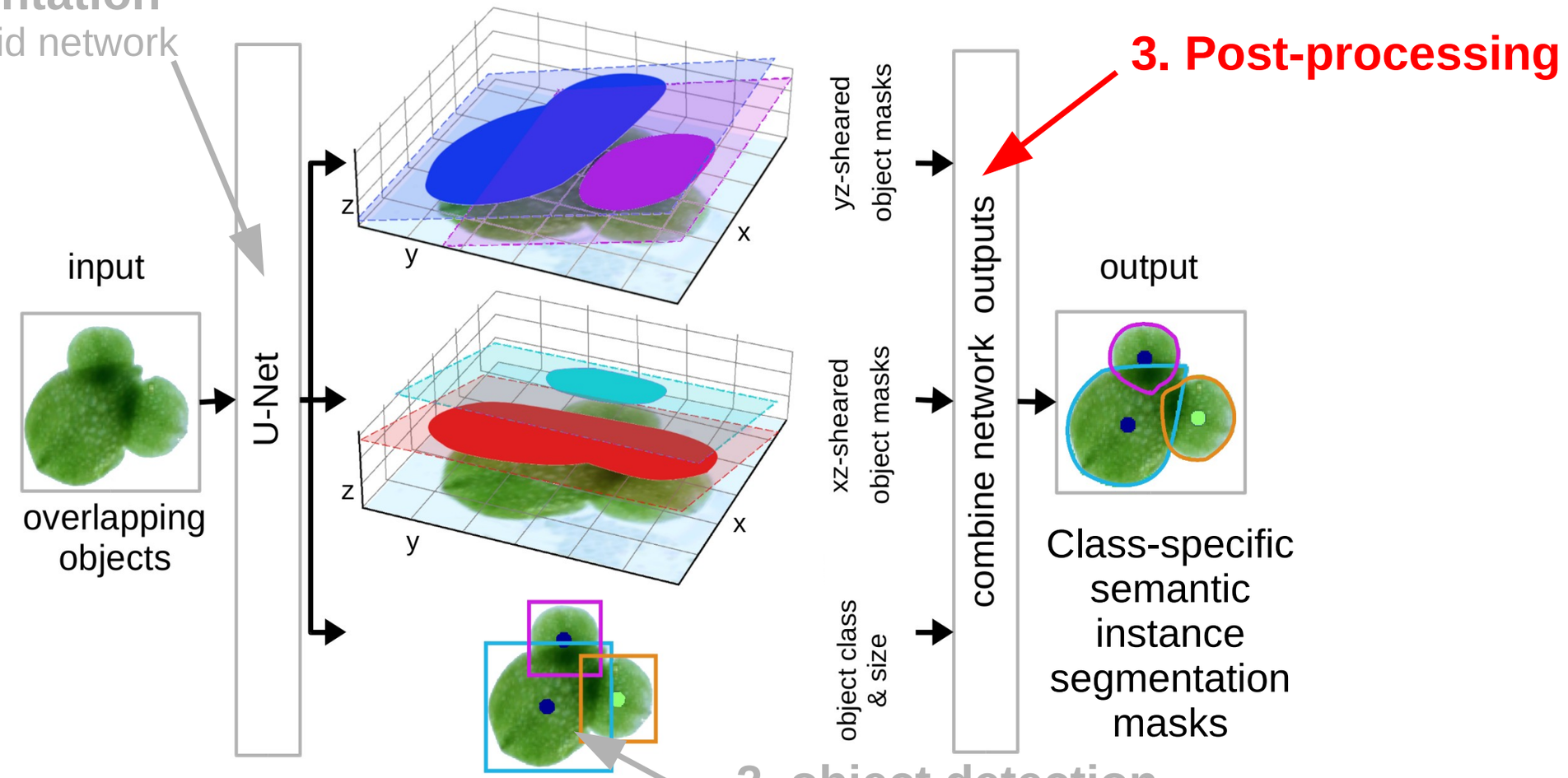
ISOO-V1:
hight (green), width (yellow)



ISOO-V2:
top (green), bottom (red),
right (yellow), left (cyan)

ISOO-V2: Outline

1. Segmentation
2D-3D hybrid network

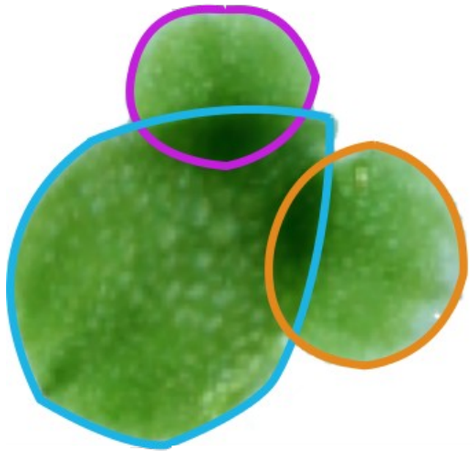


3. Post-processing

2. object detection
reference point with size adaptation
bounding box reparametrization

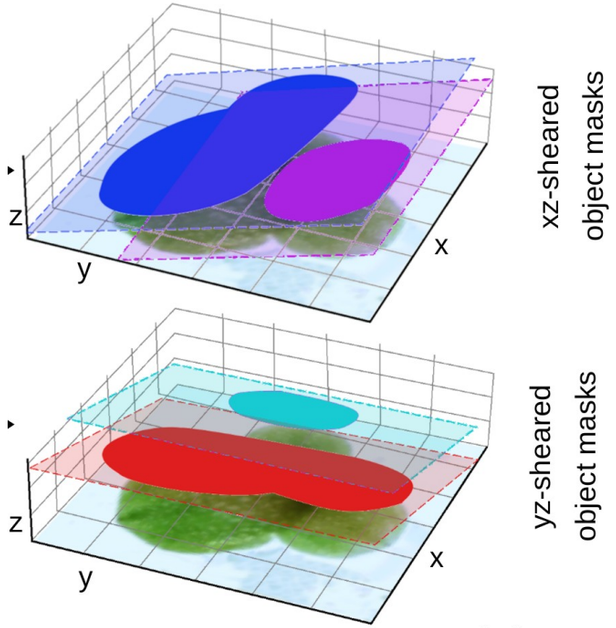
Mask representation: Encoding-Decoding

Transform
(w.r.t object bounding box centers)



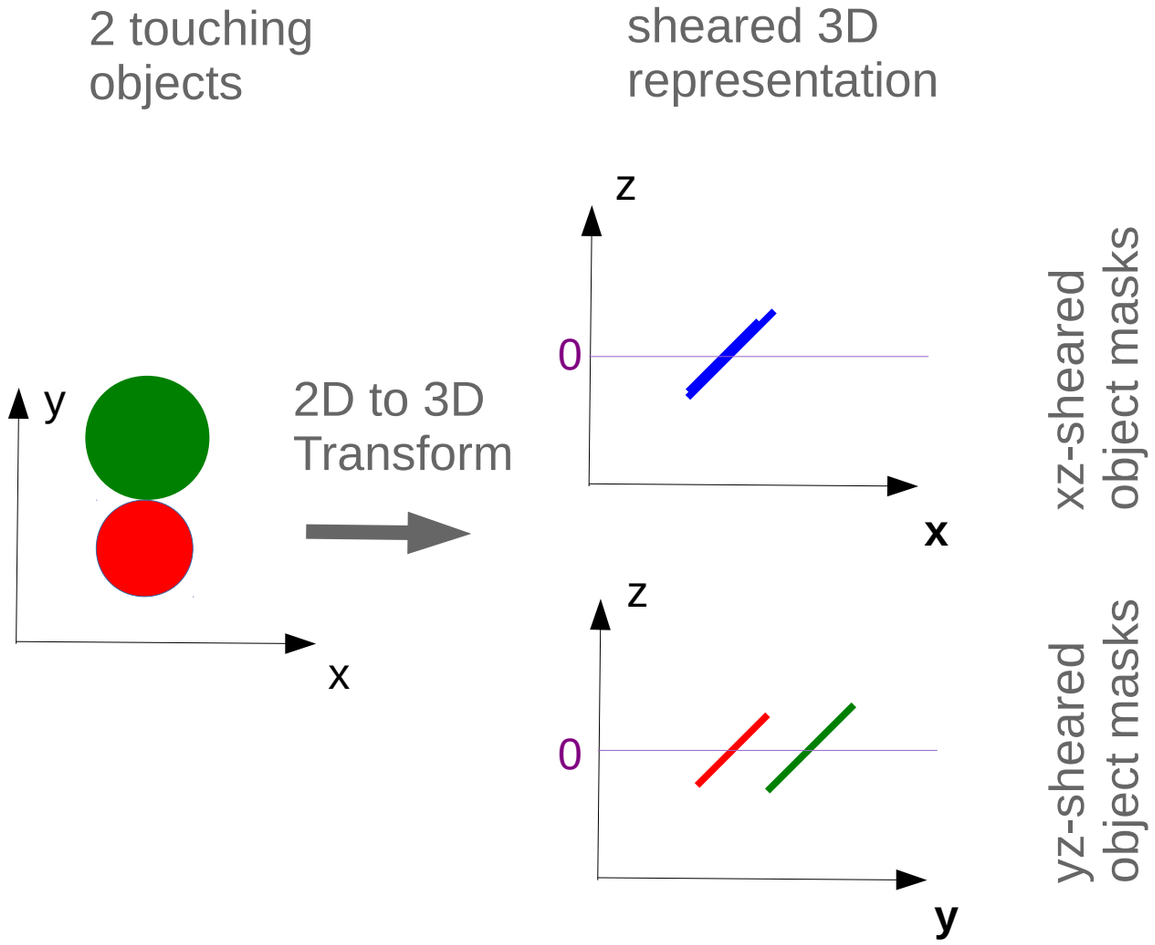
2D
instance
masks

3D
overlap-free
order-independent
masks
encoding



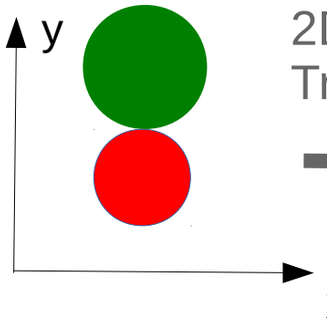
undo Transform
(**Projection**)

Post-processing: Touching objects



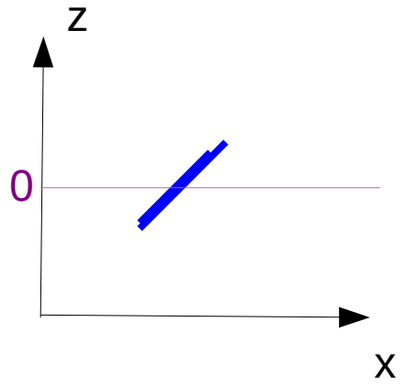
Post-processing: Touching objects

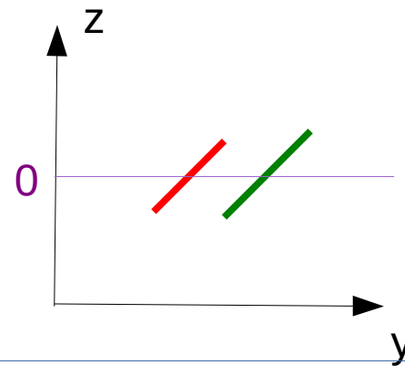
2 touching objects



2D to 3D Transform

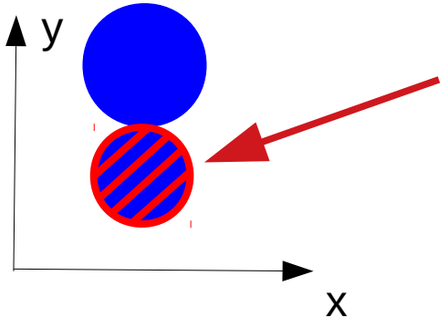
sheared 3D representation





3D to 2D Transform*

overlap (shown as blue-red cercle) of differently shaered components after projection



The desired mask is the intersection of two components

- take parwise connected componens (in the example: blue and red) from differently shaered outputs
- project (sum over z) them to xy-planes
- get the final mask after elementwise multiplikation of the planes

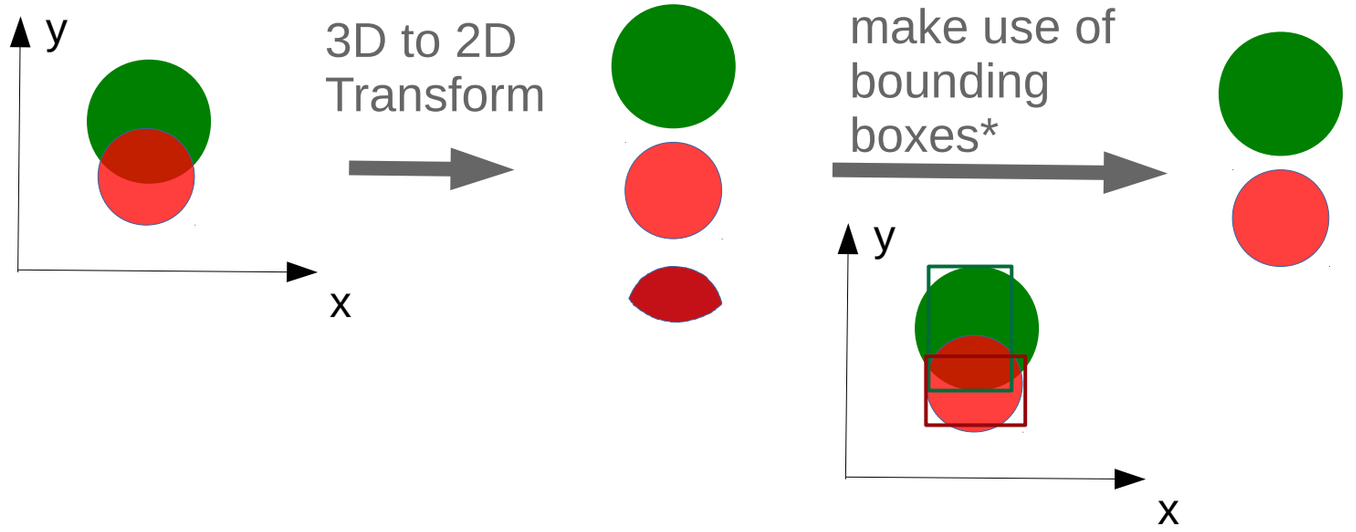
*

Post-processing: Overlapping objects

2 overlapping objects

Projection results in
- 2 desired masks
- 1 false positive

Final result

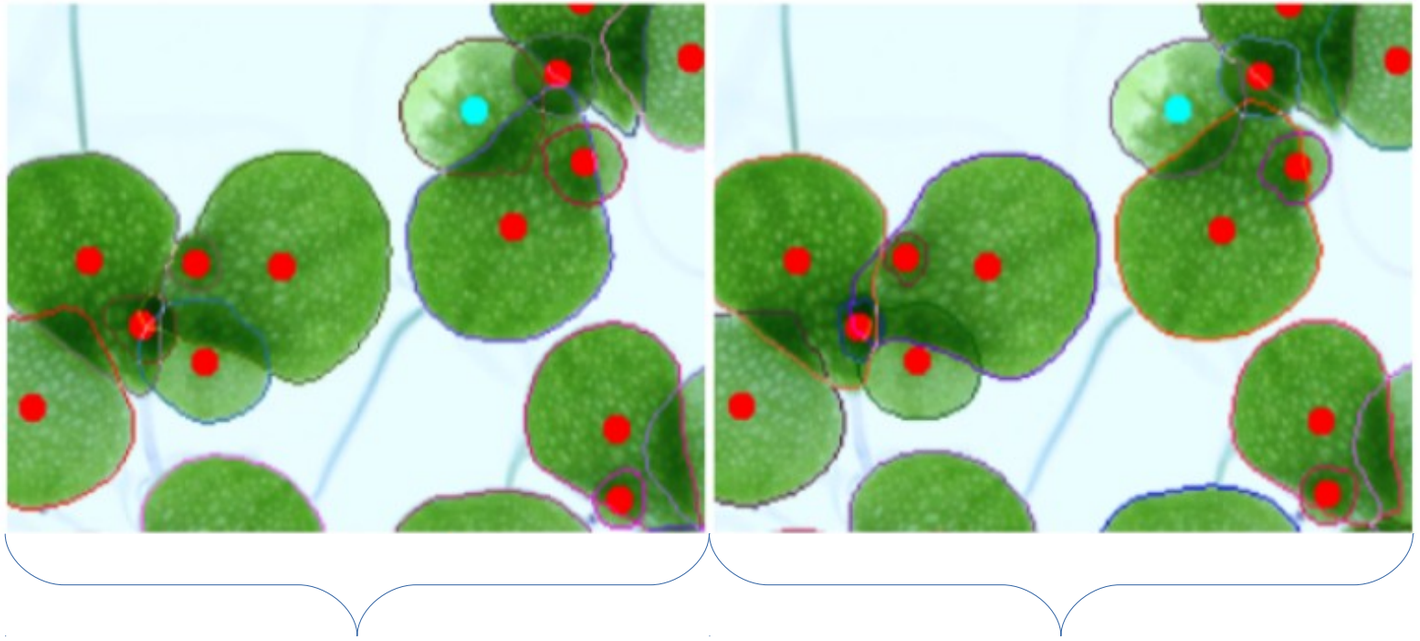


*
- find object mask - bounding box correspondences (maximum IoU)
- sort out the masks without bounding box

Resuts:
Complete pipeline

Data set: Duckweed*

- big images (1152x1728x3 pixels)
- high object size diversity
- many objects per image (max. 170 objects / image)
- Two unbalanced object classes: healthy (red disks) and deseased (cyan disks)



ground truth

predictions

** Böhm et al, ISBI, ‘18*

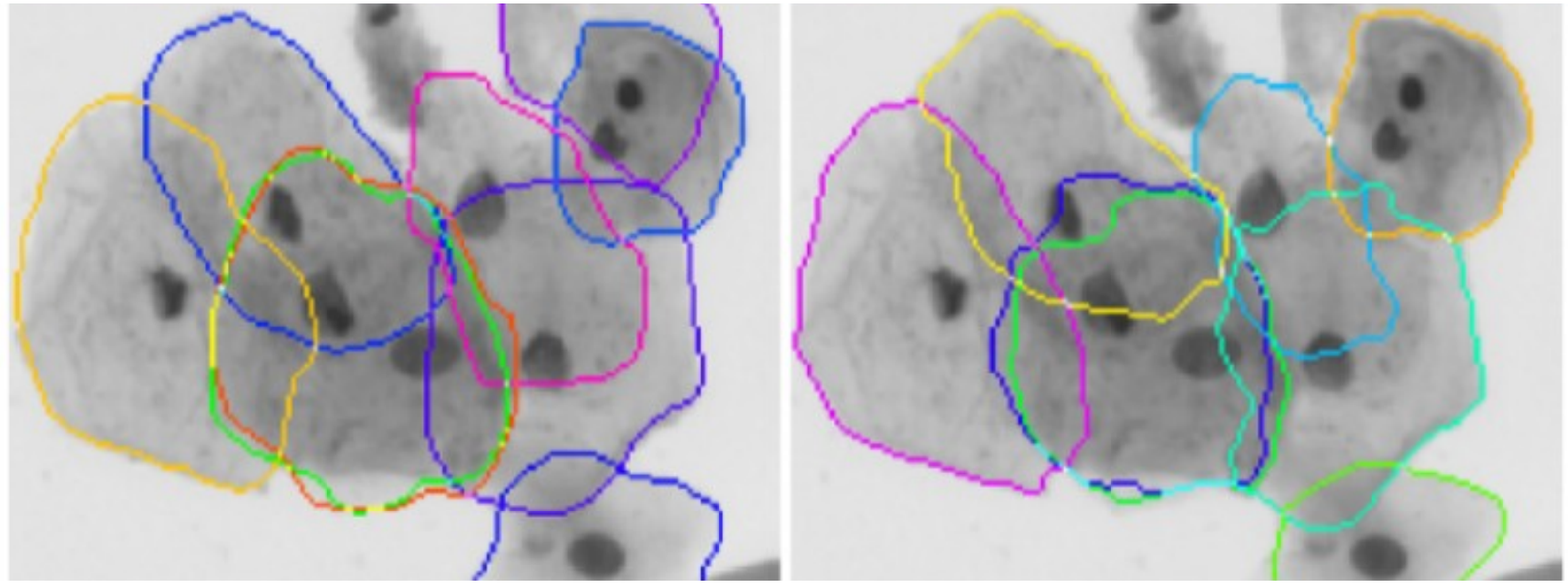
Data set: Duckweed

	DC (Mean±SD)	**oFN** (Mean±SD)	**pTP** (Mean±SD)	**pFP** (Mean±SD)
ISOO-V1 (w/o cl) Böhm et al, ISBI, ‘18	.929±.057	.129±.067	.939±.066	.000±.000
ISOO-V2 (w/o cl)	**.945±.051**	**.103±.062**	**.953±.063**	**.000±.000**

object-based **F**alse **N**egative rate;
Dice **C**oefficient; **p**ixel-based **T**rue **P**ositive, **F**alse **P**ositive rate

Data set: OSC-ISBI*

- Highly overlapped objects
- Reference points are set to the cell nuclei



ground truth

predictions

*The Second Segmentation of Overlapping Cervical Cells

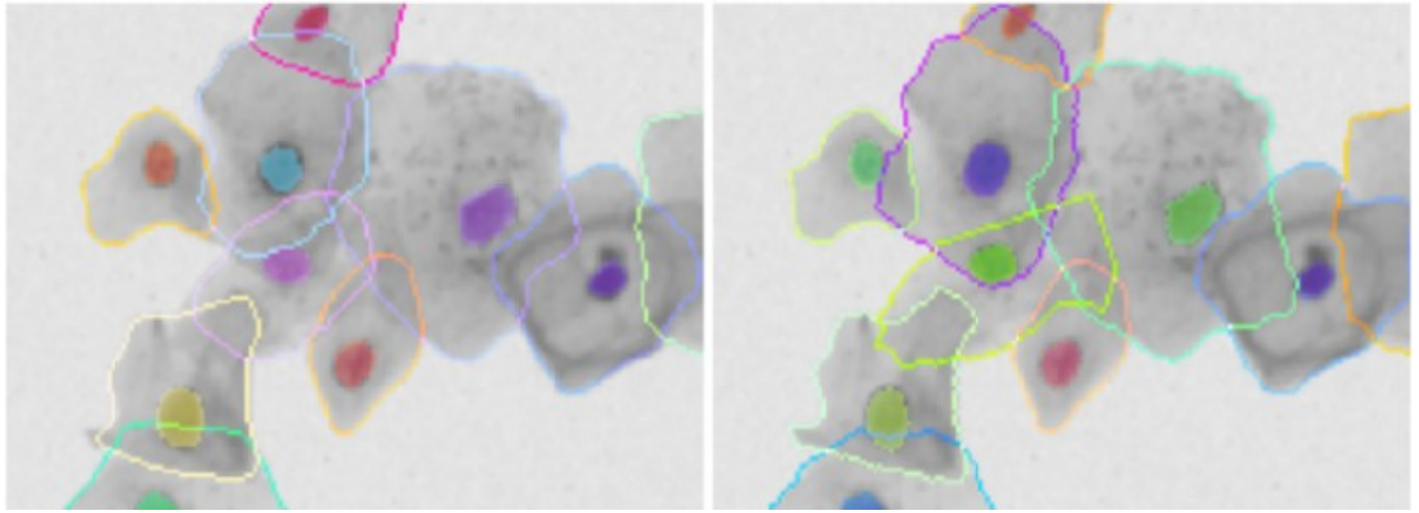
Data set: OSC-ISBI

	DC (Mean±SD)	**oFN** (Mean±SD)	**pTP** (Mean±SD)	**pFP** (Mean±SD)
Phoulady et al. *ISBI, ‘15*	.831±.079	.408±.163	**.927±.098**	.003±.002
Ramalho et al. *ISBI, ‘15*	.856±.078	.501±.180	.899±.113	.002±.001
Lee et al. *CVPR, ‘16, ws*	.879±.087	.434±.168	.877±.123	.001±.001
ISOO-V1 *Böhm et al, ISBI, ‘18*	.863±.074	.370±.141	.895±.107	.001±.001
ISOO-V2	**.895±.079**	**.290±.151**	.901±.108	**.001±.001**

object-based **F**alse **N**egative rate;
Dice **C**oefficient; **p**ixel-based **T**rue **P**ositive, **F**alse **P**ositive rate

Data set: OSC-ISBI-S*

- object subpart (semantic) segmentation: cell body (contours) and nucleus
- object contours and the corresponding sub-parts have the same color



ground truth predictions

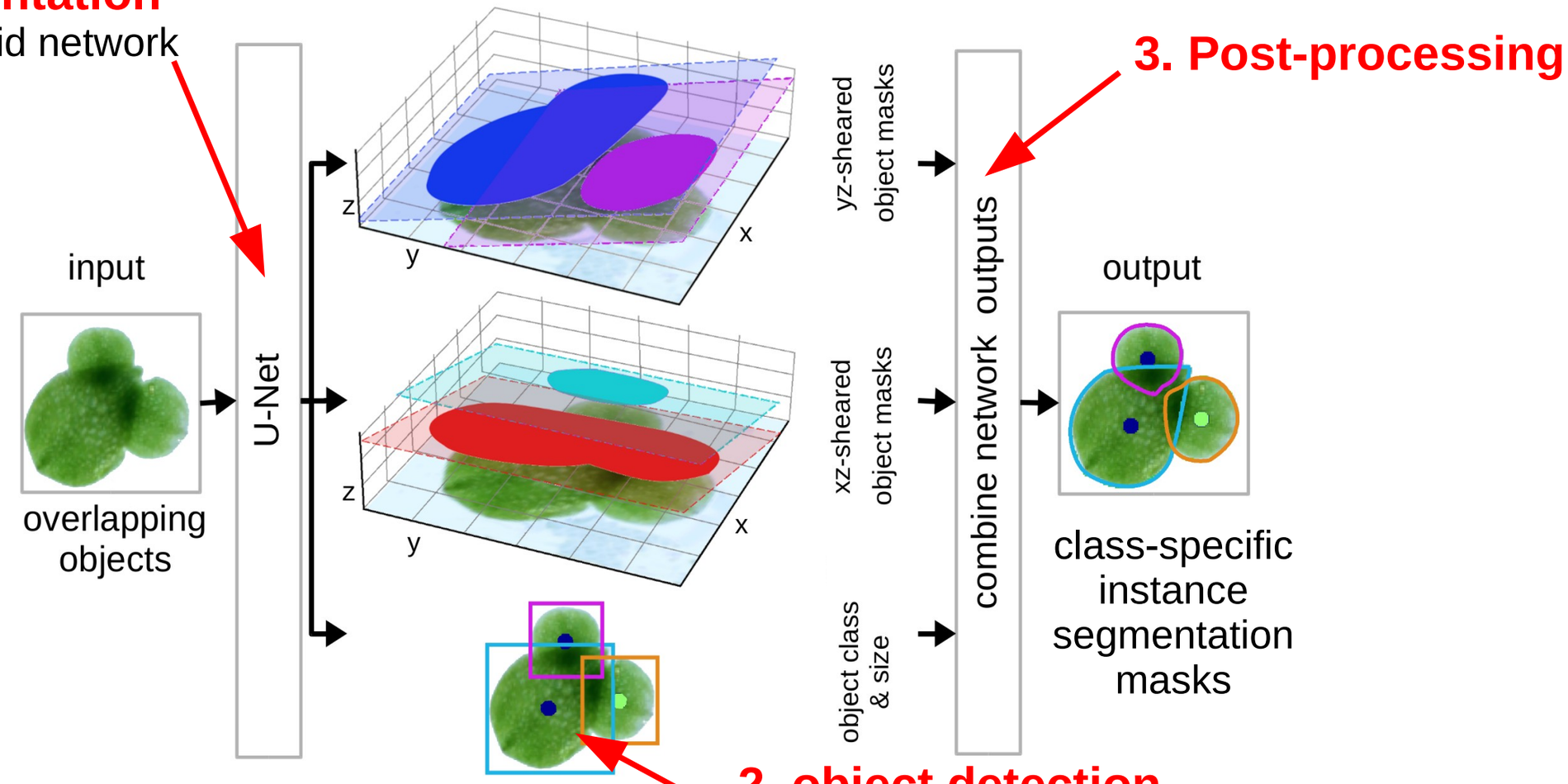
	DC
ISOO-V2 cell body (nuclei)	.924(.848) ±.069(.173)

**The first segmentation of overlapping cervical cells*

Thank you!

Questions?

1. Segmentation
2D-3D hybrid network



3. Post-processing

2. object detection
reference point with size adaptation
bounding box reparametrization