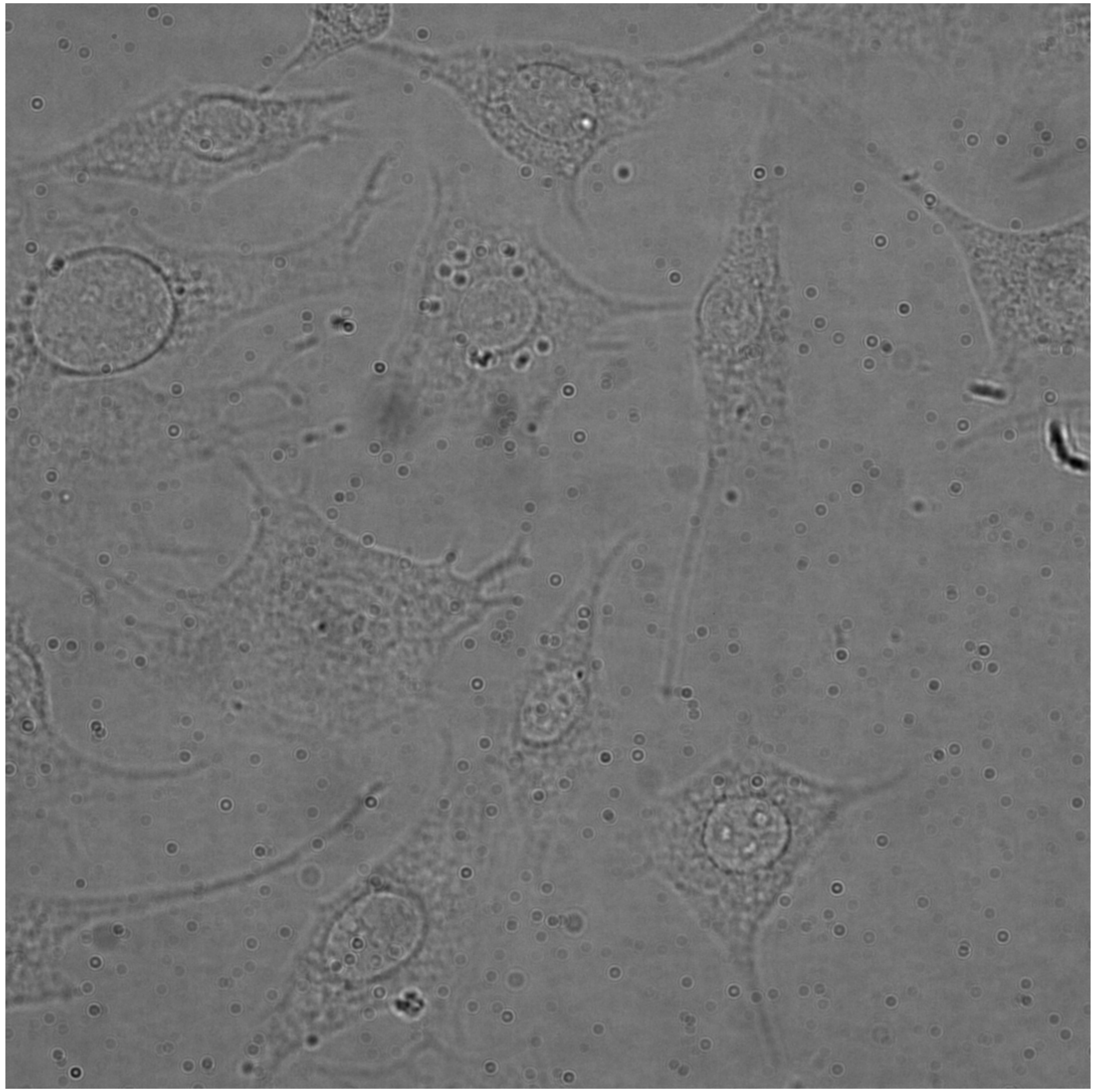
# IAUNet: Instance-Aware U-Net

Yaroslav Prytula[1,2] Illia Tsiporenko[1] Ali Zeynalli[1] Dmytro Fishman[1,3]

[1] Institute of Computer Science, University of Tartu
[2] Ukrainian Catholic University, [3] STACC OÜ, Tartu, Estonia

# Brightfield



**Microscopy** Image

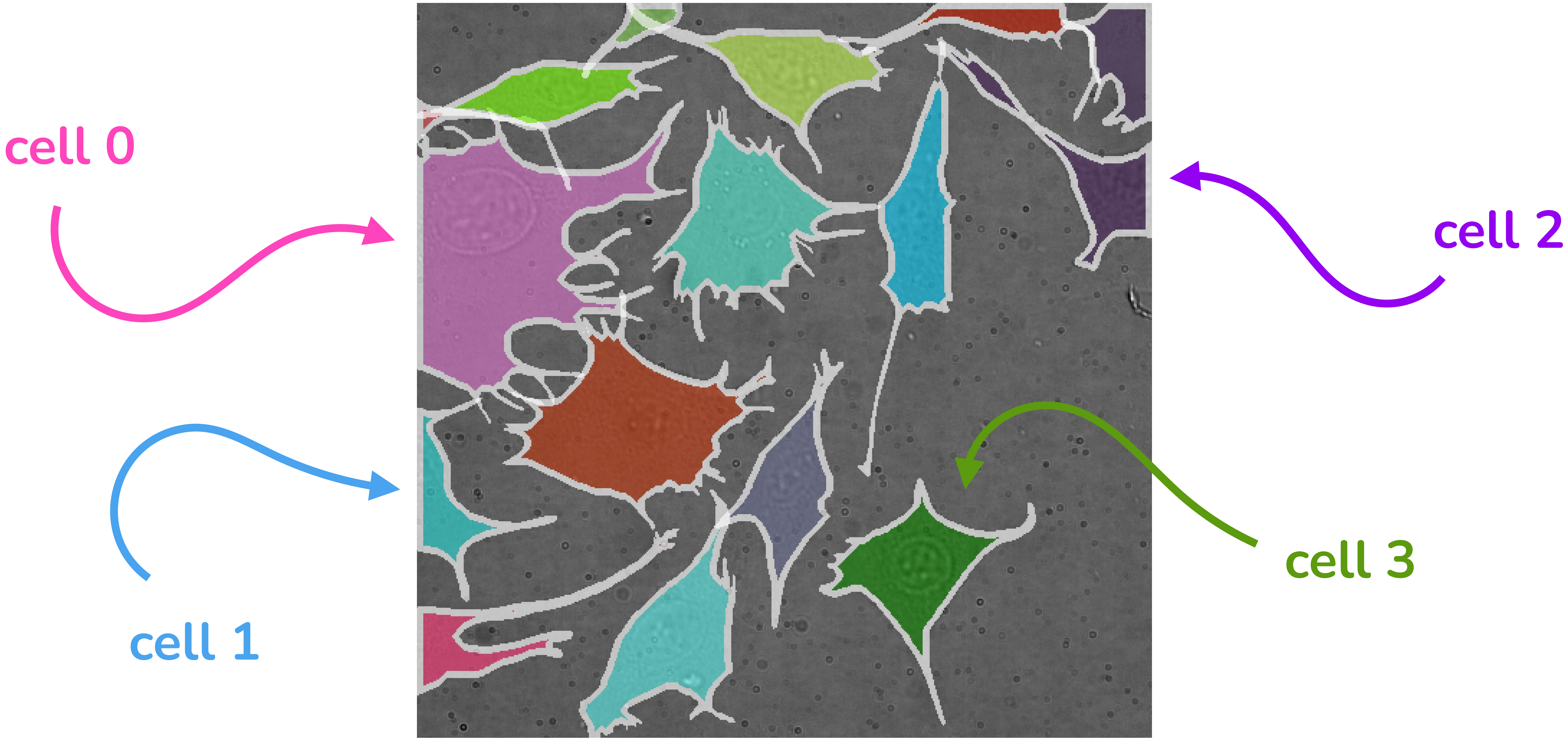
## Brightfield

Pixels of class
"cell"

Pixels of class
"cell"

**Semantic** Segmentation

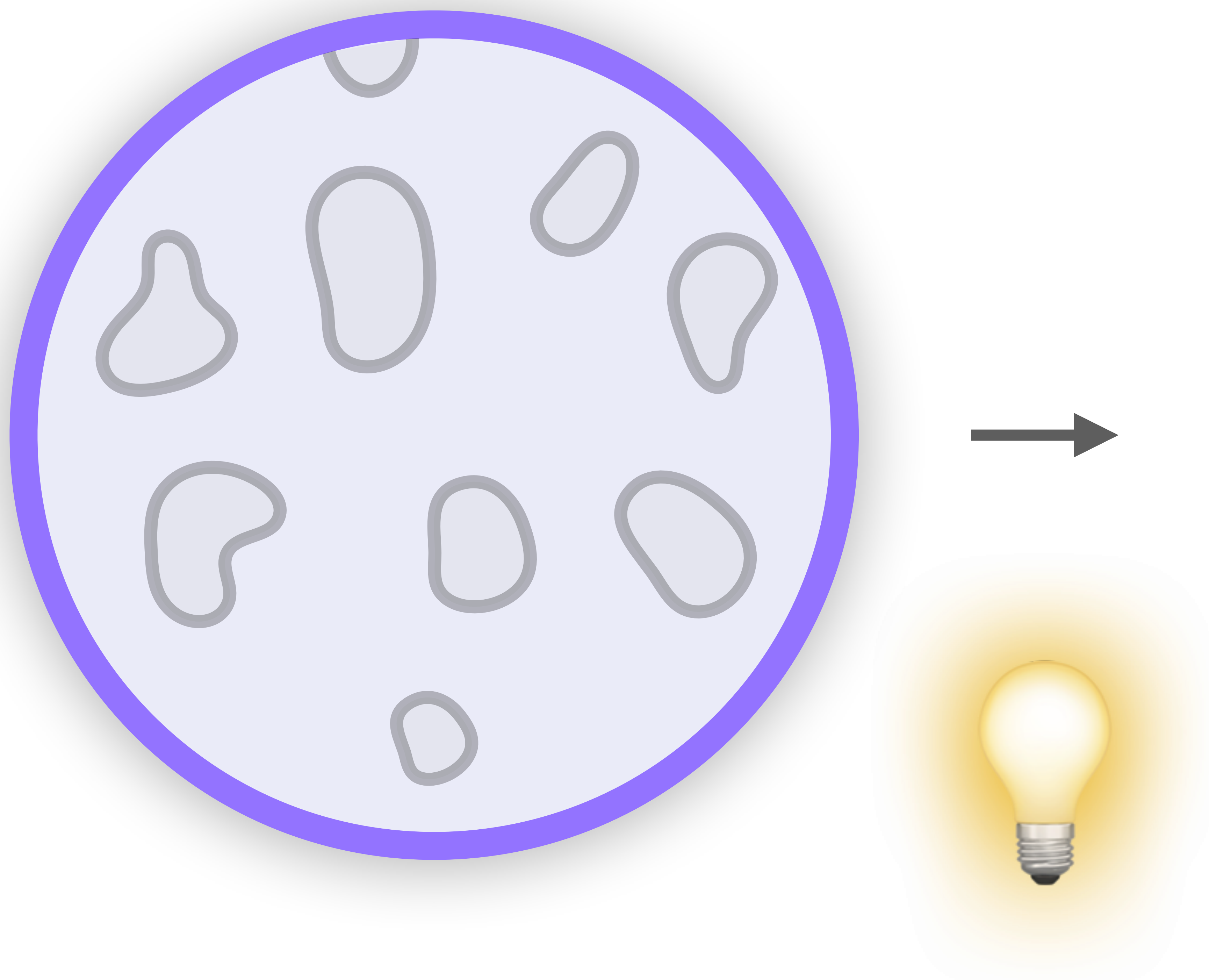
## Brightfield

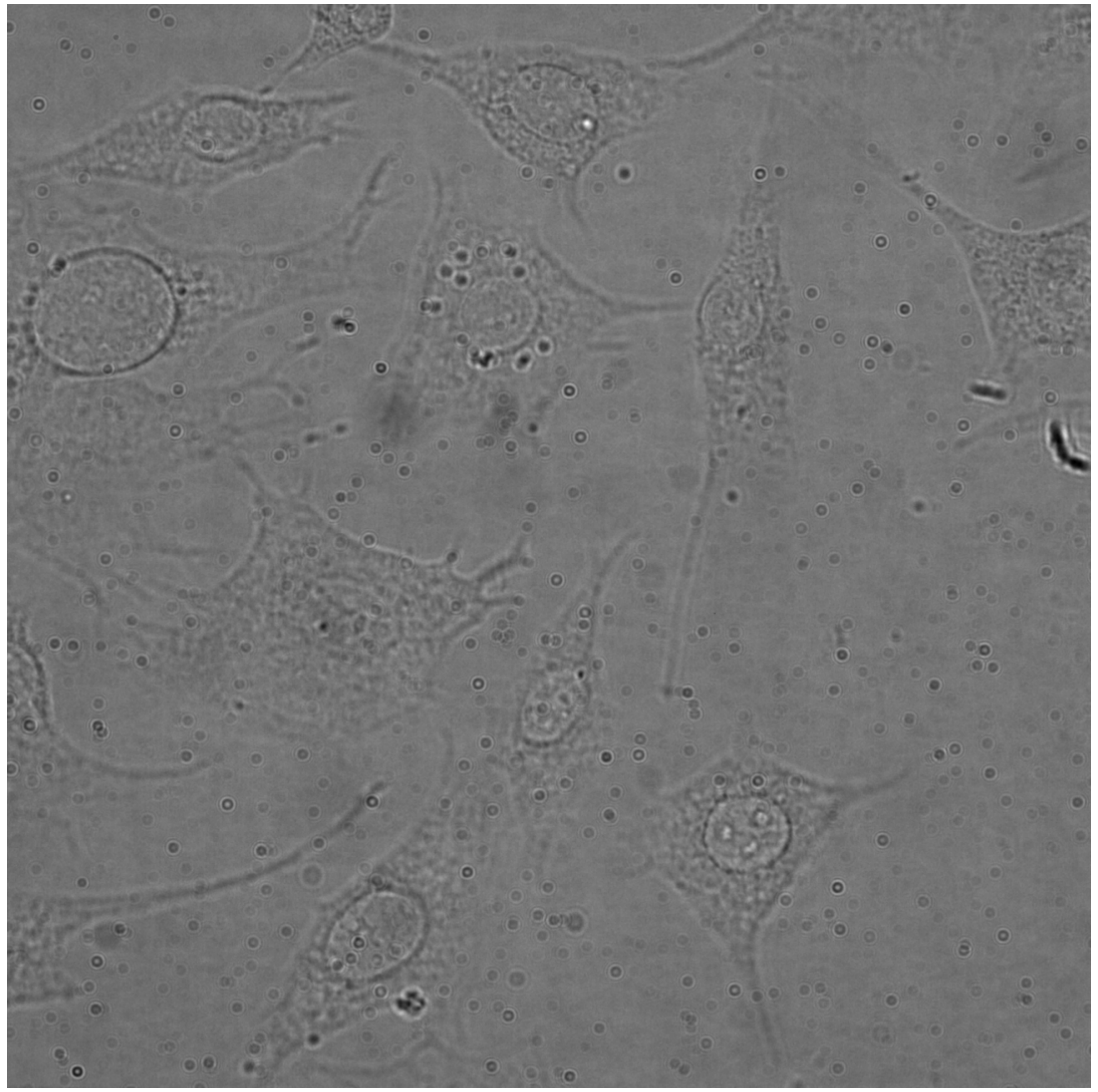



**Instance** Segmentation

# Motivation

**Brightfield** captures images by transmitting standard **white light** through the sample

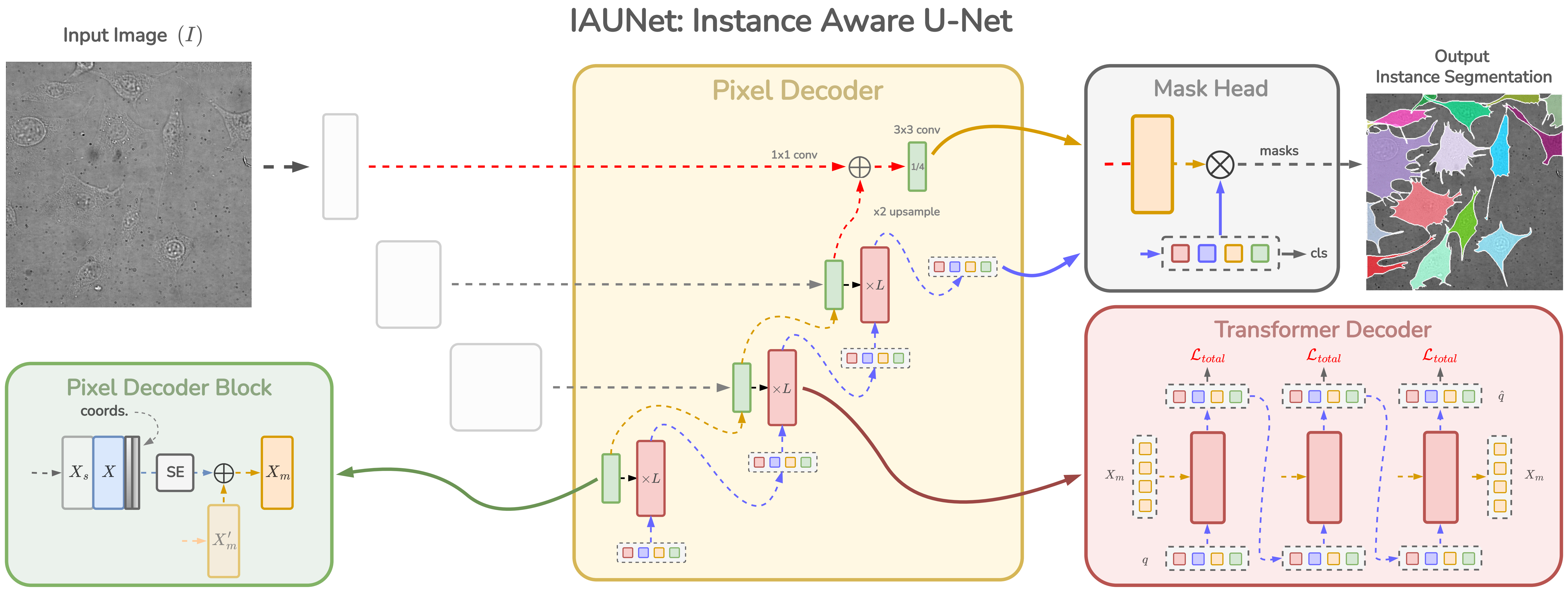


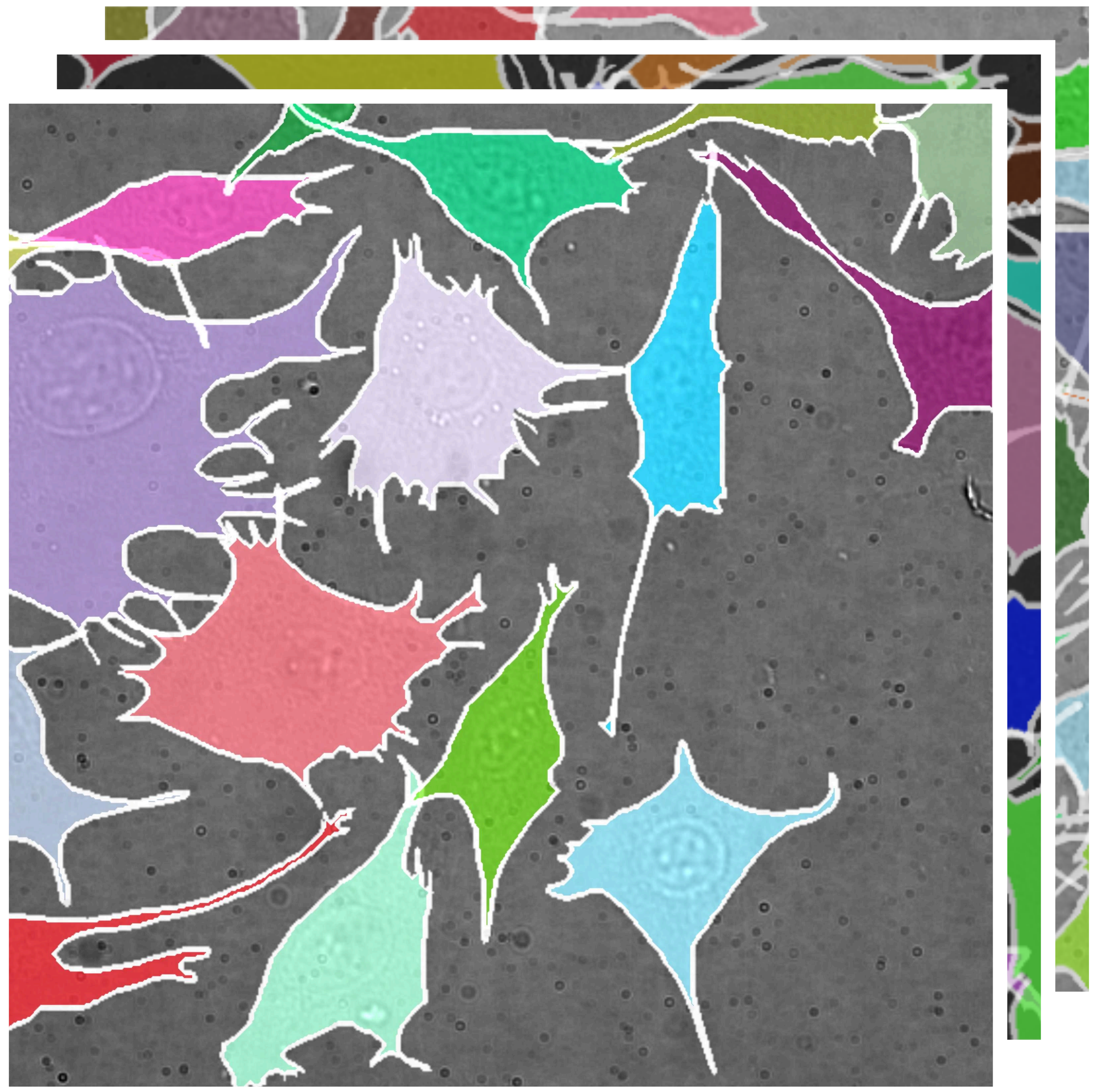
https://www.revvity.com/product/opera-phenix-plus-system-hh14001000

# Contributions

## IAUNet




## Revvity-25

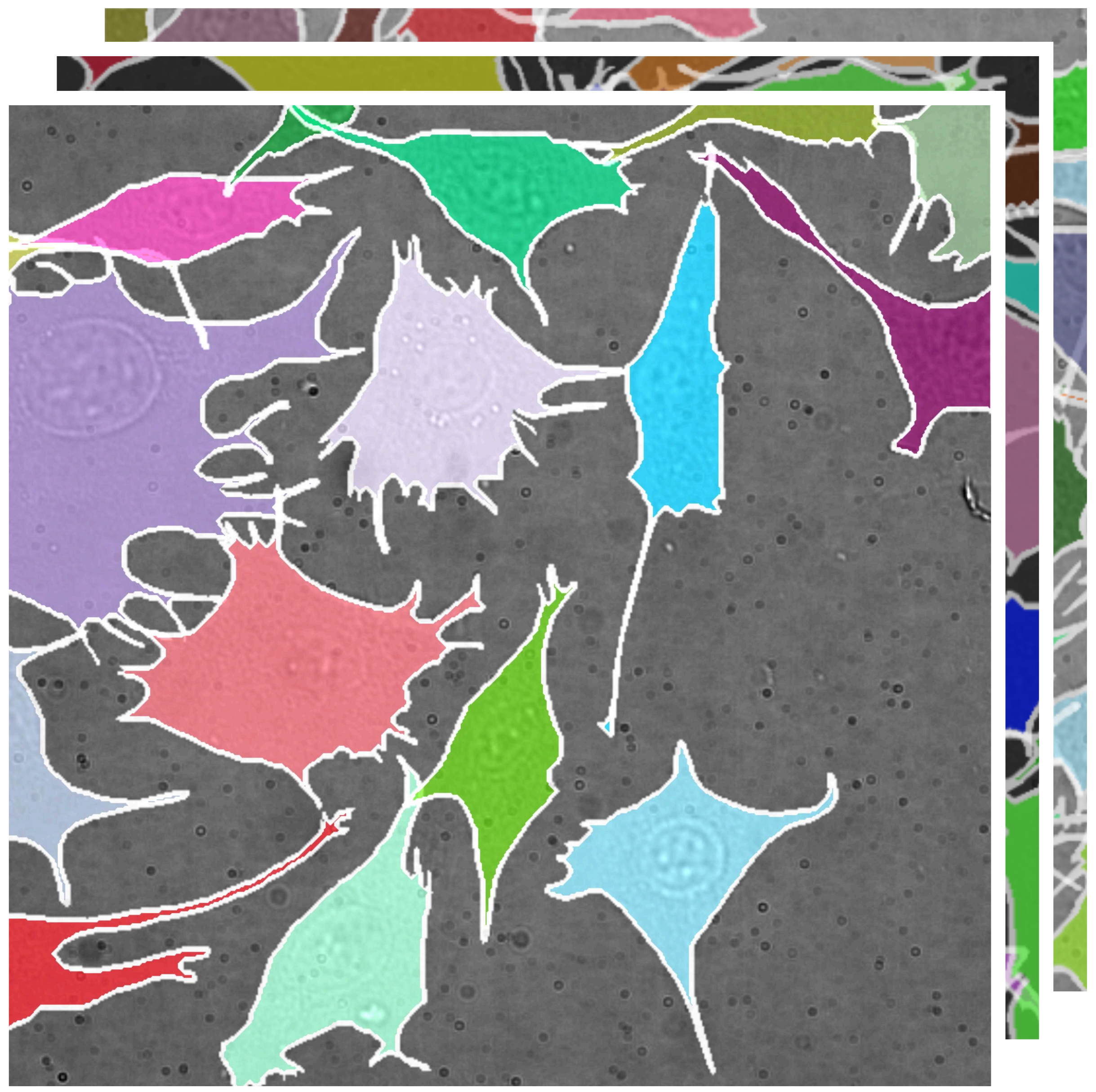


# Contributions

## IAUNet



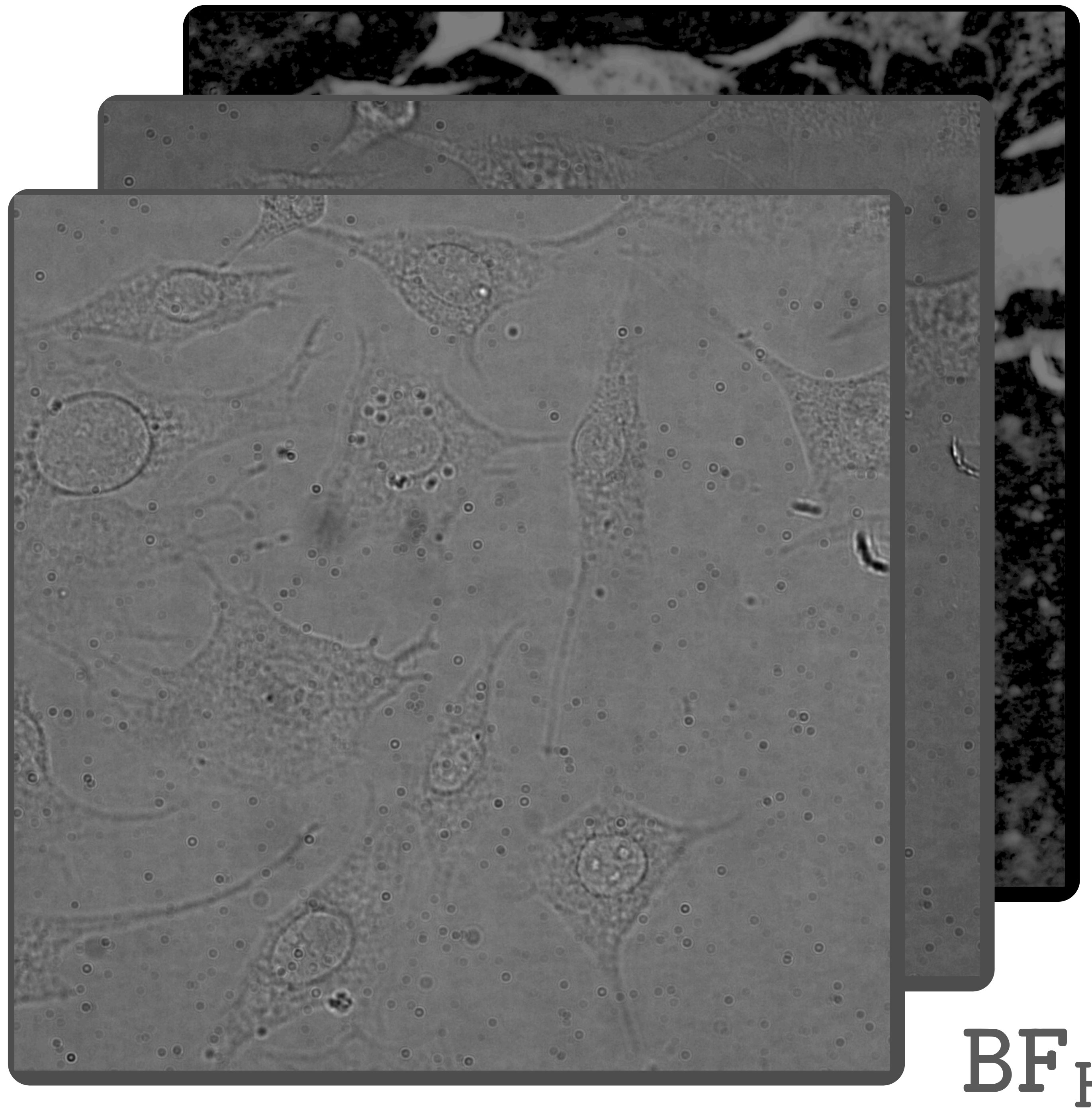

## Revvity-25




LIVECell
EVICAN
ISBI2014
Revvity-25
❌ No overlaps
❌ Missing annotations
✅ Many instances
✅ Large Dataset
✅ High visual complexity
❌ No overlaps
❌ Coarse annotations
❌ Few instances
✅ Large Dataset
❌ Low visual complexity
✅ Overlaps
❌ Simple annotations
❌ Few instances
❌ Small Dataset
❌ Low visual complexity
✅ Overlaps
✅ Precise annotations
✅ Many instances
❌ Small Dataset
✅ High visual complexity

# Revvity-25

Raw Images



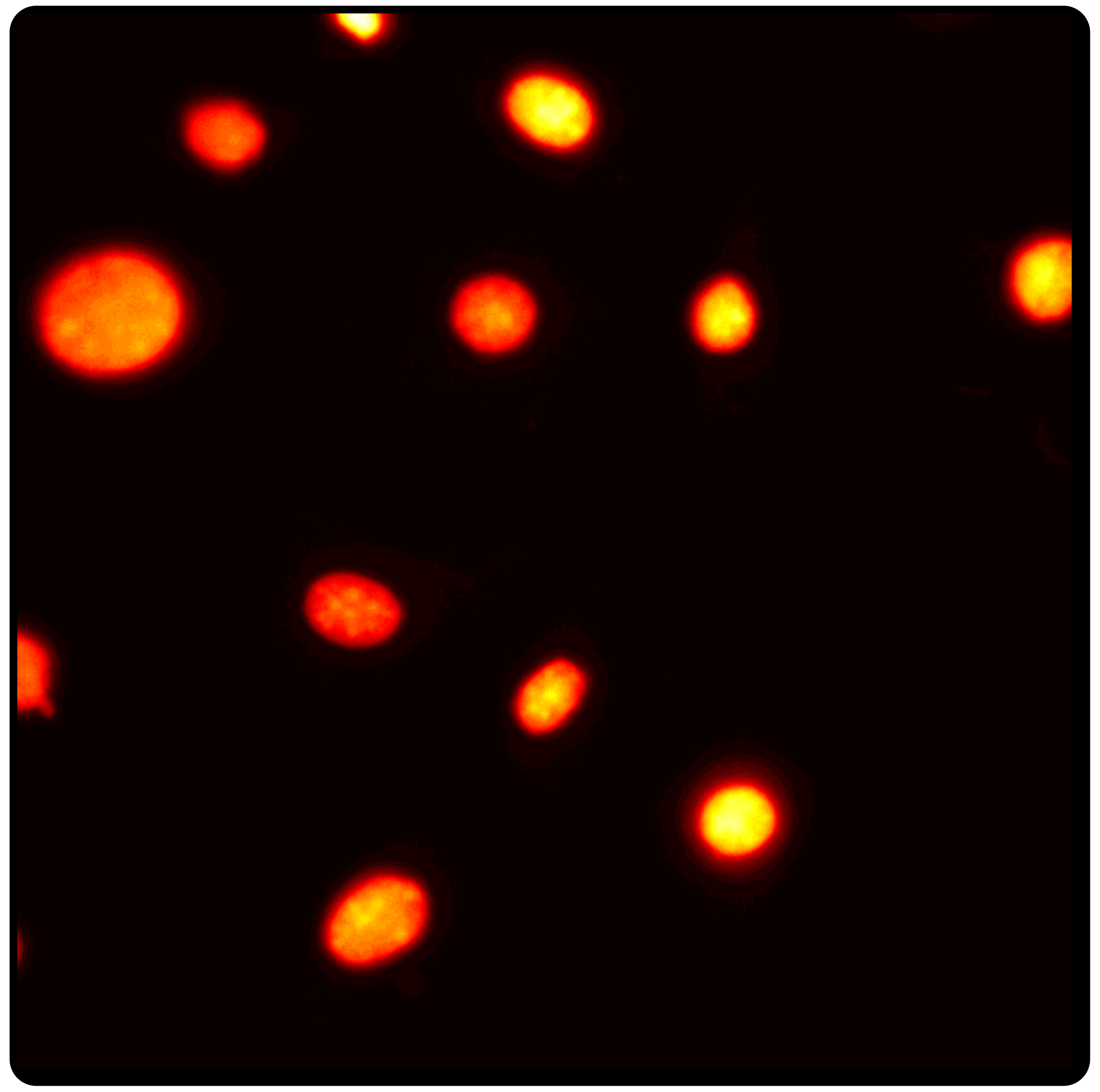
PC

$BF_{HI}$

$BF_{LO}$

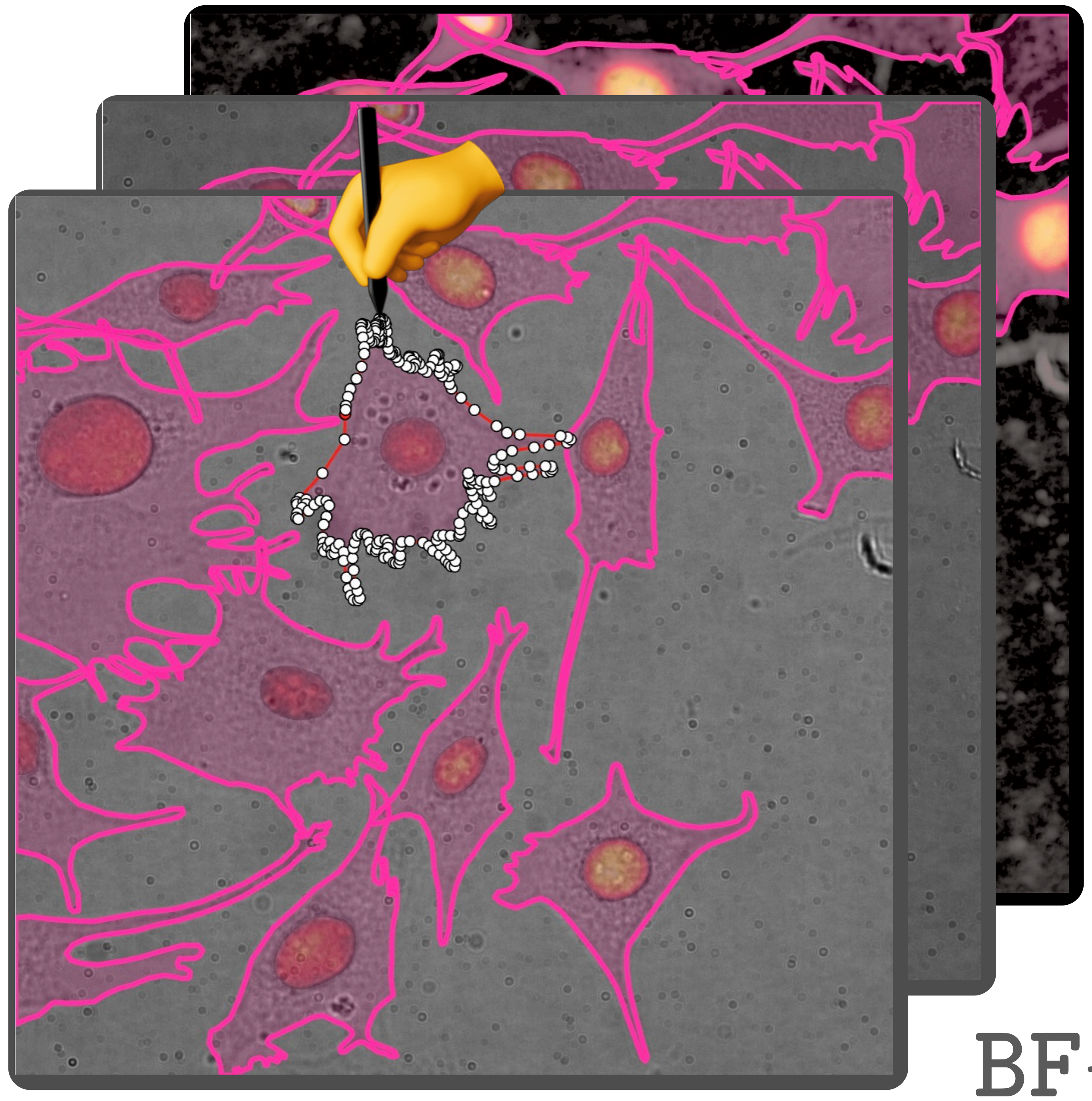
+

FL



=

Label Studio



PC

$BF_{HI}$

$BF_{LO}$

# Revvity-25

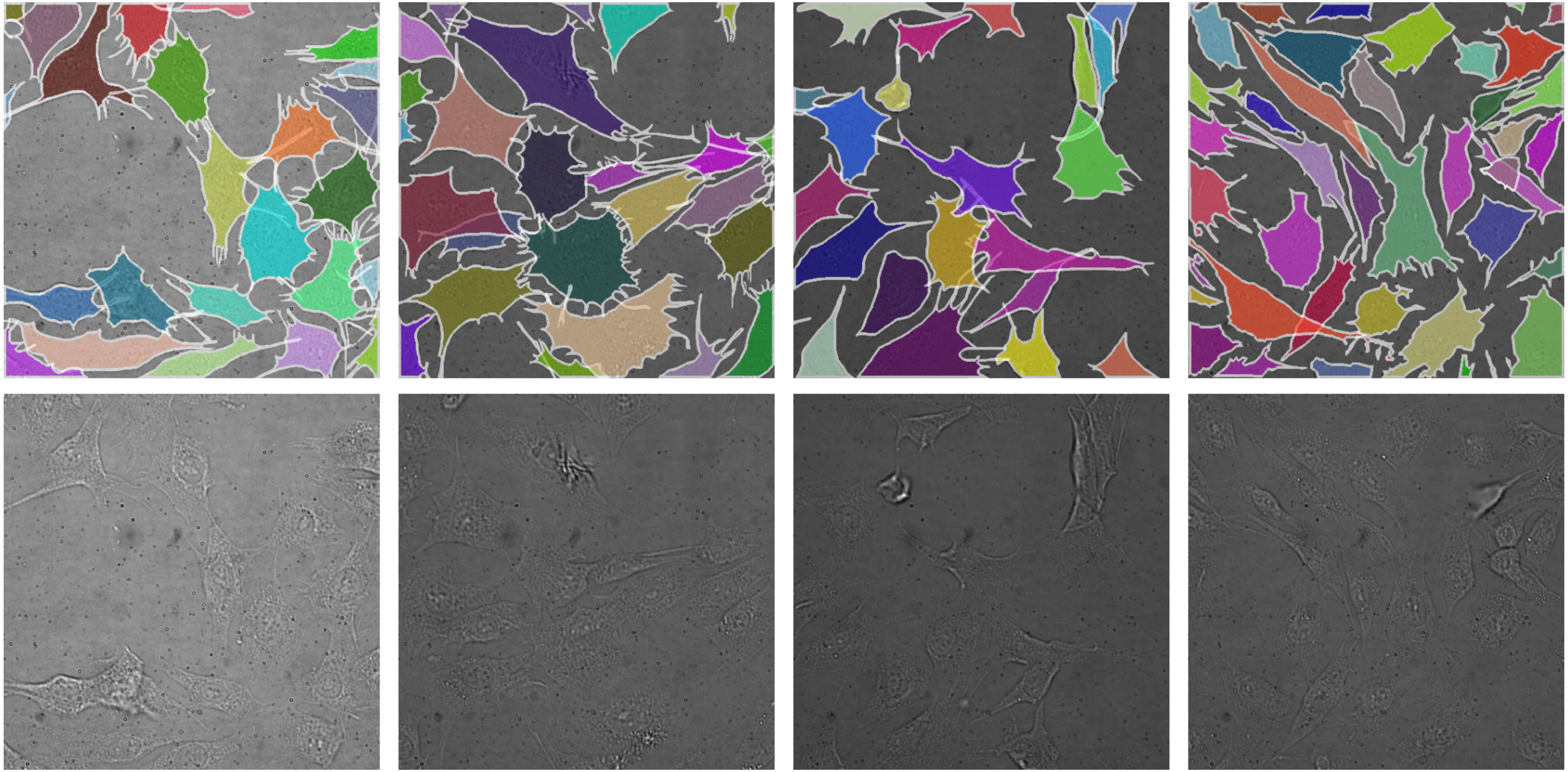
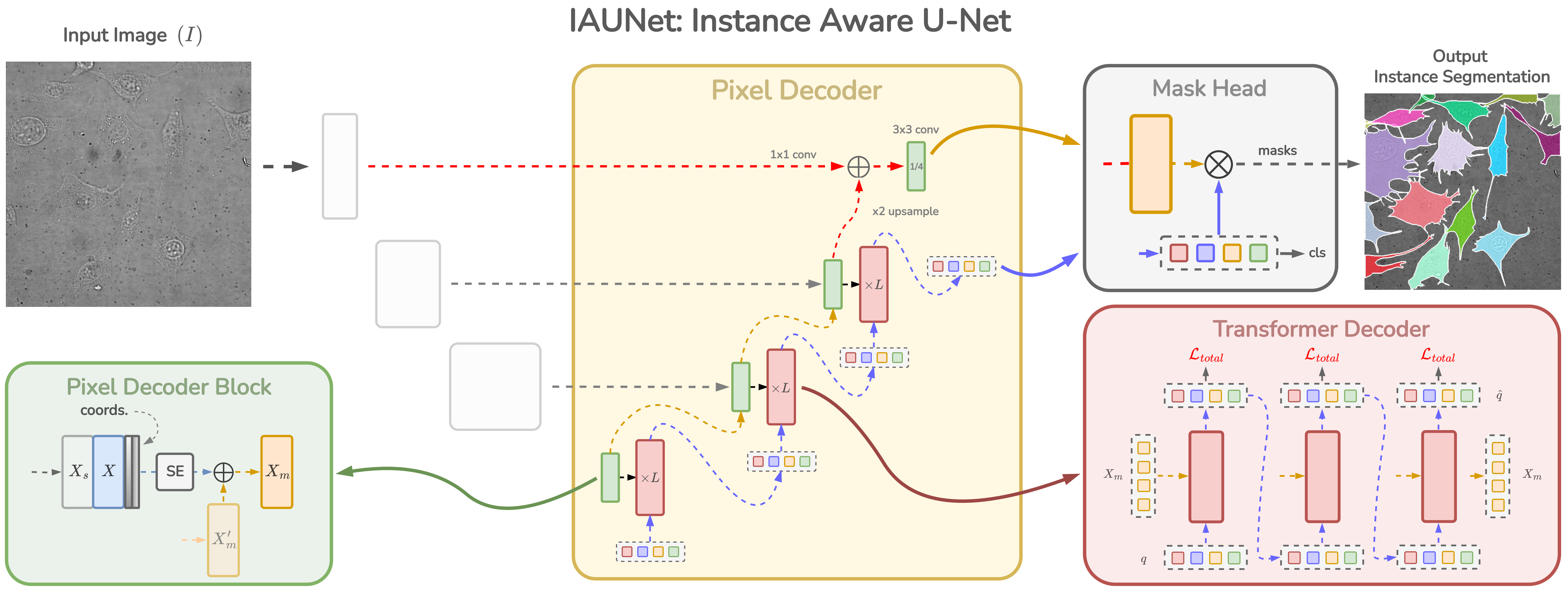


Figure 1. Revvity-25 Dataset.

# Contributions

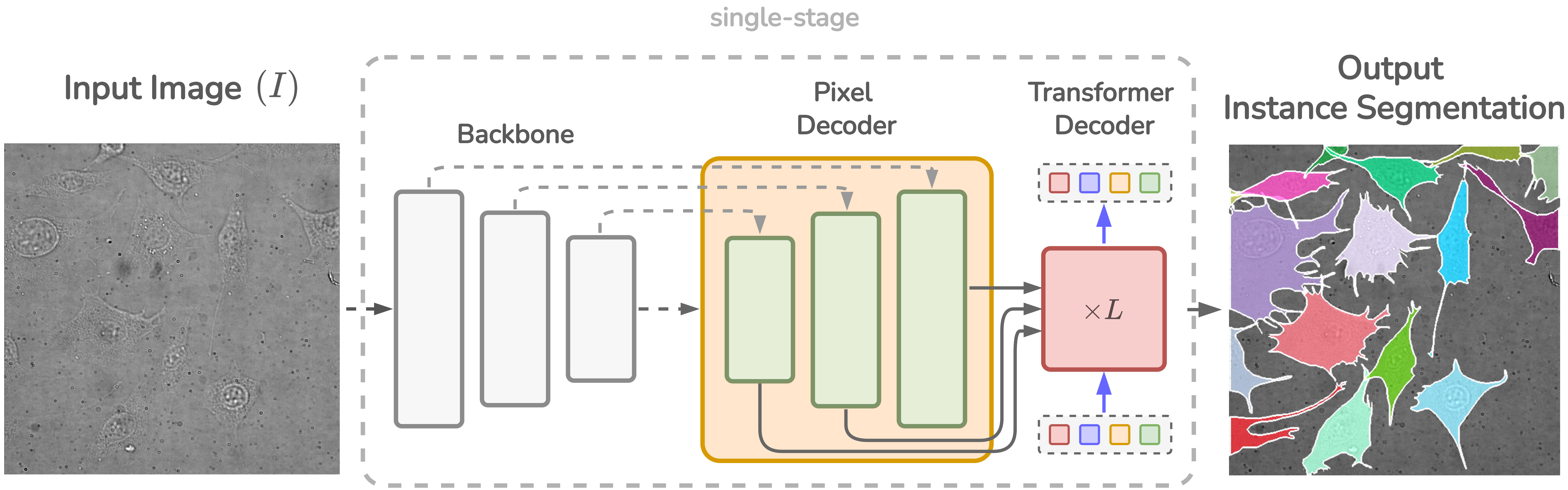
## IAUNet




## Revvity-25

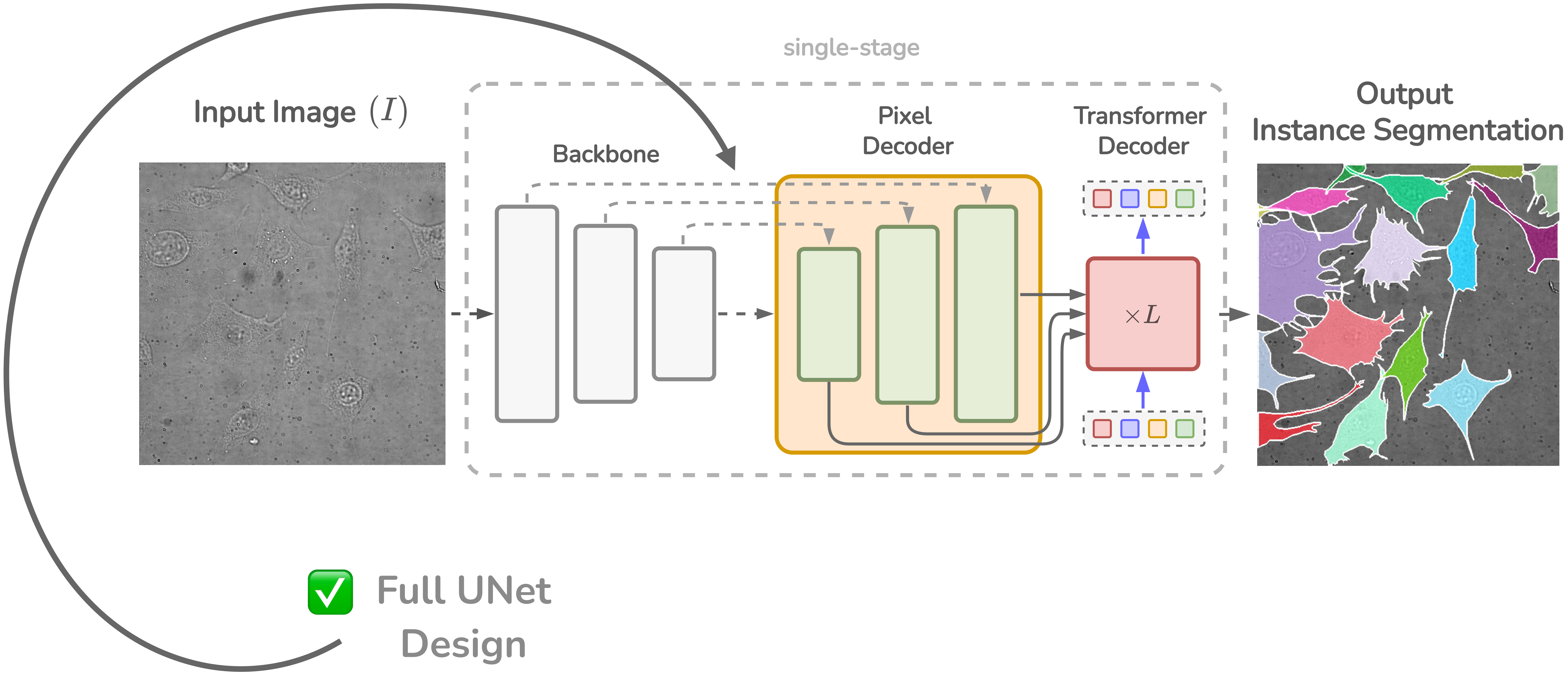
# Instance-Aware **UNet (IAUNet)**

(Query-Based)



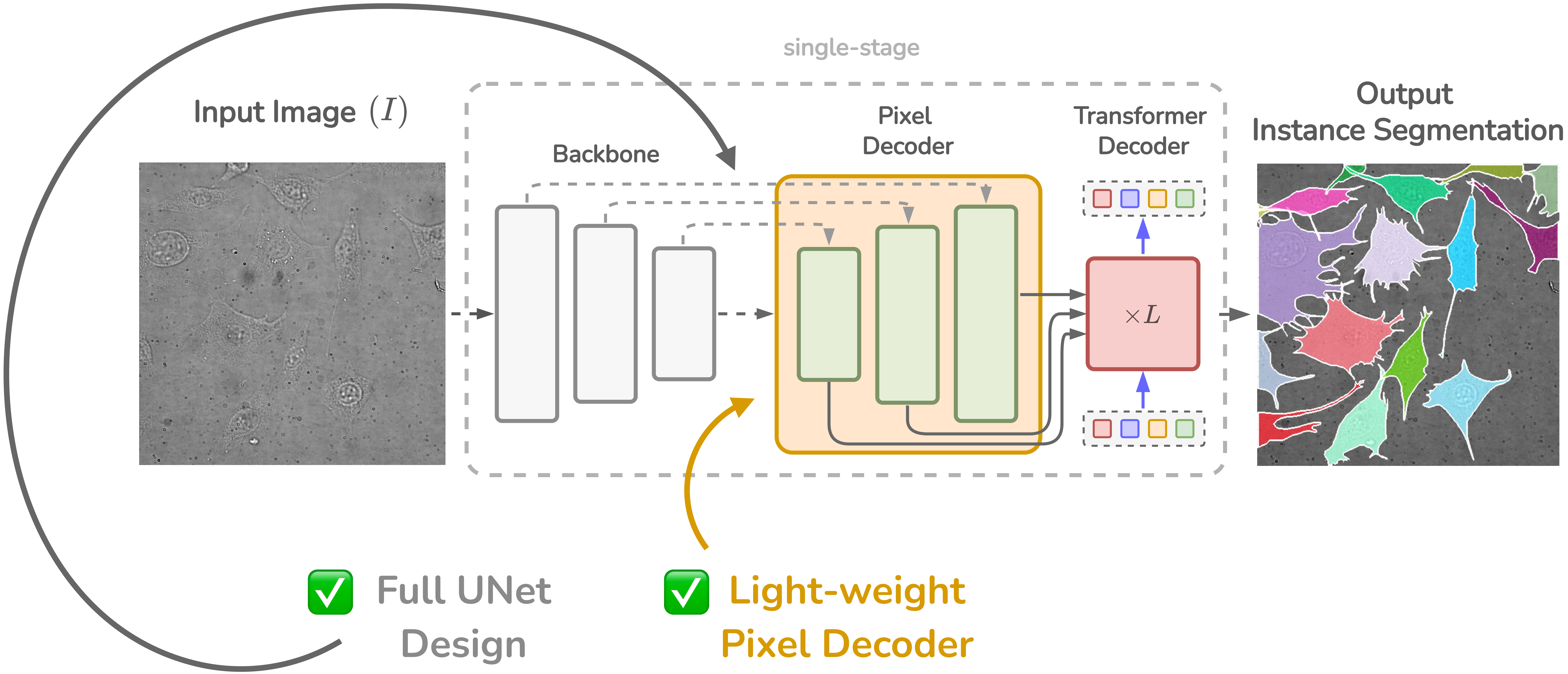
# Instance-Aware **UNet (IAUNet)**

(Query-Based)



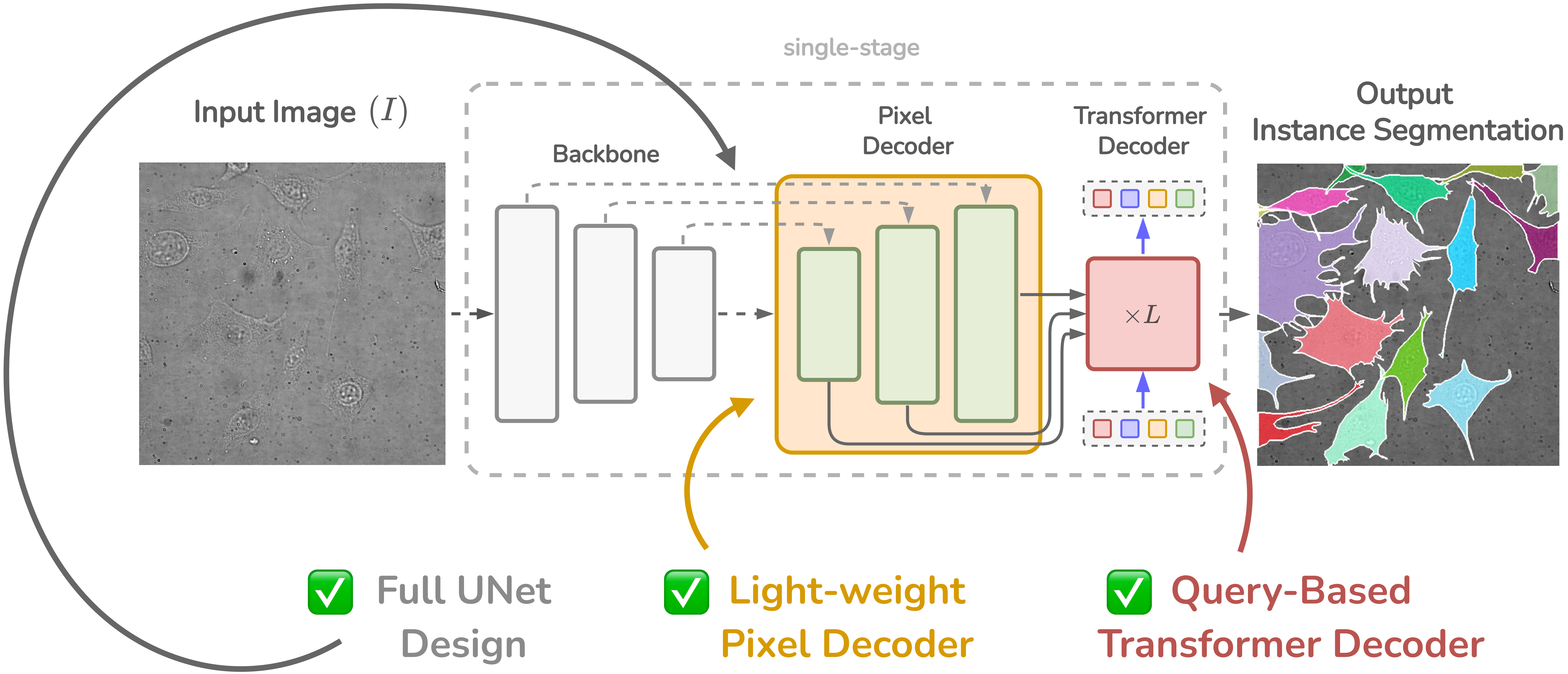
# Instance-Aware **UNet (IAUNet)**

(Query-Based)



# Instance-Aware **UNet (IAUNet)**

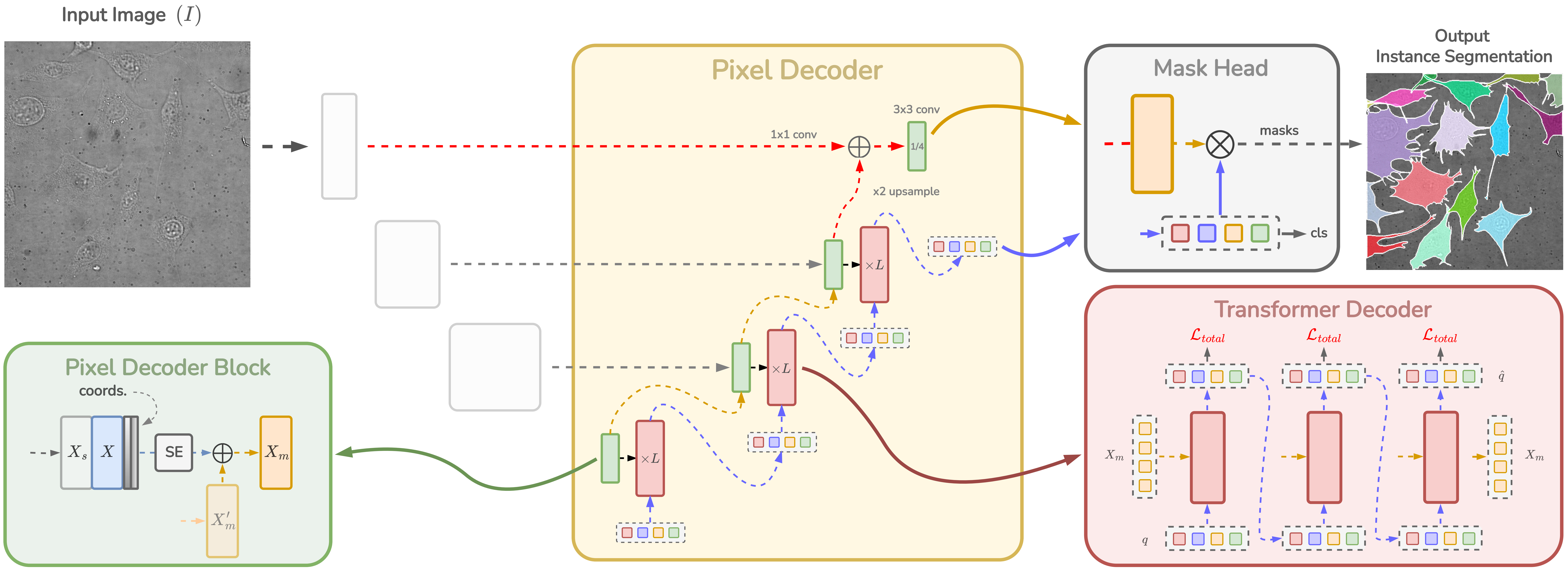
(Query-Based)



# Related Works

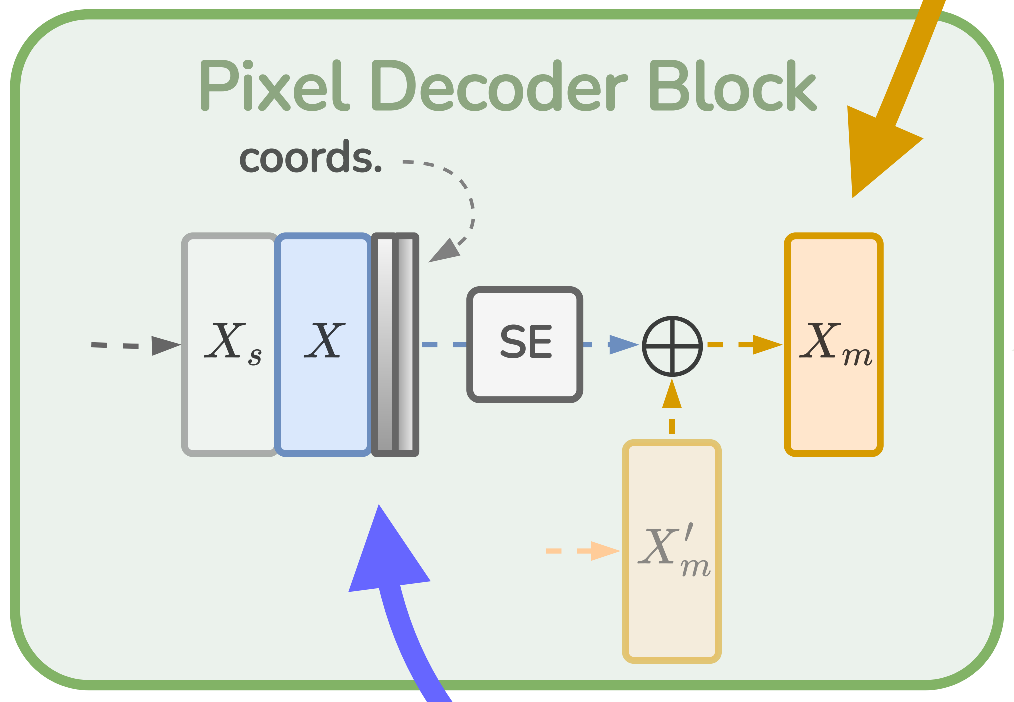
| | U-Net | Mask R-CNN | Mask2Former | MaskDINO | Cellpose | **IAUNet** |
|---|---|---|---|---|---|---|
| Instance segmentation | ❌ | ✅ | ✅ | ✅ | ✅ | ✅ |
| Overlaps | ❌ | ✅ | ✅ | ✅ | ❌ | ✅ |
| Single-stage | ✅ | ❌ | ✅ | ✅ | ❌ | ✅ |
| No NMS | - | ❌ | ✅ | ✅ | ✅ | ✅ |
| U-Net Pixel Decoder | ✅ | ❌ | ❌ | ❌ | ✅ | ✅ |
| Query selection | - | - | ❌ | ✅ | - | ❌ |

# Instance-Aware **UNet (IAUNet)**

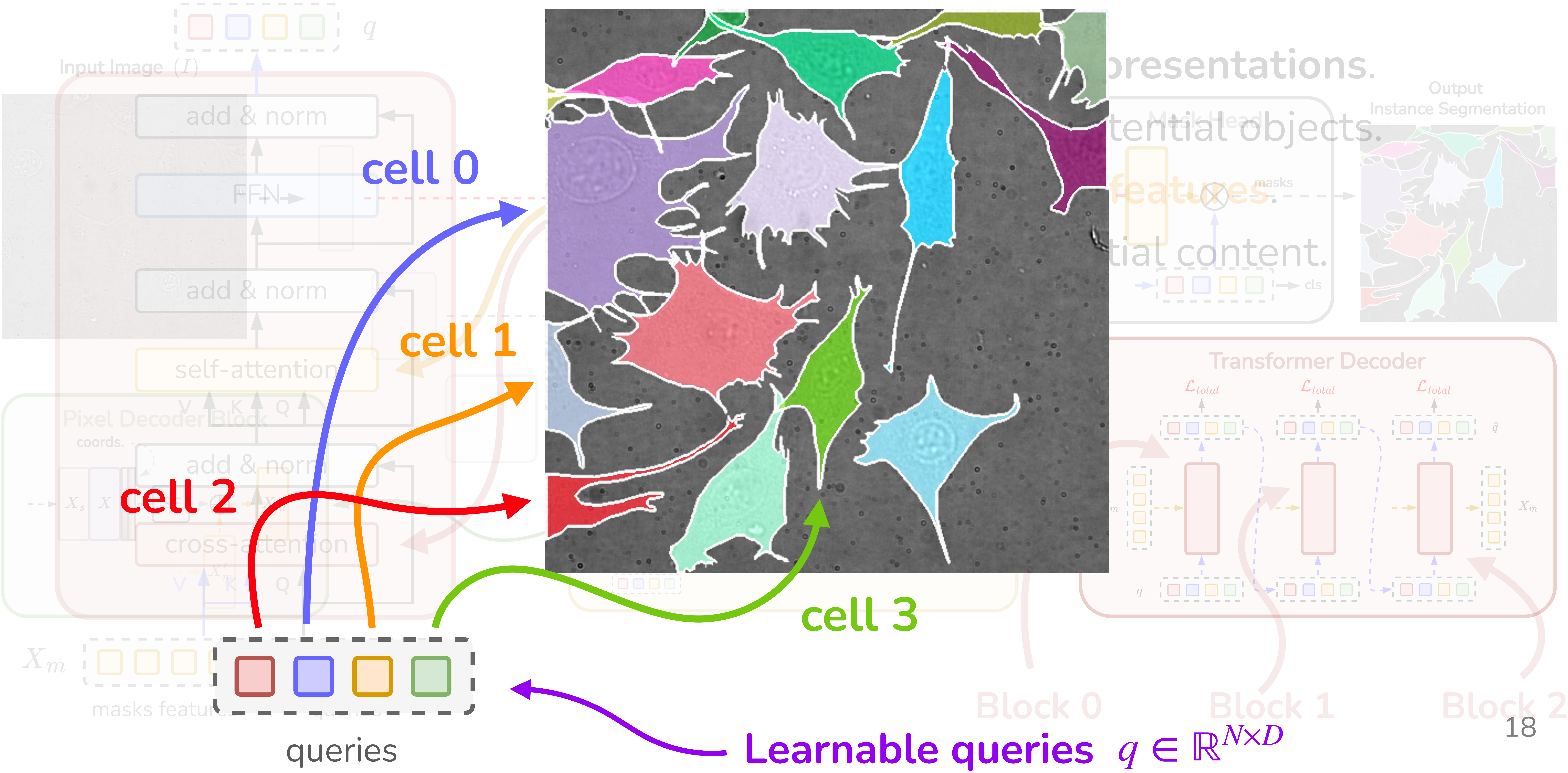


# Pixel Decoder Block

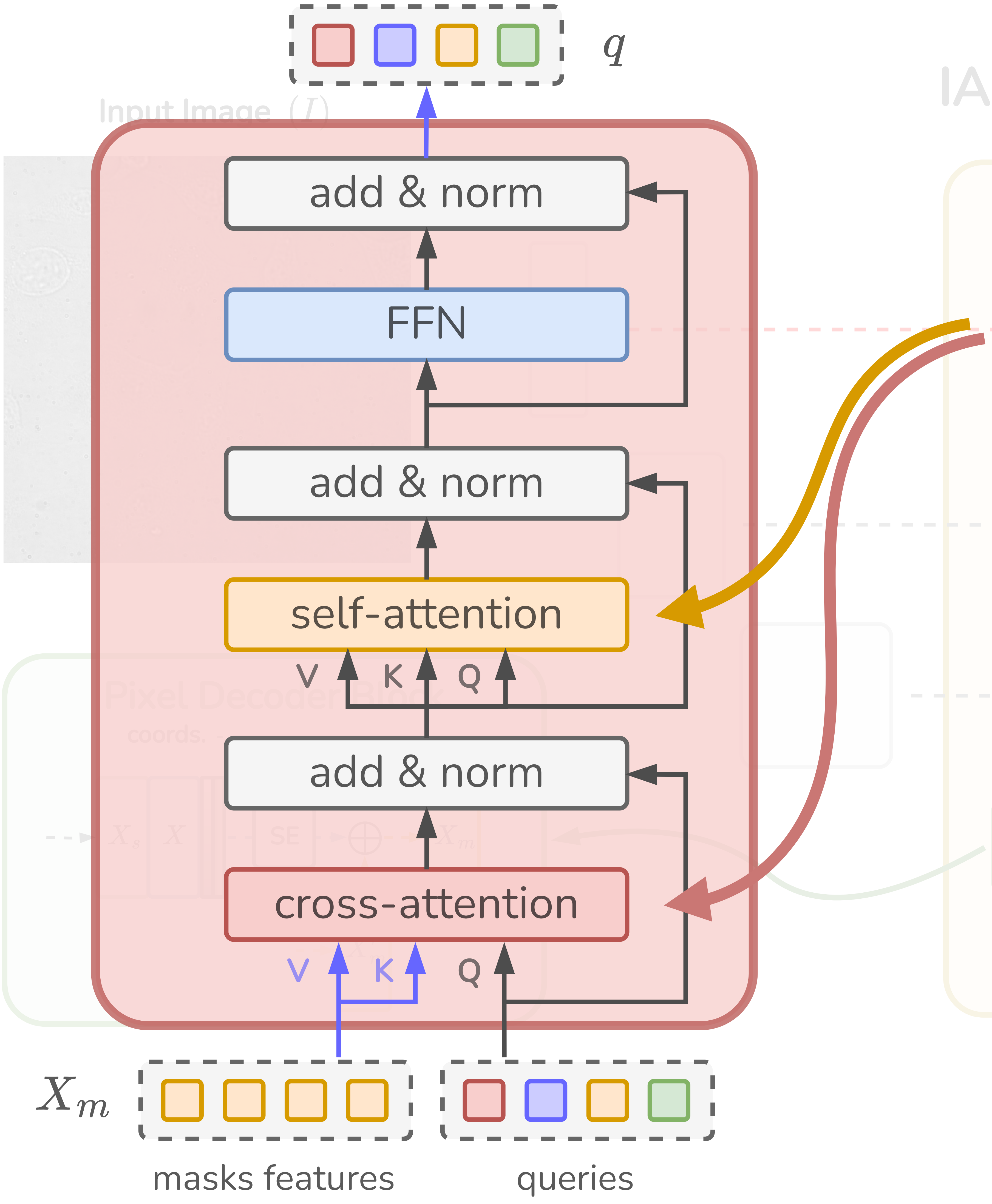
- Refines **multi-scale features**.
- Lightweight **depth-wise convolutions**.
- **CoordConv** injects explicit positional information.
- **Squeeze-and-Excitation** (SE) block enhances feature refinement.



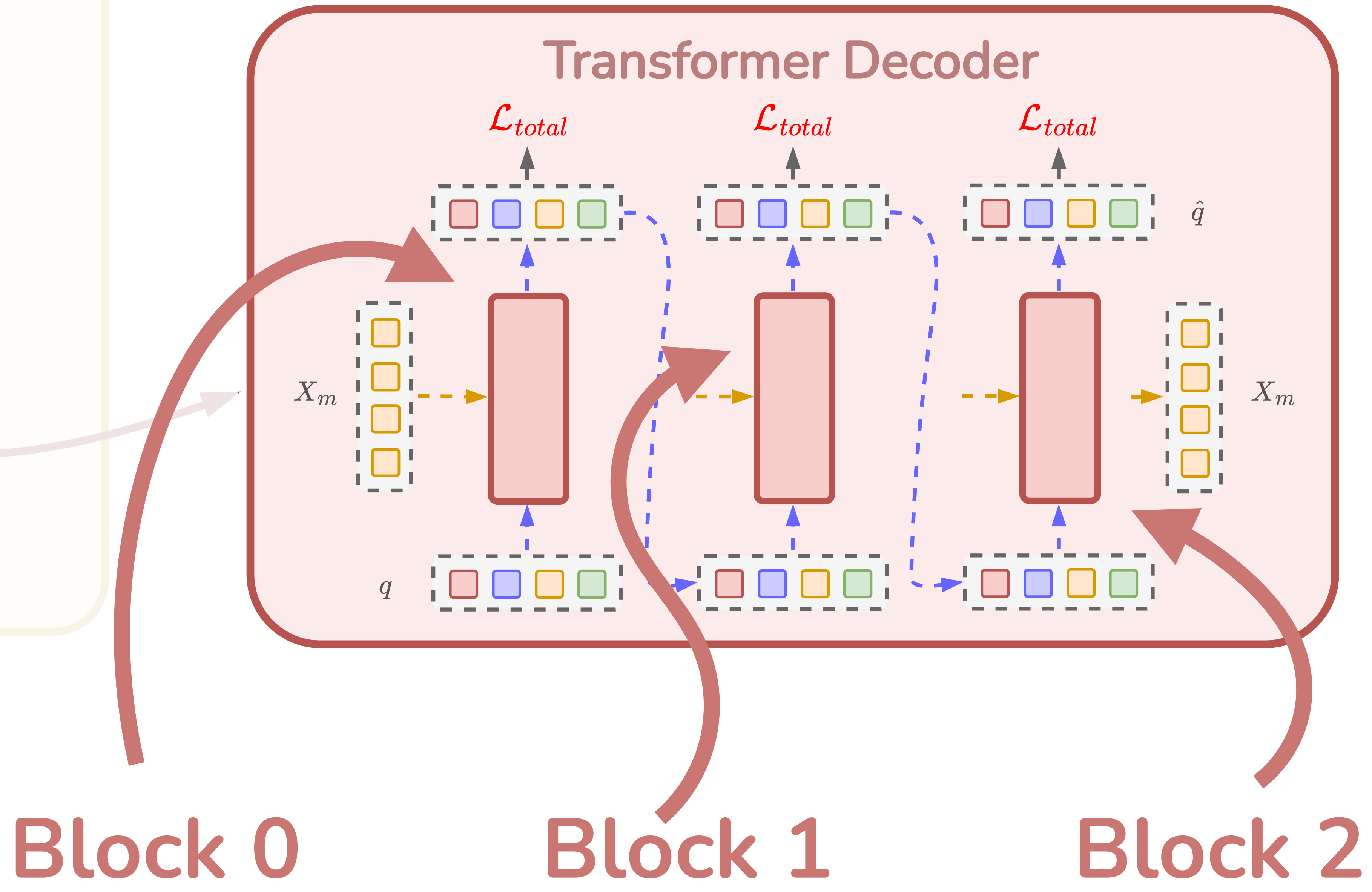
# Each **query** encodes features about a potential **object**



# Transformer Decoder Block

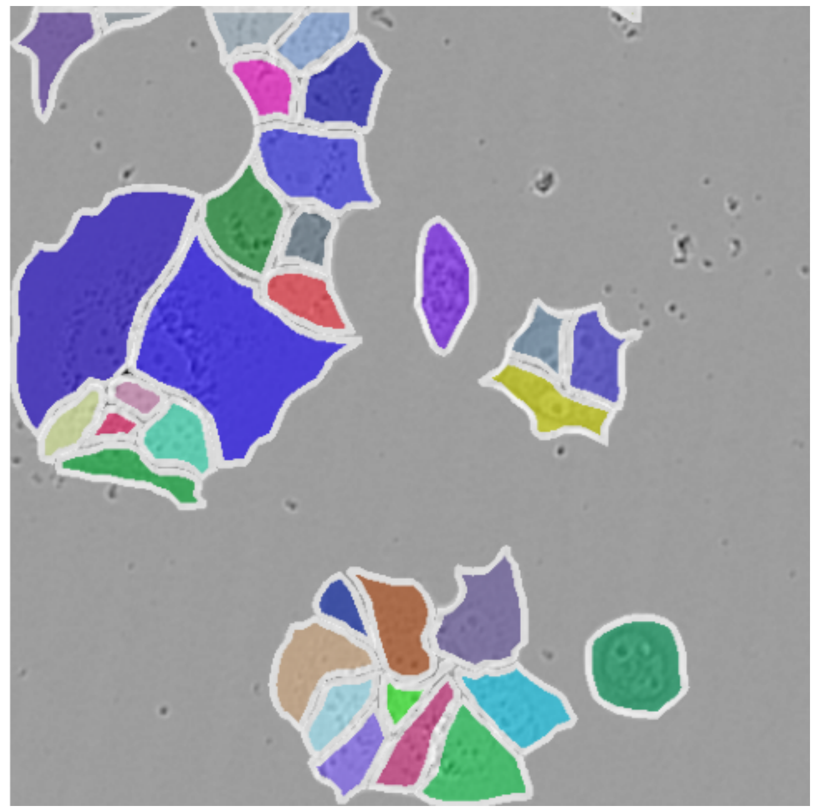



- Learns instance-level **representations**.
- **Learnable queries** as potential objects.
- **Queries attend** to **mask features**.
- Refine semantic and spatial content.

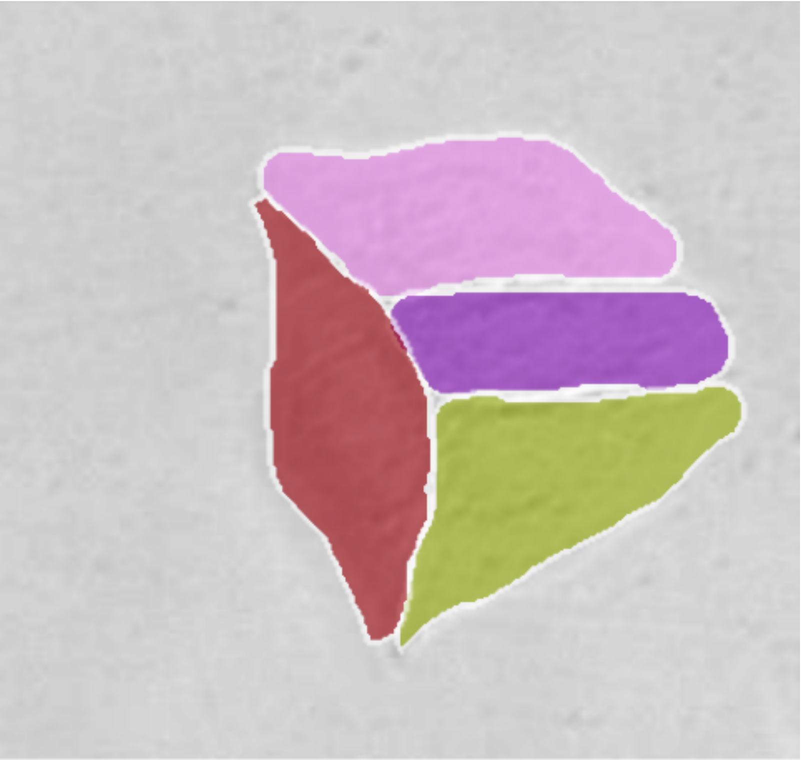


| Models | backbones | num_queries | *LIVECell* AP | *LIVECell* $AP_{50}$ | *EVICAN2$_E$* AP | *EVICAN2$_E$* $AP_{50}$ | *EVICAN2$_M$* AP | *EVICAN2$_M$* $AP_{50}$ | *EVICAN2$_D$* AP | *EVICAN2$_D$* $AP_{50}$ | *ISBI2014* AP | *ISBI2014* $AP_{50}$ | #params. | FLOPs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Models with Convolution-Based Backbones*** | | | | | | | | | | | | | | |
| Mask R-CNN [14] | R50 | 100 | 44.7 | 74.2 | 48.1 | 75.9 | 20.7 | 42.5 | 19.1 | 39.8 | 58.9 | **88.7** | 44M | 115G |
| PointRend [34] | R50 | 100 | 44.0 | 73.5 | 26.6 | 47.9 | 18.0 | 38.5 | 13.4 | 28.3 | **60.0** | **88.7** | 56M | 66G |
| Mask2Former [19] | R50 | 100 | 43.7 | 73.8 | 53.4 | 89.1 | 29.1 | 54.9 | 24.2 | **50.4** | 58.5 | 87.5 | 44M | 67G |
| MaskDINO [20] | R50 | 100 | 43.3 | 73.5 | 50.7 | 83.9 | 29.3 | 57.9 | 22.0 | 41.9 | 55.4 | 86.8 | 44M | 64G |
| **IAUNet (ours)** | R50 | 100 | **45.3** | **75.3** | **58.0** | **91.8** | **32.1** | **59.0** | **24.9** | 45.4 | 56.0 | 85.0 | 39M | 49G |
| Mask R-CNN [14] | R101 | 100 | 44.2 | 73.2 | 41.5 | 69.9 | 23.3 | 46.9 | 17.8 | 36.7 | **60.7** | 88.8 | 63M | 134G |
| PointRend [34] | R101 | 100 | 44.0 | 73.7 | 41.3 | 65.2 | 20.2 | 39.3 | 14.8 | 32.1 | 60.3 | **89.2** | 75M | 86G |
| Mask2Former [19] | R101 | 100 | 44.0 | 73.5 | 54.4 | 87.8 | 27.1 | 51.7 | 20.4 | 42.4 | 59.5 | 88.6 | 63M | 86G |
| MaskDINO [20] | R101 | 100 | 43.4 | 73.6 | 53.7 | 85.0 | 31.8 | 59.2 | **27.1** | **51.3** | 55.7 | 87.4 | 63M | 84G |
| **IAUNet (ours)** | R101 | 100 | **45.4** | **75.5** | **58.3** | **92.7** | **32.9** | **59.6** | 26.9 | 50.0 | 56.5 | 87.1 | 58M | 69G |
| ***Models with Transformer-Based Backbones*** | | | | | | | | | | | | | | |
| Mask R-CNN [14] | Swin-S | 100 | 44.3 | 73.3 | 52.6 | 91.7 | 27.0 | 59.2 | 20.2 | 50.2 | 61.9 | 90.7 | 69M | 141G |
| PointRend [34] | Swin-S | 100 | 43.9 | 73.5 | 55.1 | 89.2 | 30.1 | 61.6 | 24.4 | 54.6 | **62.1** | **91.0** | 81M | 93G |
| Mask2Former [19] | Swin-S | 100 | 44.6 | 74.3 | **65.2** | **96.8** | **36.2** | 66.7 | **30.9** | 62.7 | 57.1 | 87.3 | 69M | 93G |
| MaskDINO [20] | Swin-S | 100 | 43.9 | 73.8 | 57.0 | 86.9 | 33.6 | 64.9 | 27.6 | 56.9 | 52.7 | 85.3 | 71M | 181G |
| MaskDINO [20] | Swin-S | 300 | 44.8 | 75.1 | 56.5 | 91.8 | 35.0 | **70.7** | 30.2 | **64.3** | 51.2 | 83.4 | 71M | 187G |
| **IAUNet (ours)** | Swin-S | 100 | 45.4 | 75.4 | 58.8 | 93.1 | 32.2 | 61.9 | 27.7 | 54.1 | 61.1 | 90.1 | 64M | 76G |
| **IAUNet (ours)** | Swin-S | 300 | **45.6** | **76.4** | 60.9 | 93.6 | 33.2 | 62.0 | 29.6 | 58.0 | 61.8 | 89.8 | 64M | 87G |
| Mask R-CNN [14] | Swin-B | 100 | 44.2 | 73.1 | 52.0 | 89.0 | 26.7 | 60.3 | 24.8 | 55.5 | 62.4 | **91.5** | 107M | 186G |
| PointRend [34] | Swin-B | 100 | 44.0 | 73.7 | 58.6 | 91.0 | 34.1 | 64.6 | 25.8 | 52.0 | 62.7 | **91.5** | 119M | 137G |
| Mask2Former [19] | Swin-B | 100 | 44.9 | 74.7 | 55.0 | 92.5 | 31.4 | 60.9 | 27.7 | 56.6 | 58.1 | 88.4 | 107M | 138G |
| MaskDINO [20] | Swin-B | 100 | 44.3 | 74.1 | 57.3 | 91.1 | 37.3 | 75.7 | 30.1 | 65.6 | 53.5 | 86.6 | 110M | 226G |
| MaskDINO [20] | Swin-B | 300 | 45.2 | 75.8 | 57.9 | 91.6 | **39.1** | **78.8** | **34.0** | **72.3** | 53.3 | 84.8 | 110M | 232G |
| **IAUNet (ours)** | Swin-B | 100 | 45.5 | 75.6 | 59.6 | 93.5 | 34.2 | 65.7 | 28.9 | 56.9 | 61.5 | 90.8 | 102M | 120G |
| **IAUNet (ours)** | Swin-B | 300 | **45.8** | **76.7** | **61.2** | **94.8** | 38.0 | 69.6 | 30.7 | 59.9 | **63.0** | **91.5** | 102M | 132G |

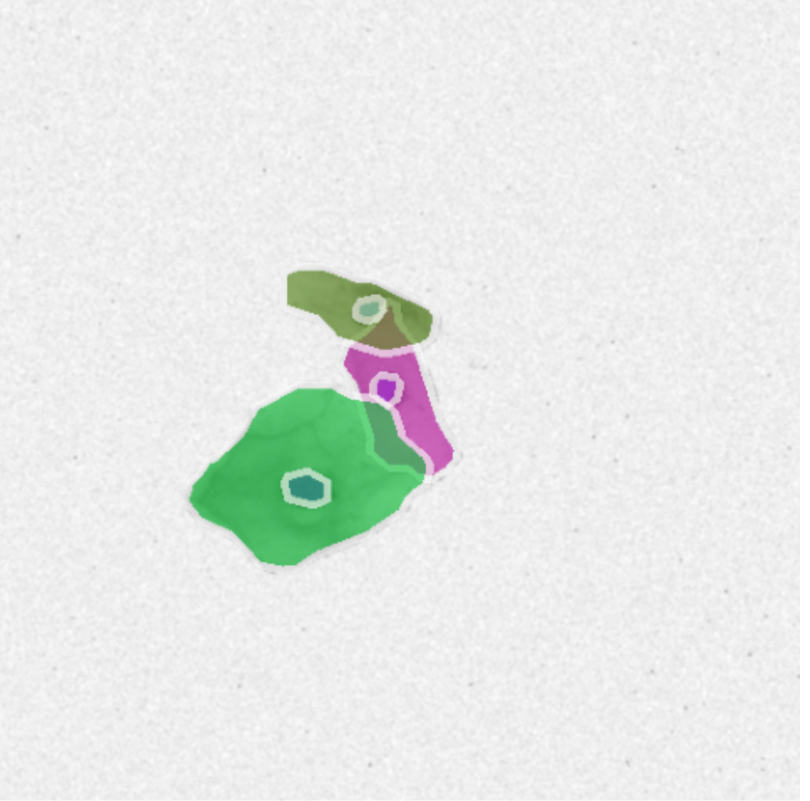
**Table 1.** Instance segmentation on LIVECell, EVICAN2 (Easy, Medium, Difficult test subsets), and ISBI2014.



LIVECell



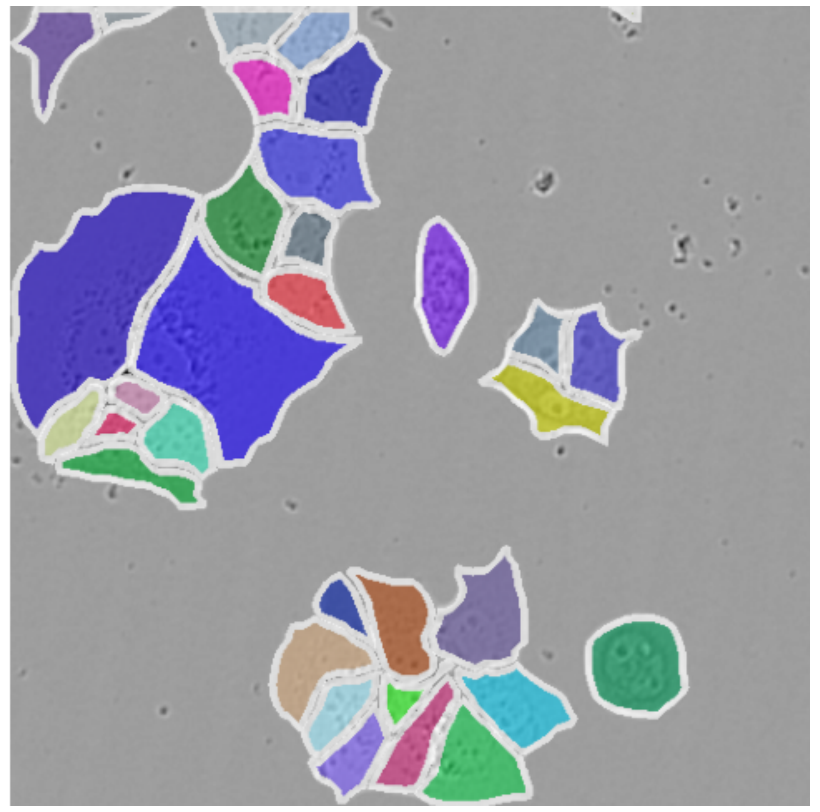
EVICAN



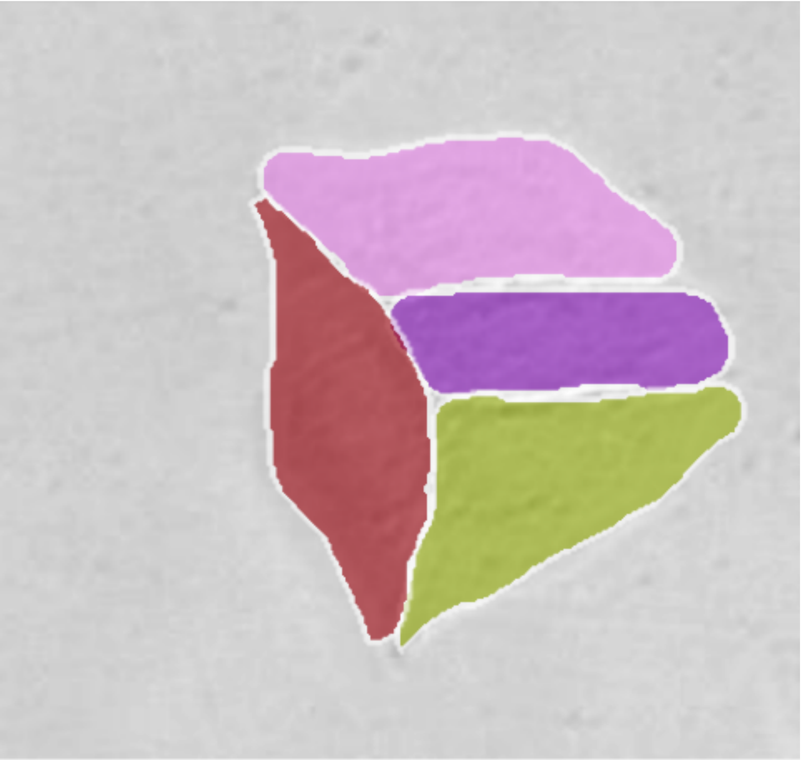
ISBI2014

| Models | backbones | num_queries | *LIVECell* AP | *LIVECell* $AP_{50}$ | *EVICAN2$_E$* AP | *EVICAN2$_E$* $AP_{50}$ | *EVICAN2$_M$* AP | *EVICAN2$_M$* $AP_{50}$ | *EVICAN2$_D$* AP | *EVICAN2$_D$* $AP_{50}$ | *ISBI2014* AP | *ISBI2014* $AP_{50}$ | #params. | FLOPs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Models with Convolution-Based Backbones*** | | | | | | | | | | | | | | |
| Mask R-CNN [14] | R50 | 100 | 44.7 | 74.2 | 48.1 | 75.9 | 20.7 | 42.5 | 19.1 | 39.8 | 58.9 | **88.7** | 44M | 115G |
| PointRend [34] | R50 | 100 | 44.0 | 73.5 | 26.6 | 47.9 | 18.0 | 38.5 | 13.4 | 28.3 | **60.0** | **88.7** | 56M | 66G |
| Mask2Former [19] | R50 | 100 | 43.7 | 73.8 | 53.4 | 89.1 | 29.1 | 54.9 | 24.2 | **50.4** | 58.5 | 87.5 | 44M | 67G |
| MaskDINO [20] | R50 | 100 | 43.3 | 73.5 | 50.7 | 83.9 | 29.3 | 57.9 | 22.0 | 41.9 | 55.4 | 86.8 | 44M | 64G |
| **IAUNet (ours)** | R50 | 100 | **45.3** | **75.3** | **58.0** | **91.8** | **32.1** | **59.0** | **24.9** | 45.4 | 56.0 | 85.0 | 39M | 49G |
| Mask R-CNN [14] | R101 | 100 | 44.2 | 73.2 | 41.5 | 69.9 | 23.3 | 46.9 | 17.8 | 36.7 | **60.7** | 88.8 | 63M | 134G |
| PointRend [34] | R101 | 100 | 44.0 | 73.7 | 41.3 | 65.2 | 20.2 | 39.3 | 14.8 | 32.1 | 60.3 | **89.2** | 75M | 86G |
| Mask2Former [19] | R101 | 100 | 44.0 | 73.5 | 54.4 | 87.8 | 27.1 | 51.7 | 20.4 | 42.4 | 59.5 | 88.6 | 63M | 86G |
| MaskDINO [20] | R101 | 100 | 43.4 | 73.6 | 53.7 | 85.0 | 31.8 | 59.2 | **27.1** | **51.3** | 55.7 | 87.4 | 63M | 84G |
| **IAUNet (ours)** | R101 | 100 | **45.4** | **75.5** | **58.3** | **92.7** | **32.9** | **59.6** | 26.9 | 50.0 | 56.5 | 87.1 | 58M | 69G |
| ***Models with Transformer-Based Backbones*** | | | | | | | | | | | | | | |
| Mask R-CNN [14] | Swin-S | 100 | 44.3 | 73.3 | 52.6 | 91.7 | 27.0 | 59.2 | 20.2 | 50.2 | 61.9 | 90.7 | 69M | 141G |
| PointRend [34] | Swin-S | 100 | 43.9 | 73.5 | 55.1 | 89.2 | 30.1 | 61.6 | 24.4 | 54.6 | **62.1** | **91.0** | 81M | 93G |
| Mask2Former [19] | Swin-S | 100 | 44.6 | 74.3 | **65.2** | **96.8** | **36.2** | 66.7 | **30.9** | 62.7 | 57.1 | 87.3 | 69M | 93G |
| MaskDINO [20] | Swin-S | 100 | 43.9 | 73.8 | 57.0 | 86.9 | 33.6 | 64.9 | 27.6 | 56.9 | 52.7 | 85.3 | 71M | 181G |
| MaskDINO [20] | Swin-S | 300 | 44.8 | 75.1 | 56.5 | 91.8 | 35.0 | **70.7** | 30.2 | **64.3** | 51.2 | 83.4 | 71M | 187G |
| **IAUNet (ours)** | Swin-S | 100 | 45.4 | 75.4 | 58.8 | 93.1 | 32.2 | 61.9 | 27.7 | 54.1 | 61.1 | 90.1 | 64M | 76G |
| **IAUNet (ours)** | Swin-S | 300 | **45.6** | **76.4** | 60.9 | 93.6 | 33.2 | 62.0 | 29.6 | 58.0 | 61.8 | 89.8 | 64M | 87G |
| Mask R-CNN [14] | Swin-B | 100 | 44.2 | 73.1 | 52.0 | 89.0 | 26.7 | 60.3 | 24.8 | 55.5 | 62.4 | **91.5** | 107M | 186G |
| PointRend [34] | Swin-B | 100 | 44.0 | 73.7 | 58.6 | 91.0 | 34.1 | 64.6 | 25.8 | 52.0 | 62.7 | **91.5** | 119M | 137G |
| Mask2Former [19] | Swin-B | 100 | 44.9 | 74.7 | 55.0 | 92.5 | 31.4 | 60.9 | 27.7 | 56.6 | 58.1 | 88.4 | 107M | 138G |
| MaskDINO [20] | Swin-B | 100 | 44.3 | 74.1 | 57.3 | 91.1 | 37.3 | 75.7 | 30.1 | 65.6 | 53.5 | 86.6 | 110M | 226G |
| MaskDINO [20] | Swin-B | 300 | 45.2 | 75.8 | 57.9 | 91.6 | **39.1** | **78.8** | **34.0** | **72.3** | 53.3 | 84.8 | 110M | 232G |
| **IAUNet (ours)** | Swin-B | 100 | 45.5 | 75.6 | 59.6 | 93.5 | 34.2 | 65.7 | 28.9 | 56.9 | 61.5 | 90.8 | 102M | 120G |
| **IAUNet (ours)** | Swin-B | 300 | **45.8** | **76.7** | **61.2** | **94.8** | 38.0 | 69.6 | 30.7 | 59.9 | **63.0** | **91.5** | 102M | 132G |

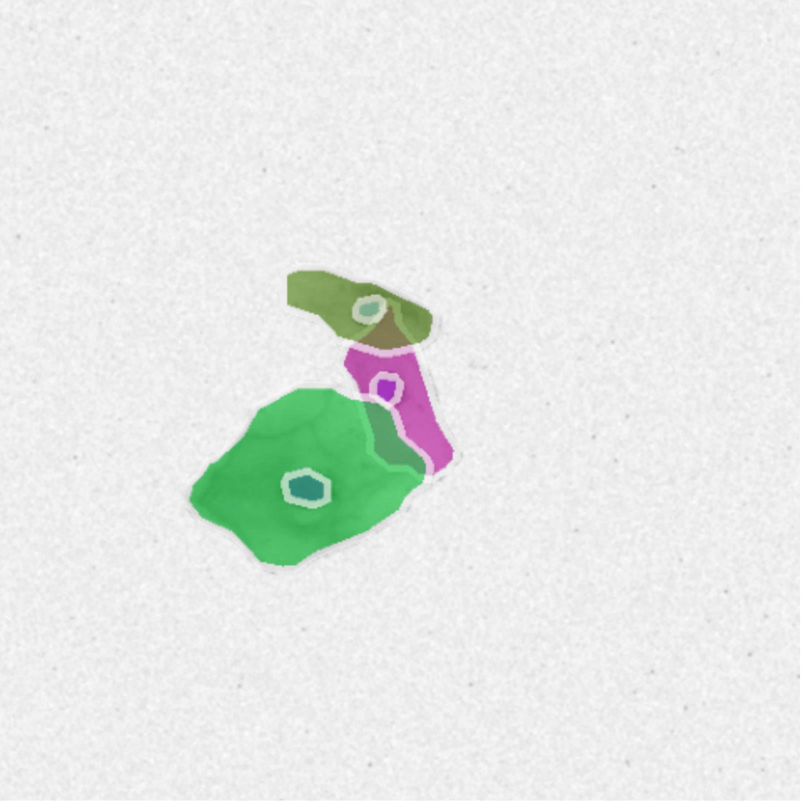
**Table 1.** Instance segmentation on LIVECell, EVICAN2 (Easy, Medium, Difficult test subsets), and ISBI2014.



LIVECell



EVICAN



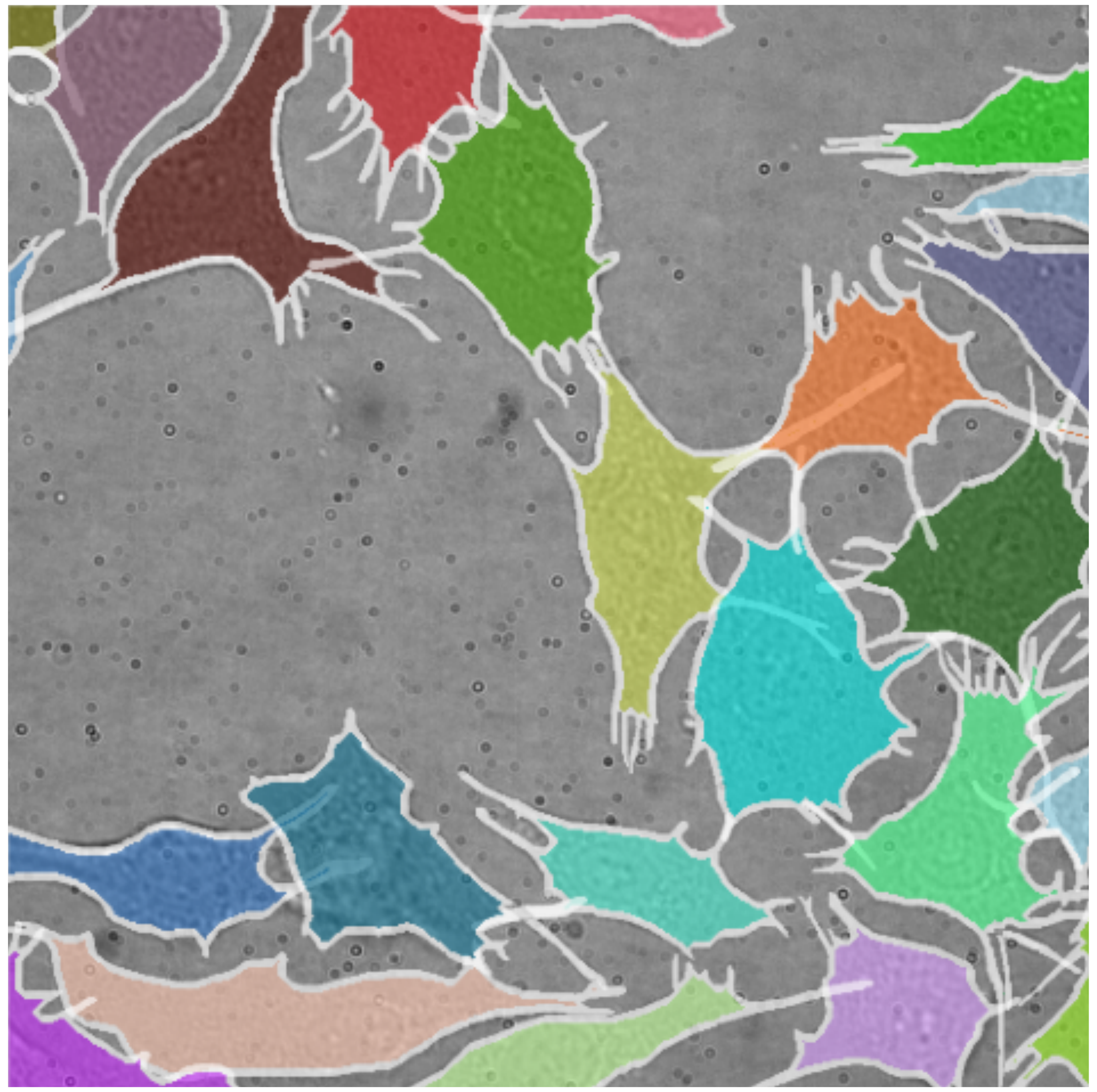
ISBI2014

*Revvity-25*

| Models | backbones | num_queries | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ | #params. | FLOPs |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Models with Convolution-Based Backbones*** | | | | | | | | | | |
| Mask R-CNN [14] | R50 | 100 | 39.7 | 77.2 | 37.4 | 0.6 | 19.0 | 44.6 | 44M | 115G |
| PointRend [34] | R50 | 100 | 42.2 | 79.4 | 40.9 | 0.4 | 21.7 | 47.3 | 56M | 66G |
| Mask2Former [19] | R50 | 100 | 46.4 | 79.8 | 49.9 | 0.7 | 25.7 | 52.8 | 44M | 67G |
| MaskDINO [20] | R50 | 100 | 45.6 | 80.4 | 48.2 | **1.8** | 22.3 | 51.8 | 44M | 64G |
| **IAUNet (ours)** | R50 | 100 | **49.7** | **82.1** | **54.8** | 0.6 | **27.3** | **56.0** | 39M | 49G |
| Mask R-CNN [14] | R101 | 100 | 40.7 | 77.5 | 39.9 | 0.4 | 20.1 | 45.8 | 63M | 134G |
| PointRend [34] | R101 | 100 | 42.9 | 79.3 | 42.5 | 0.0 | 18.4 | 48.9 | 75M | 86G |
| Mask2Former [19] | R101 | 100 | 47.2 | 80.1 | 51.8 | **1.7** | 25.7 | 53.3 | 63M | 86G |
| MaskDINO [20] | R101 | 100 | 47.3 | 81.0 | 50.4 | 0.9 | 23.0 | 53.5 | 63M | 84G |
| **IAUNet (ours)** | R101 | 100 | **51.5** | **84.7** | **56.1** | **1.7** | **29.2** | **57.8** | 58M | 69G |
| ***Models with Transformer-Based Backbones*** | | | | | | | | | | |
| Mask R-CNN [14] | Swin-S | 100 | 24.7 | 63.4 | 12.5 | 0.0 | 7.3 | 28.9 | 69M | 141G |
| PointRend [34] | Swin-S | 100 | 43.6 | 80.0 | 43.0 | 0.5 | 21.5 | 48.9 | 81M | 93G |
| Mask2Former [19] | Swin-S | 100 | 51.2 | 83.3 | 56.4 | 2.7 | 27.7 | 58.0 | 69M | 93G |
| MaskDINO [20] | Swin-S | 100 | 50.3 | 83.2 | 53.9 | **4.7** | 27.6 | 56.1 | 71M | 181G |
| MaskDINO [20] | Swin-S | 300 | 49.4 | 83.6 | 53.3 | 2.9 | 25.8 | 55.3 | 71M | 187G |
| **IAUNet (ours)** | Swin-S | 100 | 53.0 | 85.7 | 57.0 | 1.3 | **29.7** | 59.1 | 64M | 76G |
| **IAUNet (ours)** | Swin-S | 300 | **53.3** | **86.0** | **59.6** | 1.6 | 29.4 | **59.8** | 64M | 87G |
| Mask R-CNN [14] | Swin-B | 100 | 27.1 | 64.9 | 17.2 | 0.1 | 9.7 | 31.2 | 107M | 186G |
| PointRend [34] | Swin-B | 100 | 45.2 | 80.1 | 47.9 | 0.1 | 23.0 | 50.9 | 119M | 137G |
| Mask2Former [19] | Swin-B | 100 | 52.0 | 83.6 | 58.4 | 1.1 | 27.8 | 59.0 | 107M | 138G |
| MaskDINO [20] | Swin-B | 100 | 50.5 | 83.5 | 54.9 | **2.0** | 27.1 | 56.4 | 110M | 226G |
| MaskDINO [20] | Swin-B | 300 | 50.4 | 84.3 | 54.8 | 0.8 | 26.3 | 56.6 | 110M | 232G |
| **IAUNet (ours)** | Swin-B | 100 | 53.5 | 86.1 | **59.4** | 0.8 | **30.5** | 59.7 | 102M | 120G |
| **IAUNet (ours)** | Swin-B | 300 | **53.7** | **86.5** | **59.4** | 1.0 | 30.0 | **60.3** | 102M | 132G |

**Table 2. Instance segmentation on our Revvity-25 dataset.** IAUNet outperforms strong query-based baselines as well as other state-of-the-art models when training with fewer parameters

## Revvity-25



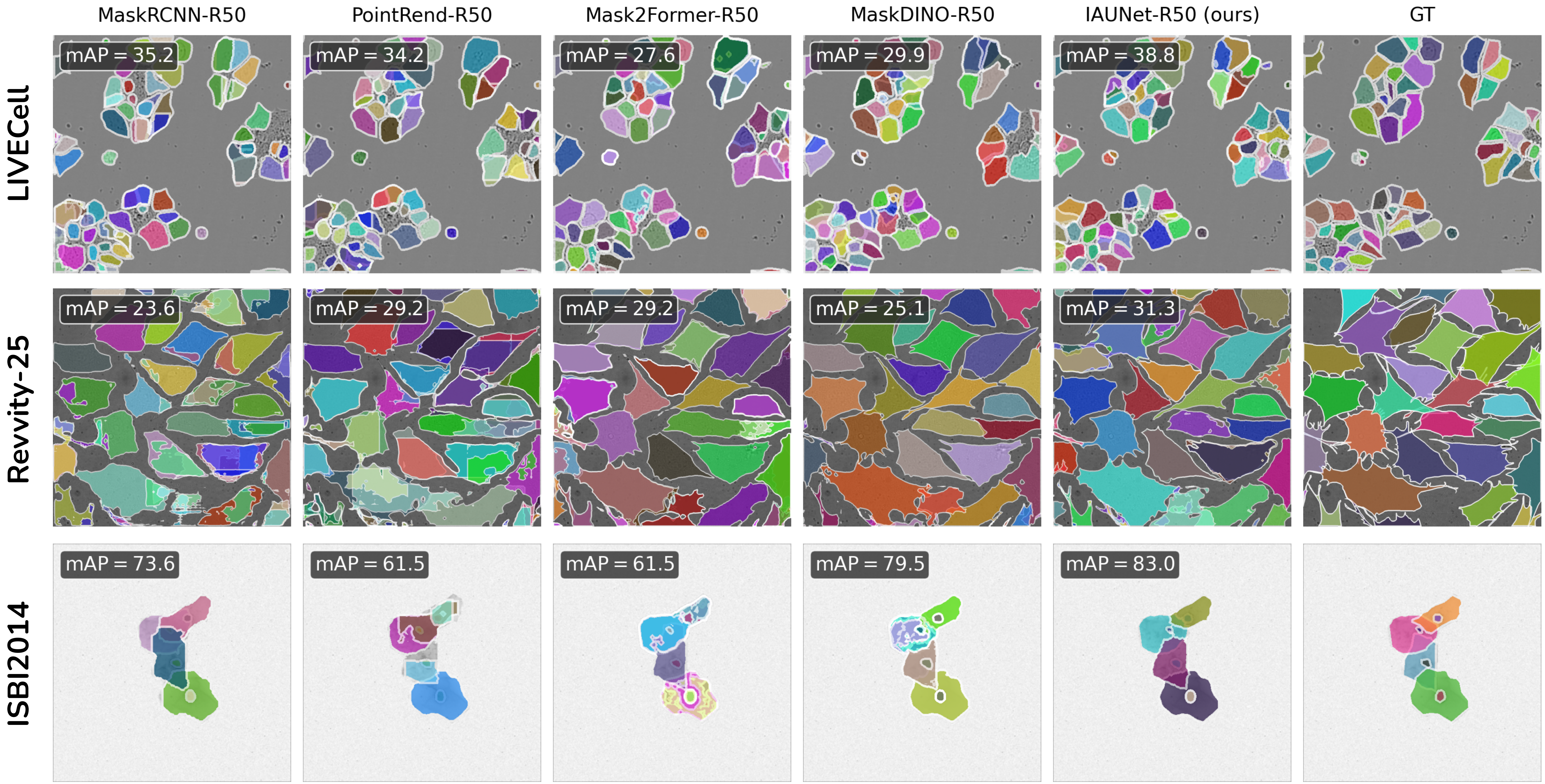



**Figure 2.** Visualization of instance segmentation predictions across different state-of-the-art models (using ResNet50 backbone). We also report per-image AP score.

# Conclusions 🇺🇦

- We introduce **IAUNet**, a novel model for cell instance segmentation that integrates a **lightweight convolutional** Pixel decoder and a Transformer decoder for efficient multi-scale object query refinement.

# Conclusions 🇺🇦

- We introduce **IAUNet**, a novel model for cell instance segmentation that integrates a **lightweight convolutional** Pixel decoder and a Transformer decoder for efficient multi-scale object query refinement.
- We present the **2025 Revvity Full Cell Segmentation Dataset**, featuring detailed and validated annotations for evaluating segmentation models on brightfield images.

# Thank you! ❤️

UNIVERSITY OF TARTU
Institute of Computer Science
1632

revvity

BCV Lab


Armed Forces
of Ukraine

Dzyga's
Paw

UKRAINIAN CATHOLIC UNIVERSITY

APPLIED
SCIENCES
FACULTY

# Appendix

# Ablations

We investigate the **benefit** of adding different **decoder components**.
Adding **CoordConv** improves object localization.

| Decoder | AP | $AP_{50}$ | $AP_{75}$ | #params. | FLOPs |
|---|---|---|---|---|---|
| **IAUNet (R50)** | 43.8 | 73.1 | 47.4 | 34M | 42G |
| + mask branch $X_m$ | 44.0 | 73.2 | 47.9 | 34M | 42G |
| + FFN (2048 $\rightarrow$ 1024) | 44.1 | 73.2 | 48.0 | 32M | 42G |
| + SE block [68] | 44.2 | 73.3 | 48.1 | 32M | 42G |
| + CoordConv [67] | 44.7 | 74.1 | <u>48.7</u> | 32M | 42G |
| + $L$ (1 $\rightarrow$ 3) (round-robin.) | 44.3 | 74.0 | 48.1 | 39M | 49G |
| + $L$ (1 $\rightarrow$ 3) (seq.) | <u>45.1</u> | <u>74.4</u> | **49.4** | 39M | 49G |
| + deep_supervision | **45.3** | **75.3** | **49.4** | 39M | 49G |

**Figure <>.** Visualization of instance segmentation predictions across different state-of-the-art models (using ResNet50 backbone). We also report per-image AP score.

# Ablations

We observe consistent gains when increasing the query count from 100 to 300 and 500

| num_queries | AP | $AP_{50}$ | $AP_{75}$ | FLOPs |
|---|---|---|---|---|
| 100 | 45.3 | 75.3 | 49.4 | 49G |
| 300 | 45.9 | 76.5 | 50.4 | 61G |
| 500 | **46.1** | **76.8** | **50.8** | 73G |
| 1000 | 45.3 | 76.3 | 50.0 | 104G |

**Figure <>.** Scaling the number of object queries benefits the model

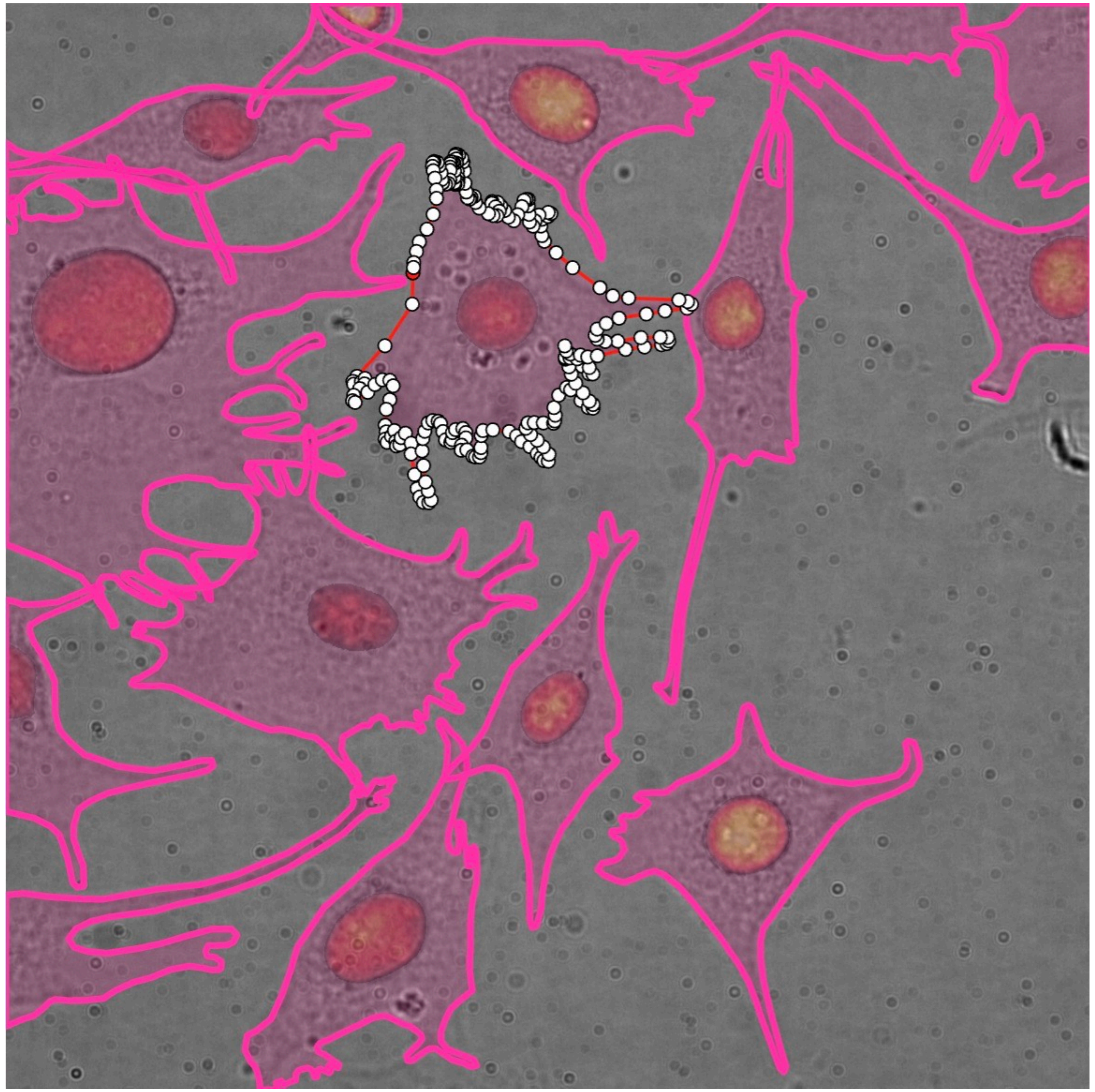
# Ablations

We observe consistent gains when increasing the query count from 100 to 300 and 500

| Pixel Decoder | AP | $AP_{50}$ | $AP_{75}$ | FLOPs |
|---|---|---|---|---|
| + full skip | **44.7** | **73.9** | **48.9** | 146G |
| + $1 \times 1$ skip concat | 44.2 | $\underline{73.8}$ | $\underline{48.3}$ | 135G |
| + $1 \times 1$ skip add | $\underline{44.3}$ | 73.3 | 48.2 | 132G |
| + light mask head | 43.8 | 73.1 | 47.4 | 42G |

**Figure <>.** Scaling the number of object queries benefits the model

# Revvity-25



- **High-resolution** (1080 × 1080)
- **110** brightfield images
- **2,937** expert-validated cell instances
- average of **60** points per cell and up to **400** points for cells with complex morphology
- On average **27** manually labeled
- 7 cell lines

  mouse fibroblasts (**NIH/3T3**)
  canine kidney epithelial cells (**MDCK**)
  human cervical adenocarcinoma (**HeLa**)
  human breast adenocarcinoma (**MCF7**)
  human lung carcinoma (**A549**)
  human hepatocellular carcinoma (**HepG2**)
  human fibrosarcoma (**HT1080**)