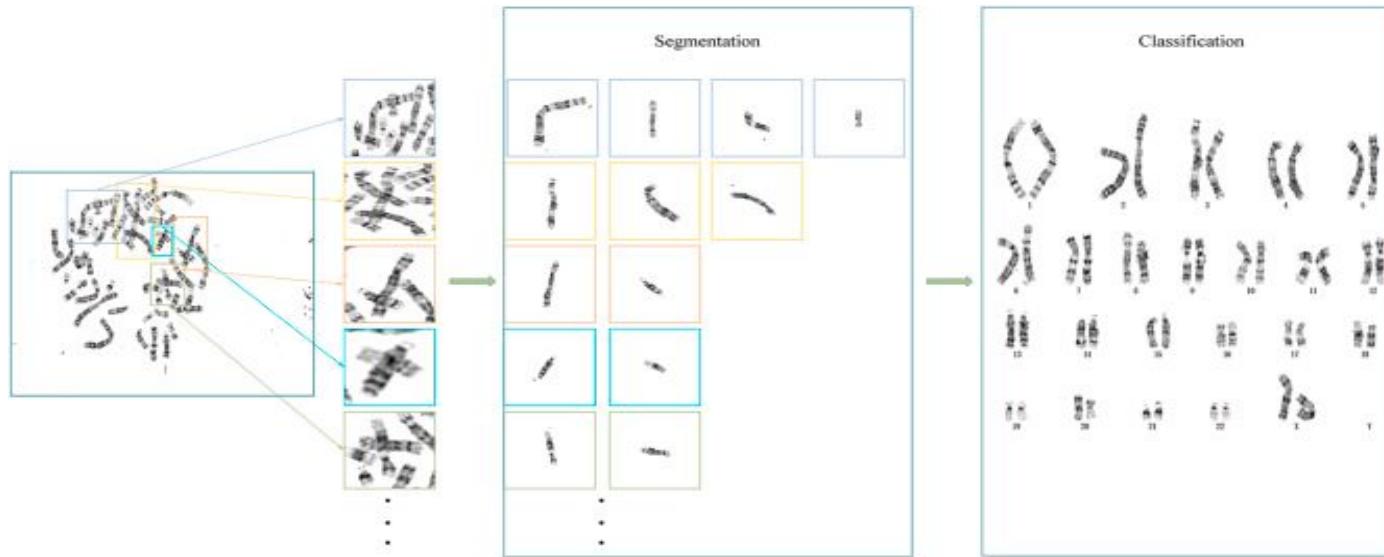


Metaphase images instance segmentation task.



Expected output per image: ~46 individual chromosomes (23 pairs)

Classification: 24 classes representing chromosomes 1-22, X, and Y

Datasets

Dataset 1 — 24-class bounding boxes

- 5,000 metaphase images with 24-class annotations (YOLO-format bounding boxes)
- 229,852 total chromosome instances
- Median image size: ~640 px (median)

Dataset 2 — 24-class instance masks

- 612 metaphase images with instance masks + class labels for all 24 chromosome types
- ~27,000 chromosome instances total
- Image size: 1600×1600
- Split:
 - Train: 547 images
 - Test: 65 images
- Patient-level separation: images from the same patient never appear in both train and test

Dataset 3 — unlabeled metaphase images

- 2,000 metaphase images
- Includes a subset of difficult images (listed separately)

Training setup

- GT:
 - Dataset 2: masks + boxes
 - Dataset 1: boxes
- Train is further split into train/validation subsets 90%/10%
- 40 epochs
- Input image size: 640 px (640×640 crops)

Evaluation

- Instance Seg Test (65 images with instance masks and classification)
- 2000 images unlabeled using the sanity metric

Sanity metric (unlabeled evaluation)

Evaluates biological plausibility of unlabeled predictions for 24 classes (1–22 autosomes, 23=X, 24=Y). Produces a 0–1 score as the mean of:

- **Count:** should be close to 46
- **Class balance:** autosomes (1–22) exactly 2 copies each
- **Sex:** best match to XX (2 X, 0 Y) or XY (1 X, 1 Y)
- **Size:** fraction of boxes within a tolerance factor (default 3×) of the median area (chromosomes are sane size)

Long vs short chromosomes

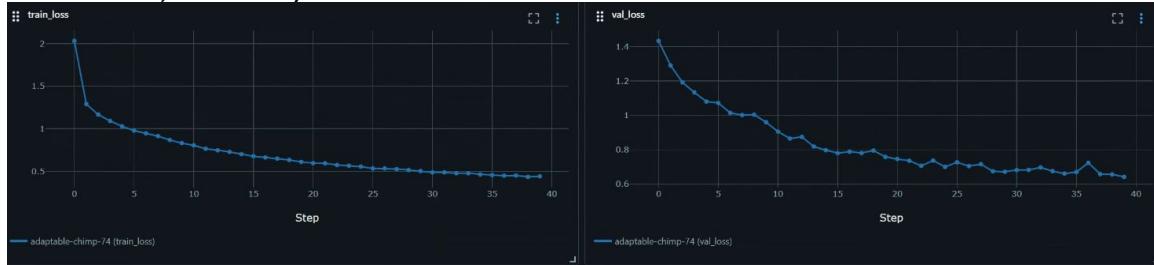
Report metrics separately for long vs short chromosomes, since short chromosomes are much harder to split.

Models

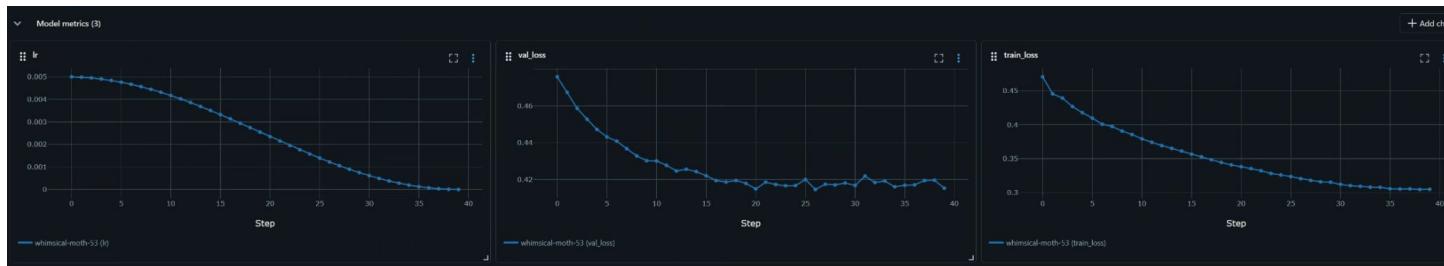
1. **Simple Mask R-CNN:** trained/validated only on the mask-annotated set (612 images, ~27k instances; train 547, test 65).
2. **Mask R-CNN + Attention RPN:** trained on both the box-annotated set (5,000 images, 229,852 instances) and the mask-annotated set (612 images), alternating batches (box batch, mask batch), EMA, LR scheduler, and warmup.

Models

1. **Simple Mask R-CNN:** trained/validated only on the mask-annotated set (612 images, ~27k instances; train 547, test 65).



2. **Mask R-CNN + Attention RPN:** trained on both the box-annotated set (5,000 images, 229,852 instances) and the mask-annotated set (612 images), alternating batches (box batch, mask batch), EMA, LR scheduler, and warmup.



1) Autokary test set (supervised evaluation)

| Training setup | mAP50 | PQ_all | mPQ | AJL |
|---|--------------|-----------|-----------|-----------|
| Mask-supervised (v0) | 0.885 | 0.582 | 0.581 | 0.504 |
| └ short chromosomes | 0.678 | 0.524 | 0.422 | 0.445 |
| └ long chromosomes | 0.770 | 0.625 | 0.534 | 0.534 |
| Mixed supervision (masks + boxes) + Attention RPN (v1) | 0.936 | 0.744 | 0.738 | 0.712 |
| └ short chromosomes | 0.786 | 0.689 | 0.541 | 0.641 |
| └ long chromosomes | 0.864 | 0.794 | 0.736 | 0.750 |

2) Unlabeled metaphase set (2,000 images) — sanity metric

| Model | Total | Count | Balance | Sex | Size |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|
| Mask-supervised (v0) | 0.302 | 0.001 | 0.289 | 0.043 | 0.874 |
| Attention model (v1) | 0.696 | 0.552 | 0.611 | 0.691 | 0.931 |

