NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS
Faculty of Informatics, Mathematics, and Computer Science
Department of Information Systems and Technology
.

Chromosome automatic segmentation and classification.

Nizhniy Novgorod
2023

Task: Metaphase images instance segmentation

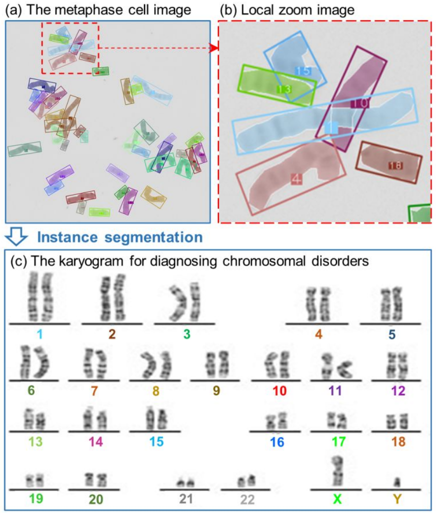
Goal: automatically analyze a metaphase microscopy image.

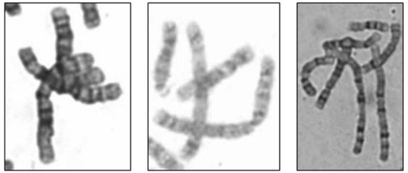
Input: one grayscale metaphase image containing a full chromosome spread.

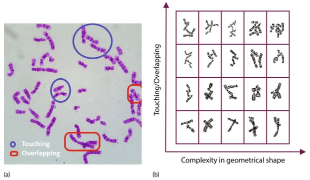
Output (per image)
Instance segmentation: detect and segment ~46 individual chromosome instances (≈23 pairs).
Per-instance classification: assign each instance one chromosome ID from 24 classes: {1–22, X, Y}.

Difficulty:

- Chromosomes are thin and deformable, often blurred and low-contrast.
- Many chromosomes overlap or touch, so a single connected region may contain multiple true instances.
- Several chromosome classes are visually very similar, so classification errors are easy even when segmentation is correct.
- Properly labeled datasets are very few.




(a) (b) (c)



Datasets

Dataset 1 — Cell Image Library “24 chromosomes” (boxes, 24-class)

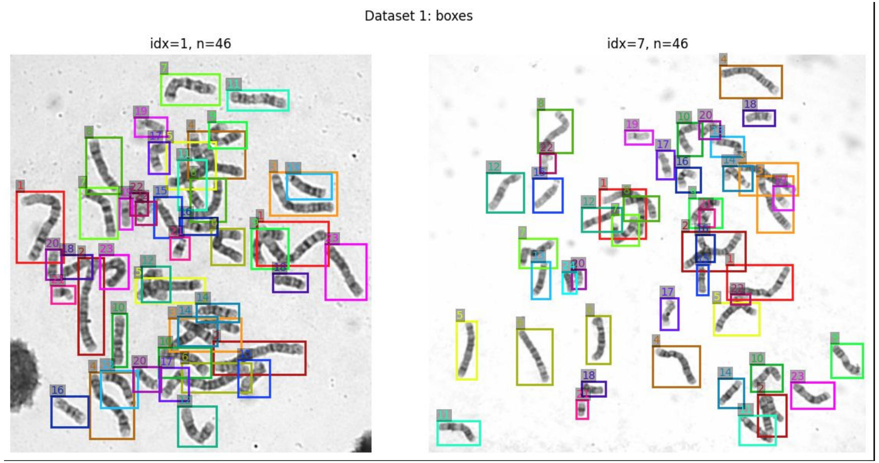
- **Source:** Kaggle
- **Type:** metaphase spreads with **24-class bounding boxes** (1–22, X, Y)
- **Scale:** 5,000 images, 229,852 chromosome instances
- **Annotations:** box-only (Pascal-VOC XML → converted)
- **Typical size:** ~640 px (median)

Dataset 2 — AutoKary2022 / Chromosome Instance Segmentation (masks, 24-class)

- **Source:** GitHub)
- **Type:** metaphase spreads with instance masks + 24-class label per chromosome
- **Scale:** 612 images, ~27,000 chromosome instances
- **Image size:** 1600×1600
- **Split:** 547 train / 65 test, patient-level separation (no patient overlap)

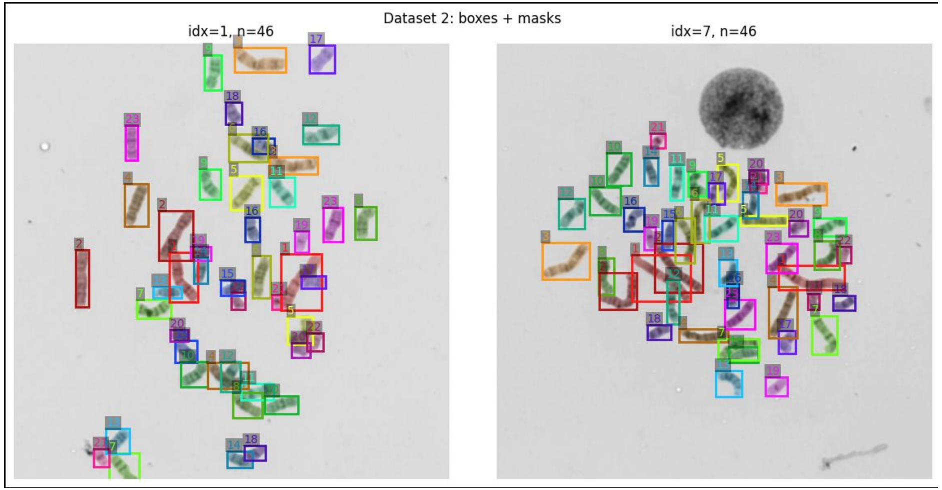
Dataset 3 — Small custom-made (masks,no classification)

- **Type: 10** metaphase images with instance masks only (no chromosome IDs)


Dataset 1: boxes

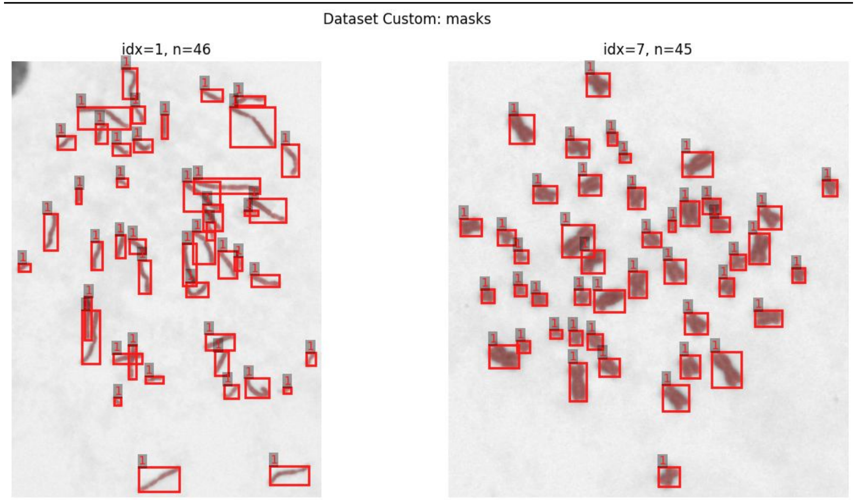
idx=1, n=46

idx=7, n=46


Dataset 2: boxes + masks

idx=1, n=46

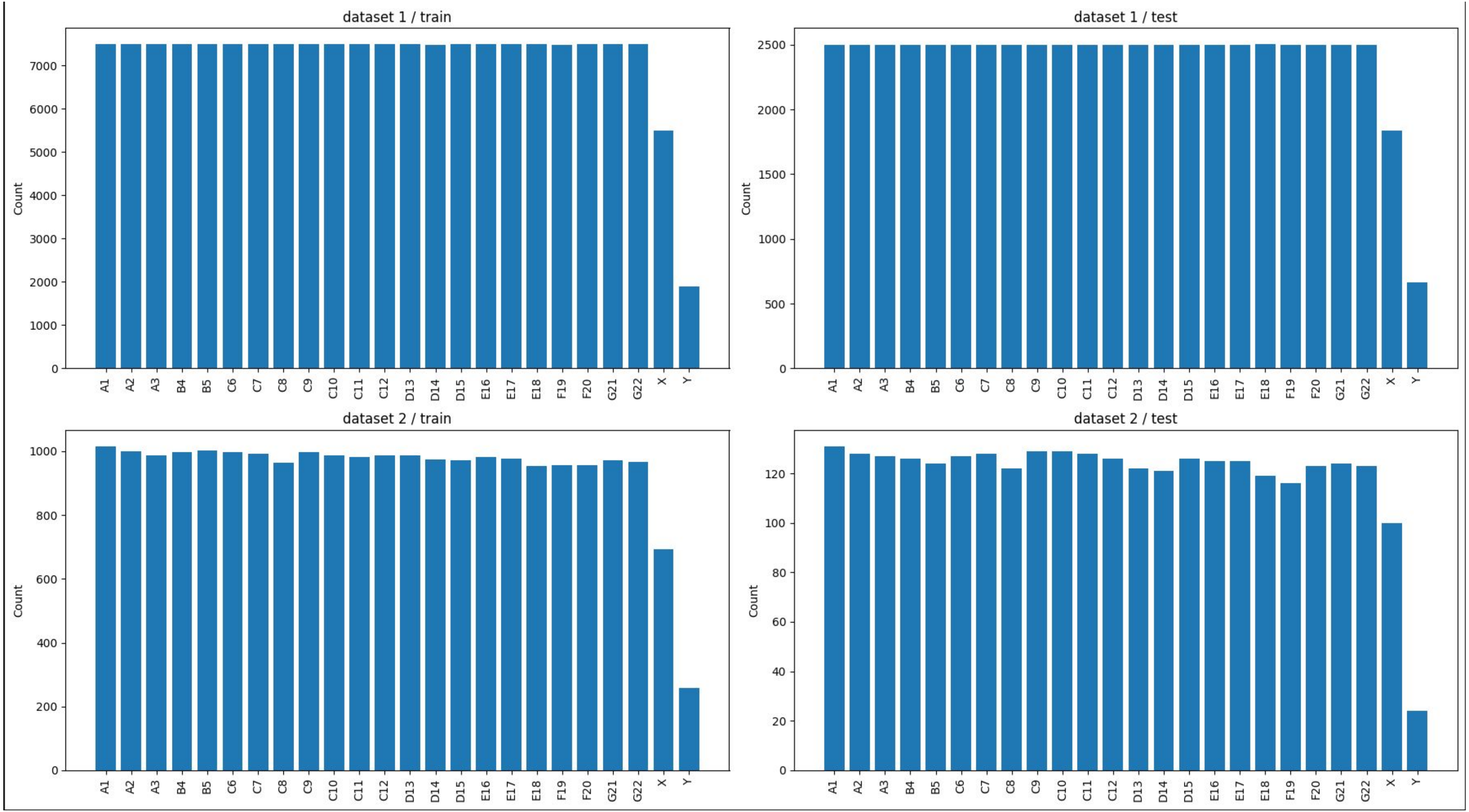
idx=7, n=46


Dataset Custom: masks

idx=1, n=46

idx=7, n=45

Class counts for individual chromosomes



Models

1) Simple Mask R-CNN (Baseline)

Train + validate only on Dataset 2 (612 images with instance masks + 24-class IDs).

2) Mask R-CNN — 2-Stage Training (boxes, masks)

Stage 1: pretrain detector/classifier on Dataset 1 (box-only, 24 classes).

Stage 2: fine-tune full Mask R-CNN (incl. mask head) on Dataset 2 (instance masks).

3) Mask R-CNN + Attention FPN — Alternating Batches

Train on Dataset A + Dataset B together, alternating batch types:

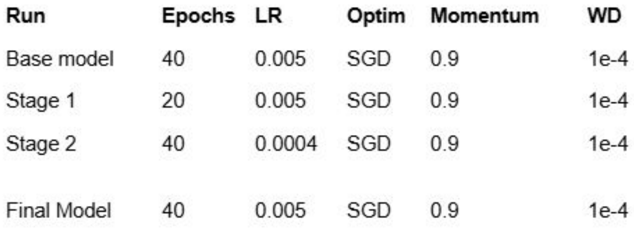
Box batch (A): optimize detection losses only (RPN + box cls/reg).

Mask batch (B): optimize detection + mask losses.

Purpose: train detector on larger box dataset while learning masks from smaller dataset.

Training setup

Model: MaskRCNN
Batch size: 4
Backbone: ResNet-50
Input size: 640x640, grayscale image
Train is further split into train=0.9, validation=0.1 each dataset.
Basic augmentation is always applied (vertical/horizontal flip, rotation, brightness/contrast)

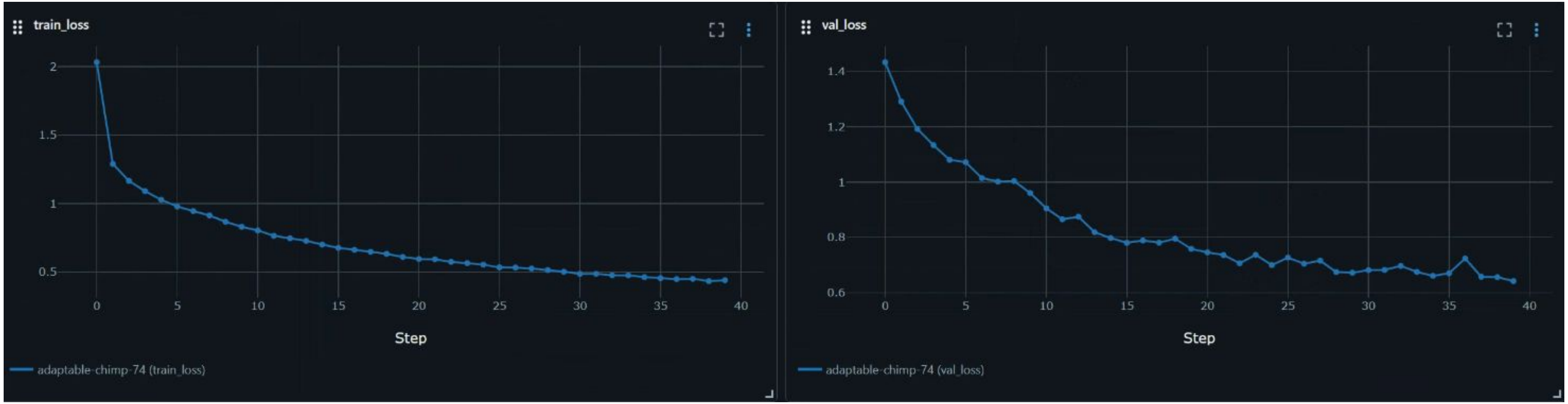


Run	Epochs	LR	Optim	Momentum	WD
Base model	40	0.005	SGD	0.9	1e-4
Stage 1	20	0.005	SGD	0.9	1e-4
Stage 2	40	0.0004	SGD	0.9	1e-4
Final Model	40	0.005	SGD	0.9	1e-4

Metric (From AutoKary2022: A Large-Scale Densely Annotated Dataset for Chromosome Instance Segmentation):
mAP@50
Average precision across classes for detections with IoU ≥ 0.5 and correct class; measures object detection accuracy (localization + classification).

AJI (aggregated jaccard index)
Sum of intersection areas of greedily matched instances (correct class) divided by sum of their unions plus all unmatched predictions; measures instance segmentation accuracy, especially for dense/overlapping objects.

1. Simple Mask R-CNN: trained/validated only on the mask-annotated set images.

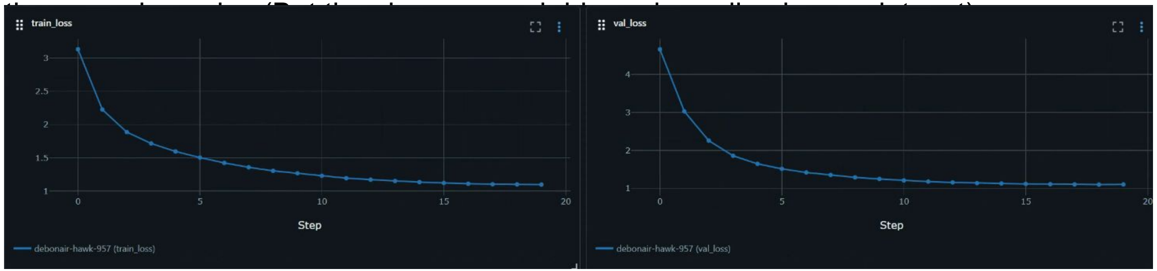


2. Two-stage Model

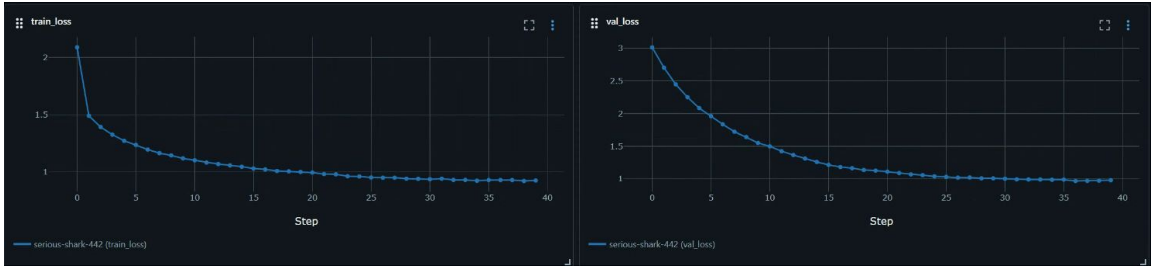
Pretraining instance segmentation models with bounding box annotations (Agnew et al., 2024)

The idea from that paper is to train in two stages.

- **Stage 1 (weak):** turn bounding boxes into coarse polygon masks (basically the 4 box corners) and pretrain an instance-seg model on [illegible]



- **Stage 2 (strong):** fine-tune on real masks.



3. Train one Mask R-CNN on mixed supervision

We would like to train simultaneously so that model learns detection from both and mask quality from B. So we use alternating batching.

Papers:

1) "Learning to Segment Every Thing" (Ronghang Hu et al., CVPR 2018)

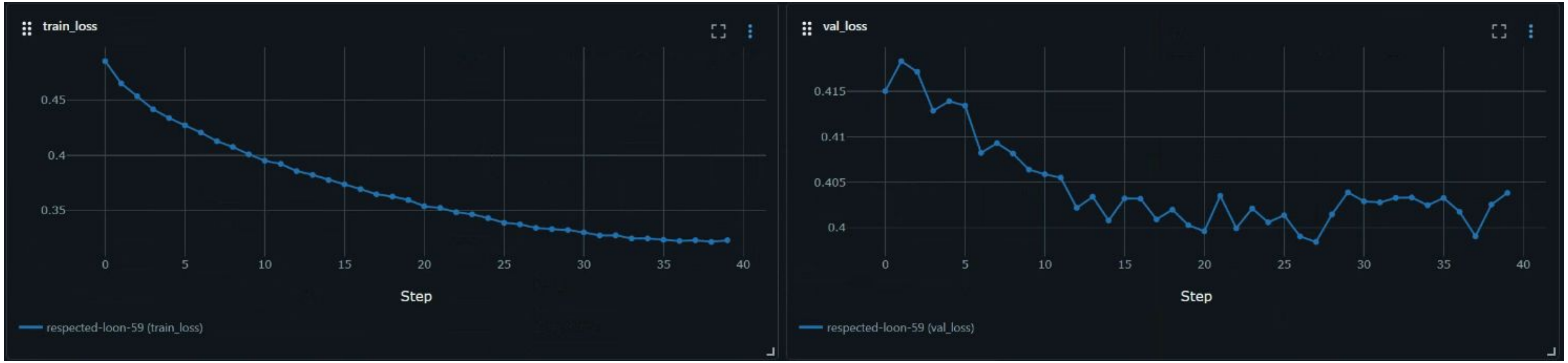
Their problem is class mismatch: some classes have masks, other classes only have boxes. That can bias training, so they stop gradients from the mask loss into class-specific detection weights. In our case (same classes in both datasets) this specific issue mostly doesn't apply.

2) "DaTaSeg: Taming a Universal Multi-Dataset Multi-Task Segmentation Model" (Xiuye Gu et al., NeurIPS 2023)

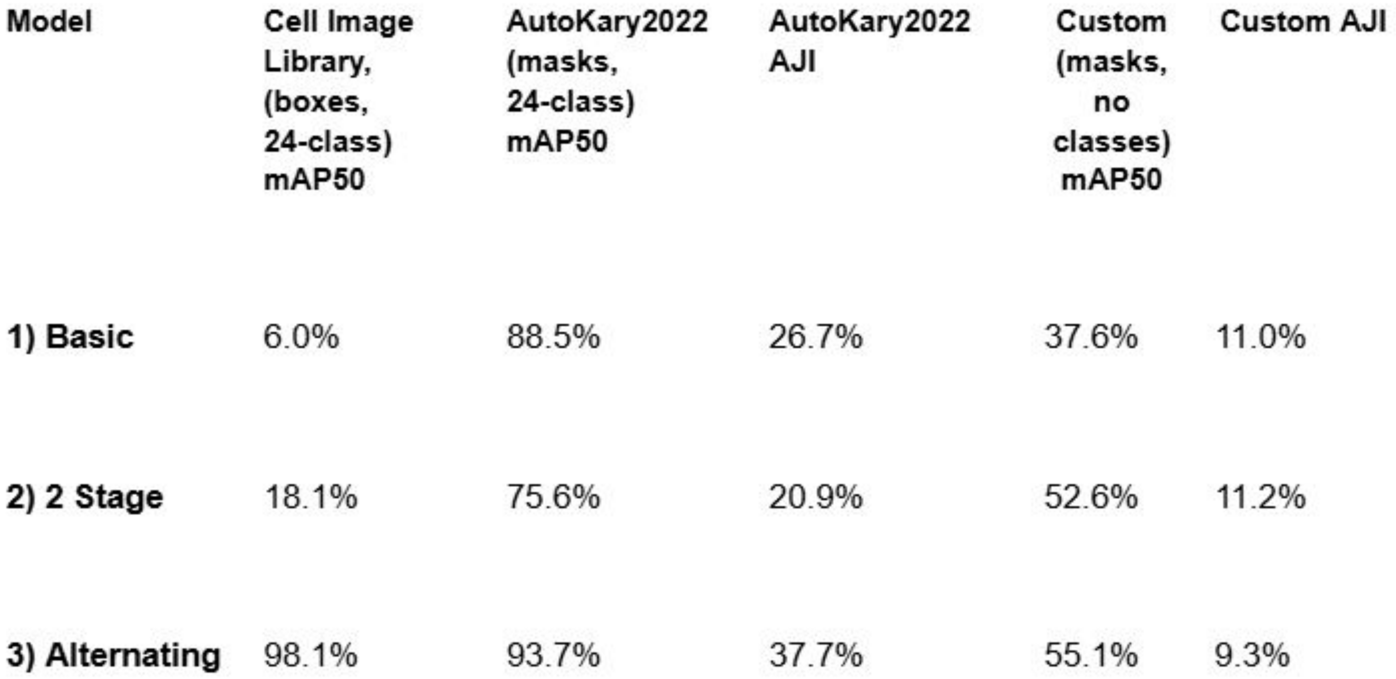
The idea is to alternate batches from different datasets in one run. Compute only the losses that exist for that batch (box-only - detection losses, mask batch - detection + mask). For box-only supervision, they add a projection loss to give the mask branch a weak training signal from boxes.

3. Train one Mask R-CNN on mixed supervision

- Each iteration: pick dataset (1 or 2) by a ratio, train on that batch.
- Losses depend on what supervision exists for that batch.
- If larger dataset dominates or has different image statistics, the shared backbone can drift toward it; mask quality will be worse.

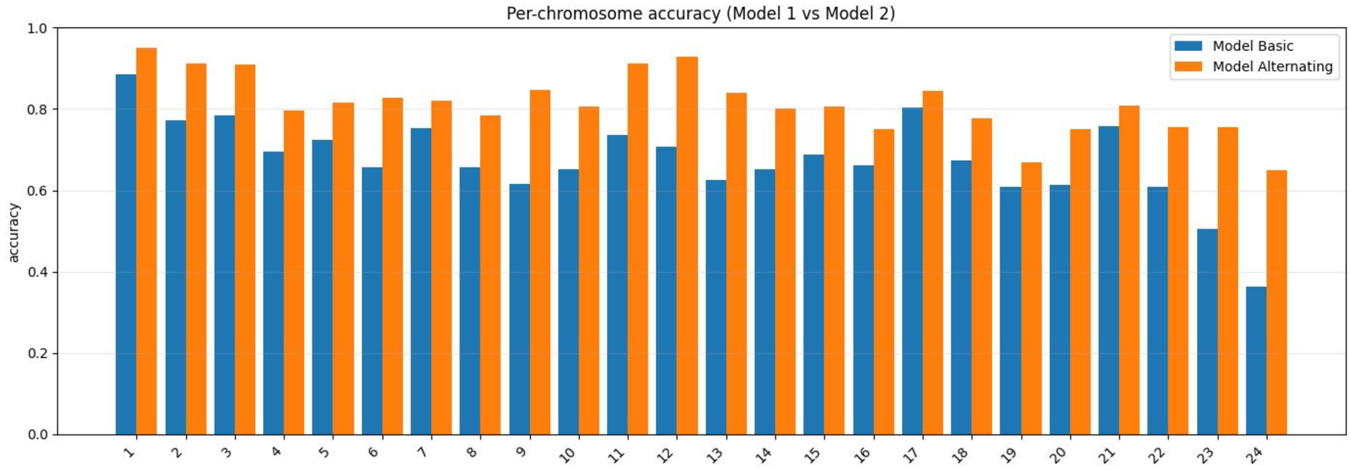


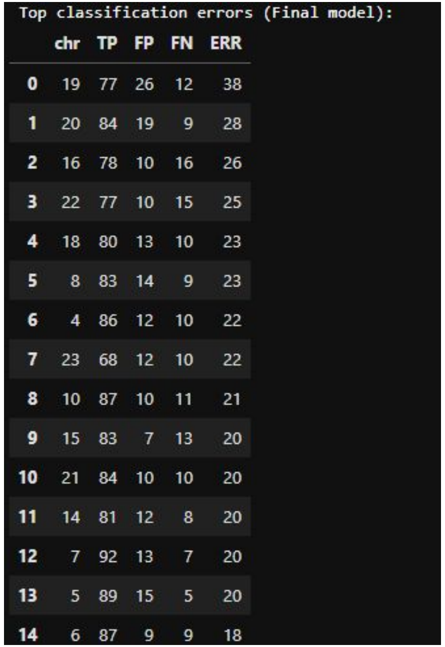
Evaluation on Test Sets



Model	Cell Image Library, (boxes, 24-class) mAP50	AutoKary2022 (masks, 24-class) mAP50	AutoKary2022 AJI	Custom (masks, no classes) mAP50	Custom AJI
1) Basic	6.0%	88.5%	26.7%	37.6%	11.0%
2) 2 Stage	18.1%	75.6%	20.9%	52.6%	11.2%
3) Alternating	98.1%	93.7%	37.7%	55.1%	9.3%

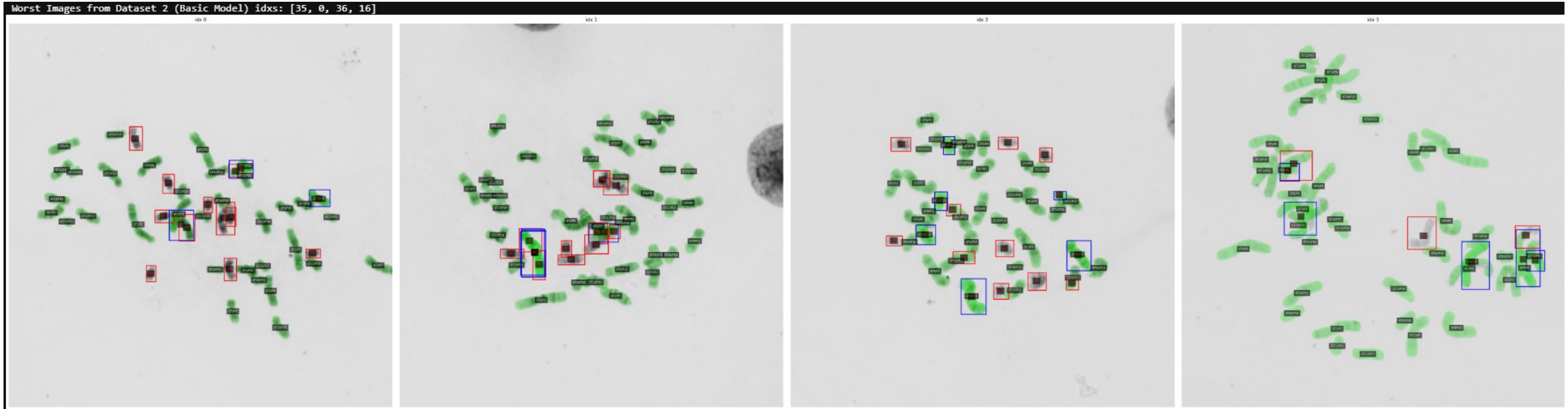
Classification accuracy for each chromosome



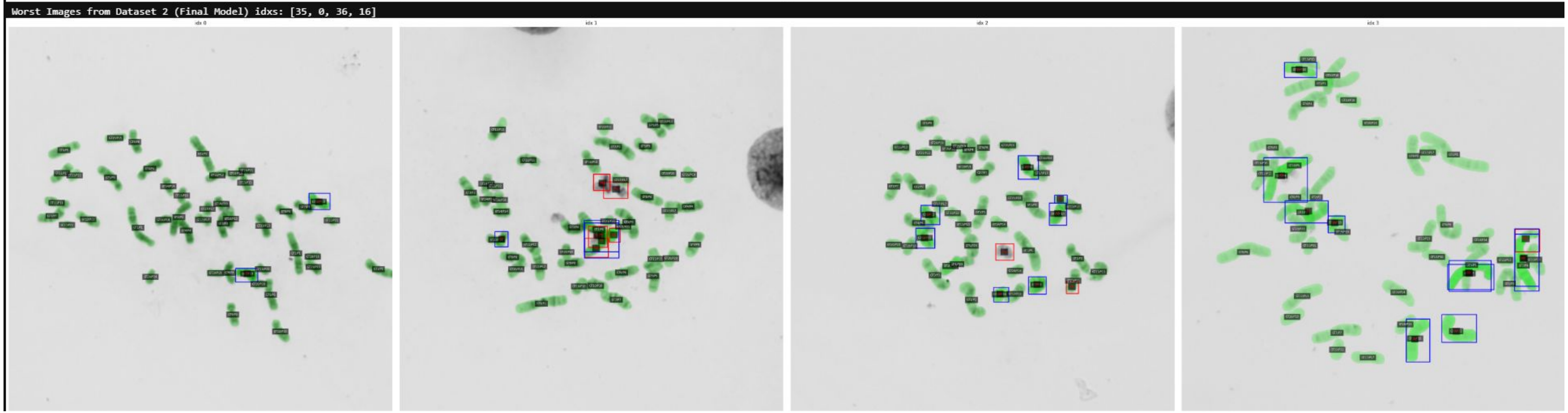

```
Top classification errors (Final model):
```

	chr	TP	FP	FN	ERR
0	19	77	26	12	38
1	20	84	19	9	28
2	16	78	10	16	26
3	22	77	10	15	25
4	18	80	13	10	23
5	8	83	14	9	23
6	4	86	12	10	22
7	23	68	12	10	22
8	10	87	10	11	21
9	15	83	7	13	20
10	21	84	10	10	20
11	14	81	12	8	20
12	7	92	13	7	20
13	5	89	15	5	20
14	6	87	9	9	18

Worst images from Dataset 2 (FN & FP) Basic vs Final Model

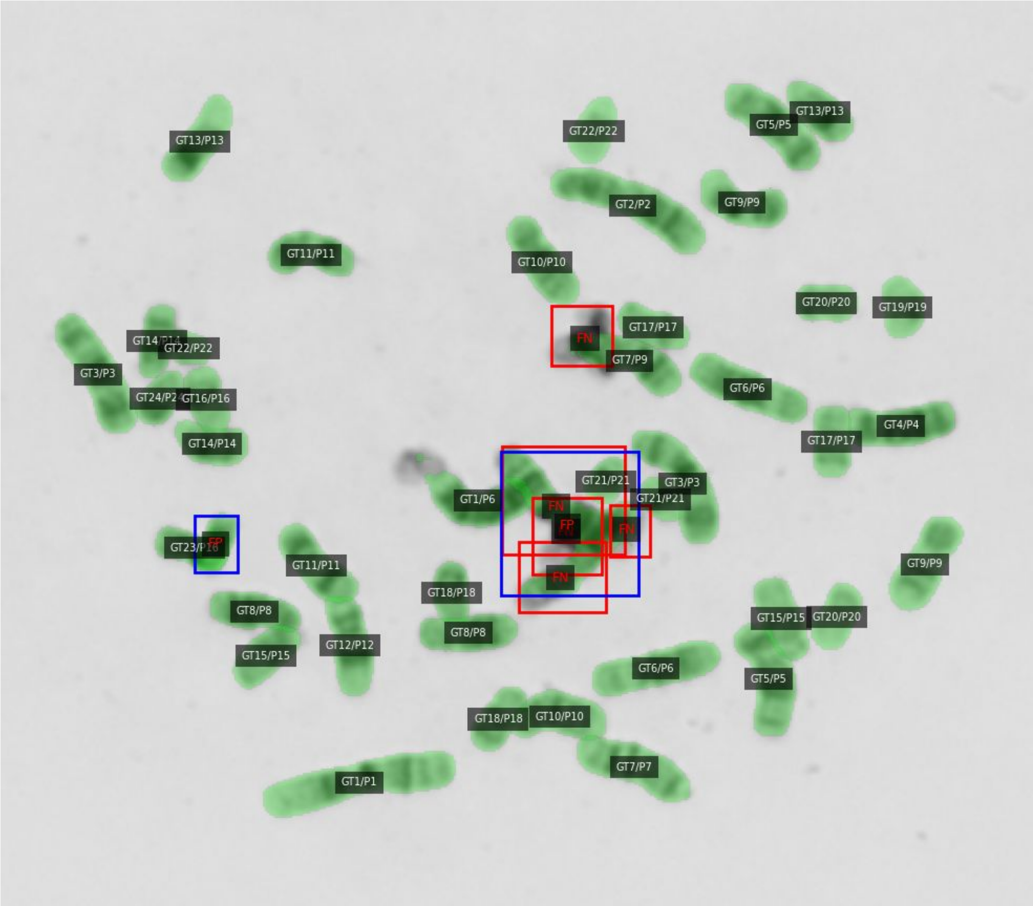


Worst Images from Dataset 2 (Basic Model) idxs: [35, 0, 36, 16]

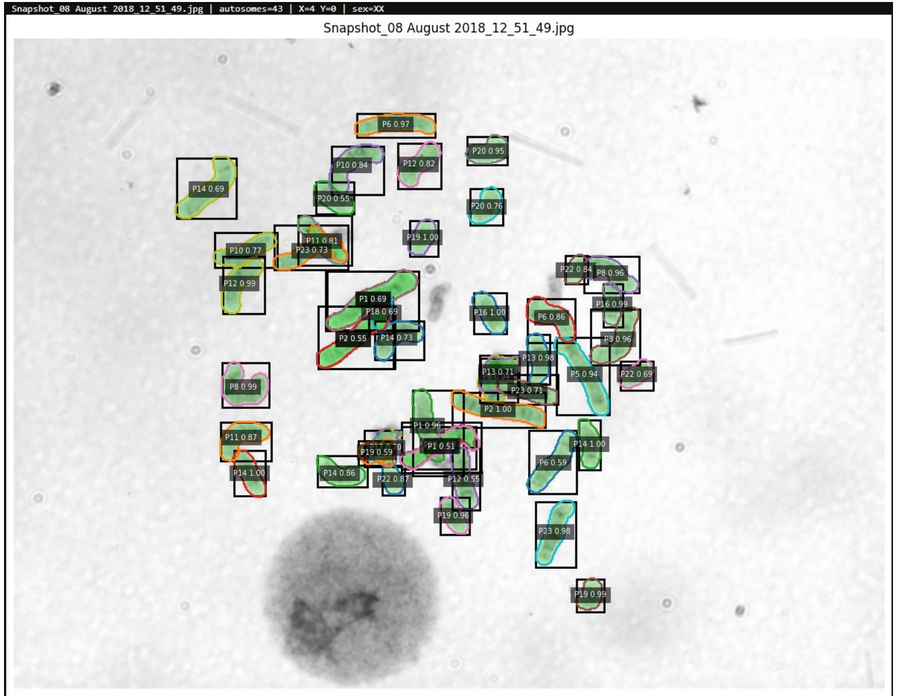


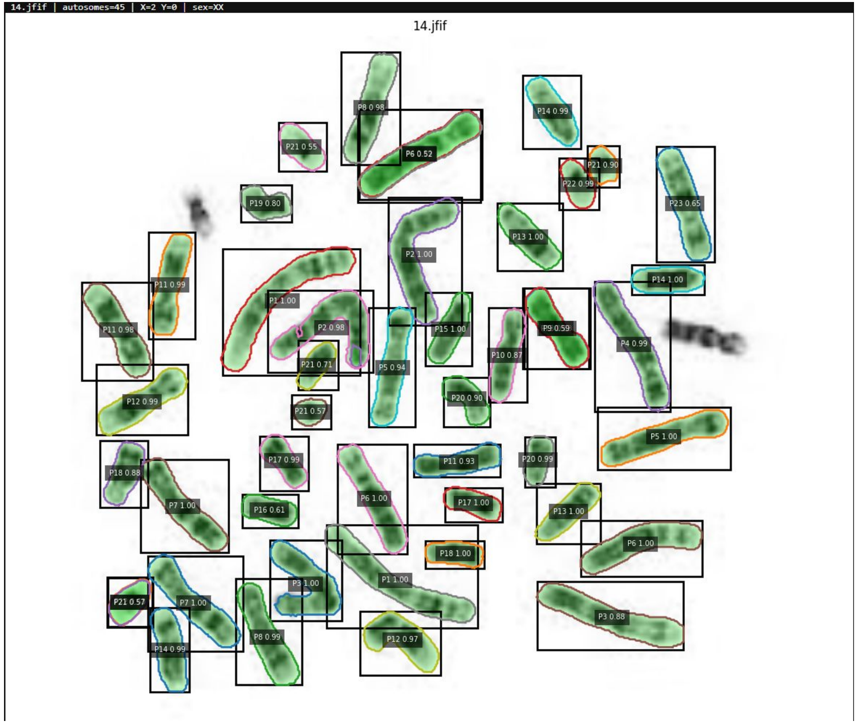
Worst Images from Dataset 2 (Final Model) idxs: [35, 0, 36, 16]

Errors for short chromosomes



Evaluation on random metaphase images collected from Google





Current Issues

- Overlaps: 2–3 chromosomes often merge into one mask.
- Short chromosomes are harder. More misses and bad splits.

Next steps

- Multi-stage inference: main model, overlap model (trained on overlap dataset), classification model (trained on classification dataset).
- Other instance-segmentation models beyond Mask R-CNN (Cascade RCNN, HTC, PolarMask++)
- More sane mixed model training
- Postprocessing