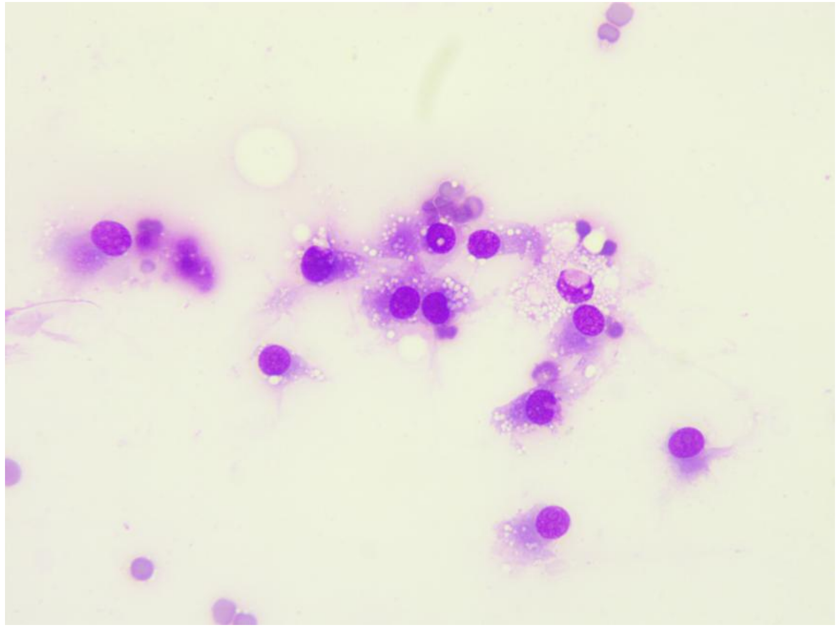
Deep Learning for Cancer Cell Detection in Veterinary Cytology

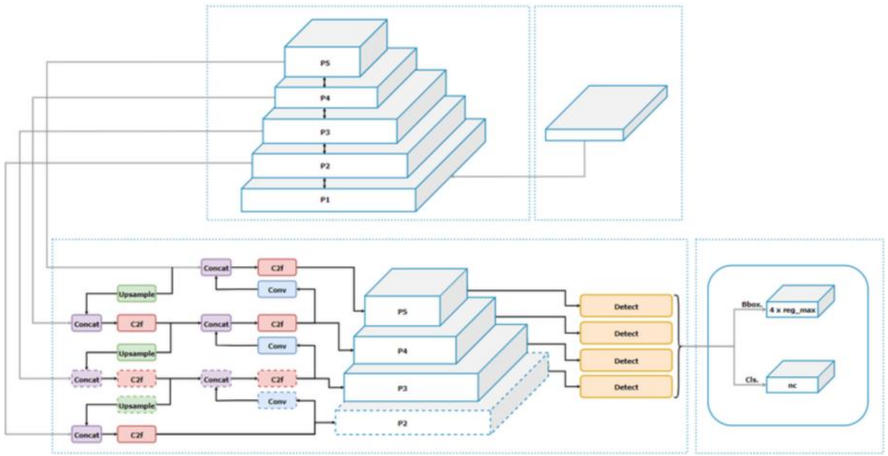
Jan Krupiński, Ernest Jamro, Maciej Wielgosz, Paweł Russek,
Agnieszka Dąbrowska-Boruch, Kazimierz Wiatr

Cytological Examination of Skin Lesions



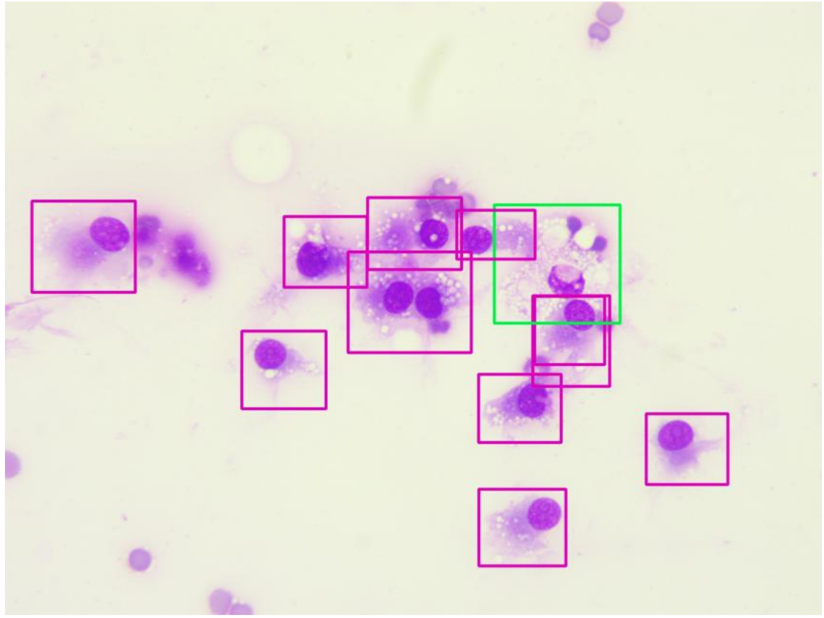
- Quick and minimally invasive skin cancer diagnostics
- Requires expert knowledge

Cancer Cell Detection with Deep Learning



http://dx.doi.org/10.1109/ACCESS.2023.3293056

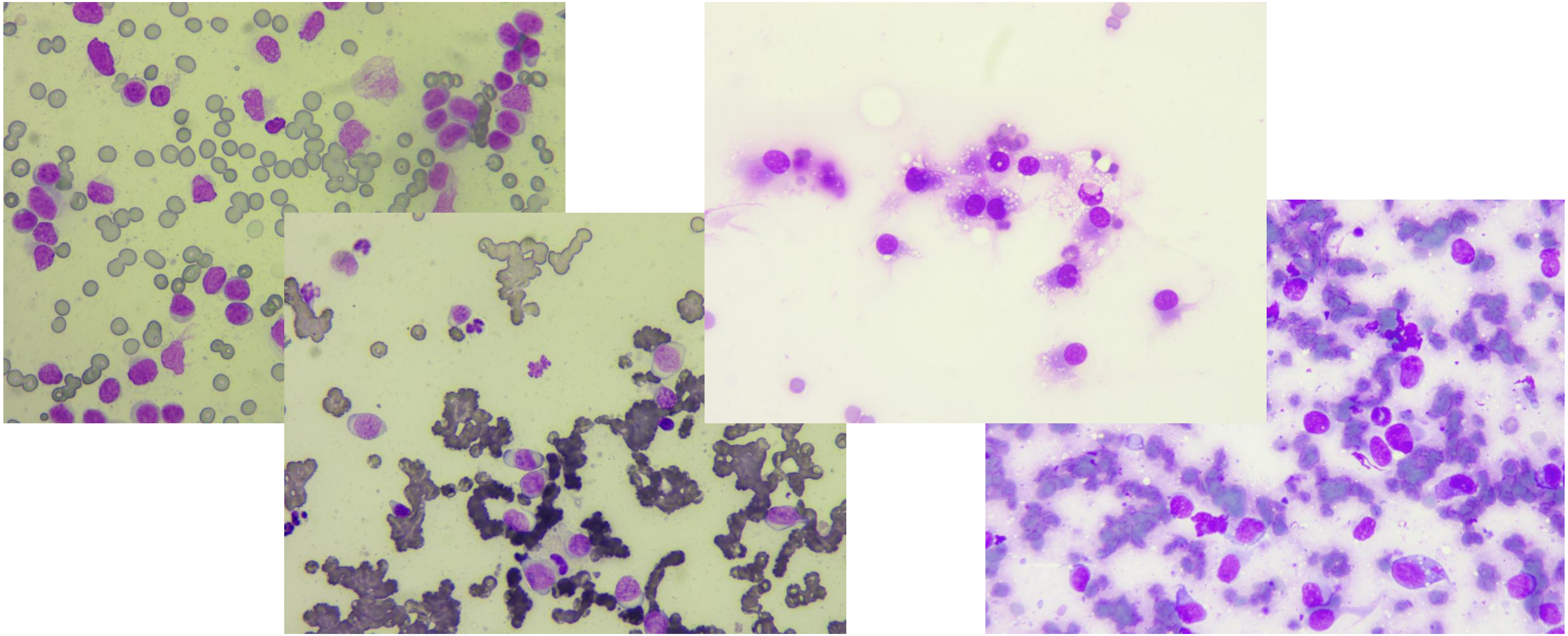
Deep Learning model



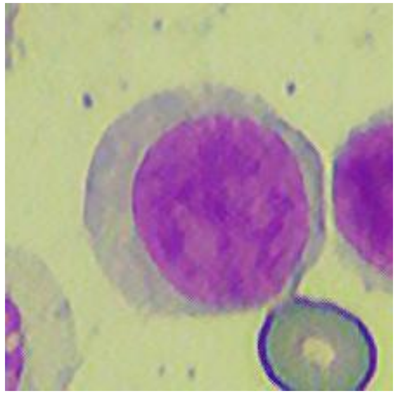
Example detection:
Histiocytic Sarcoma cells

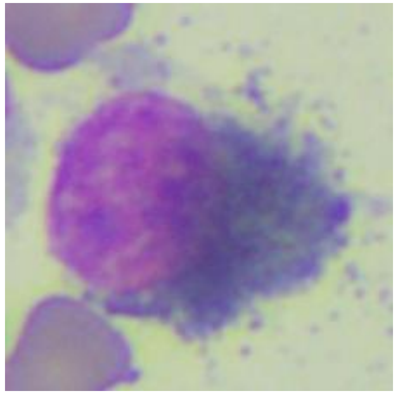
Dataset of Cytological Images

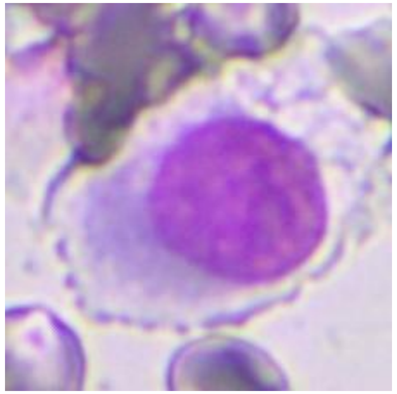
2,146 images (split 60-20-20% training, validation, testing)

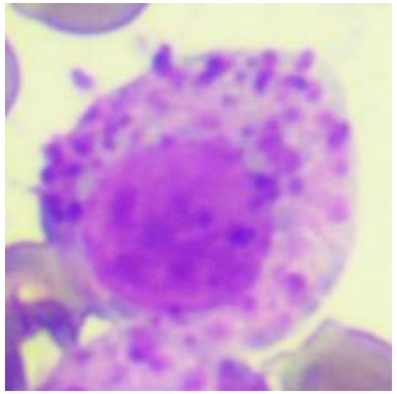


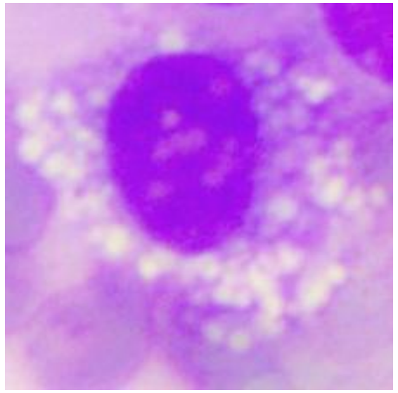
Cancer Cell Types

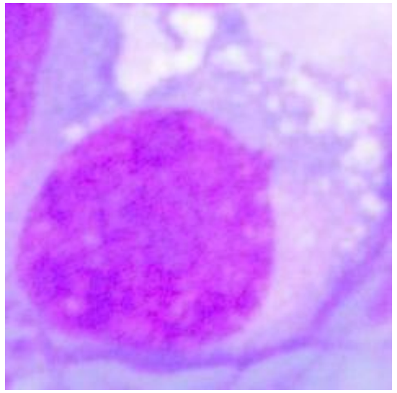

Lymphoma

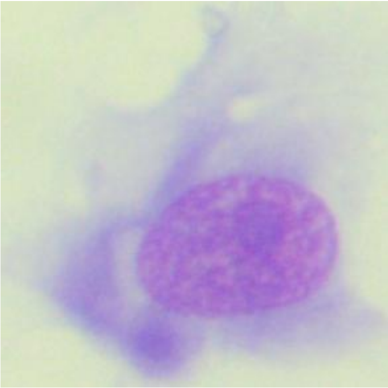

Melanoma

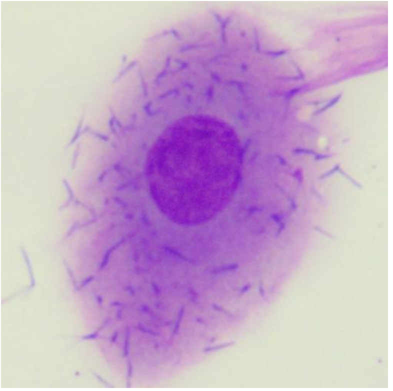

Histiocytoma


MCT

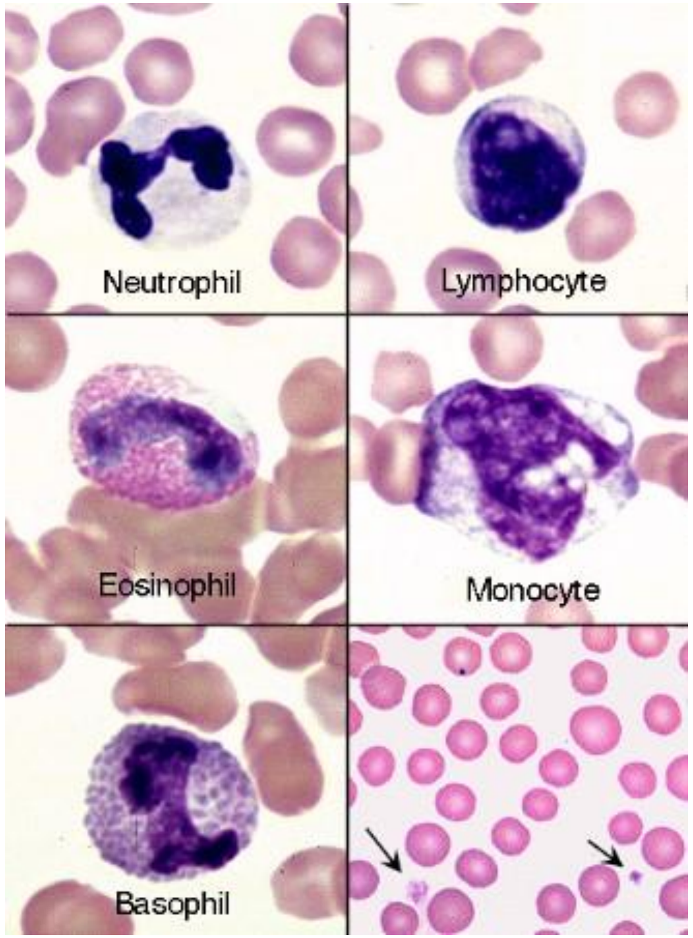

Histiocytic
Sarcoma


Plasmacytoma

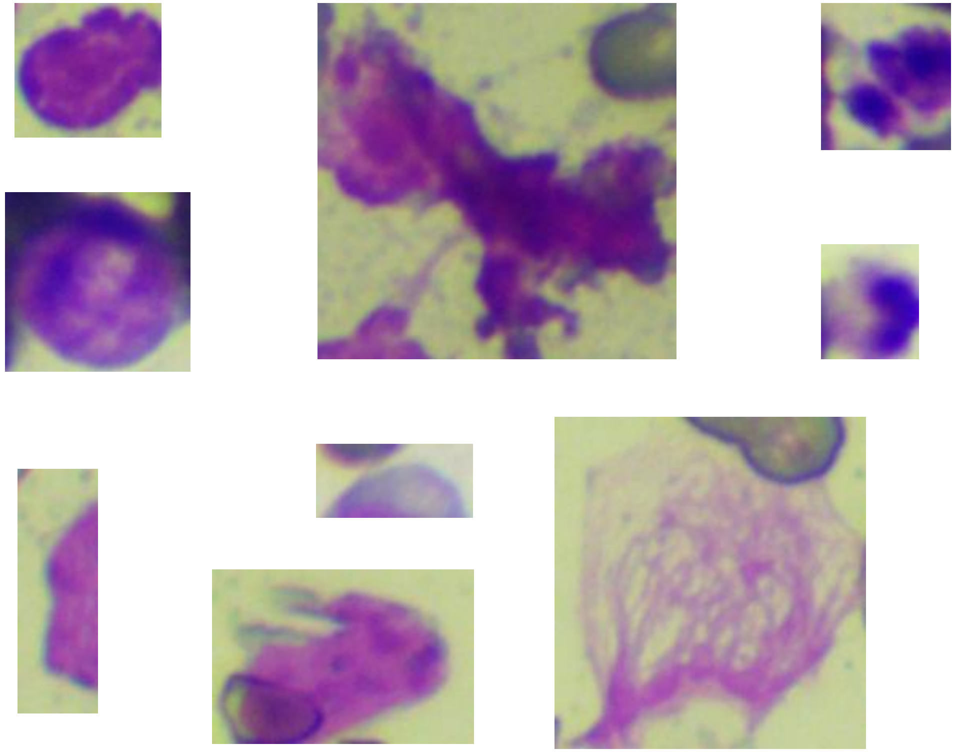

STS


Hepatoid

Other Cell Types



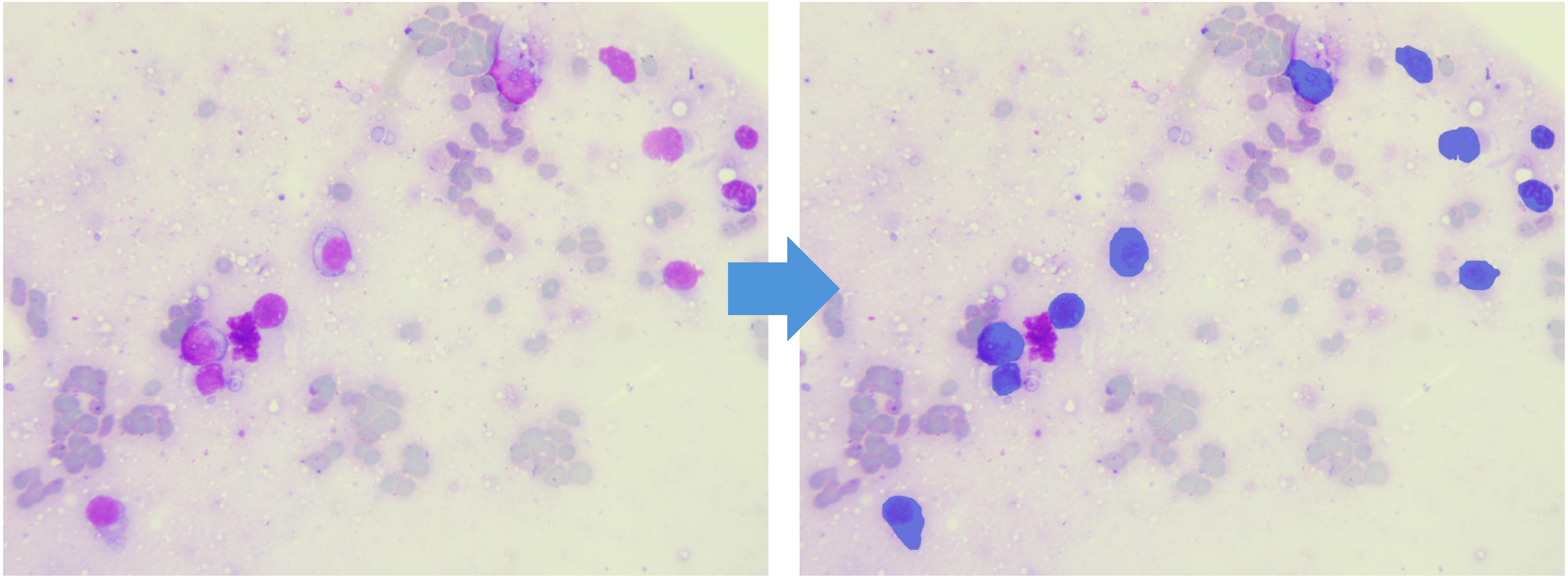
white blood cells



non-diagnostic category
(40% of all cells)

Labelling Cells for Segmentation

65,684 cells in **13 classes**



Segmentation Results

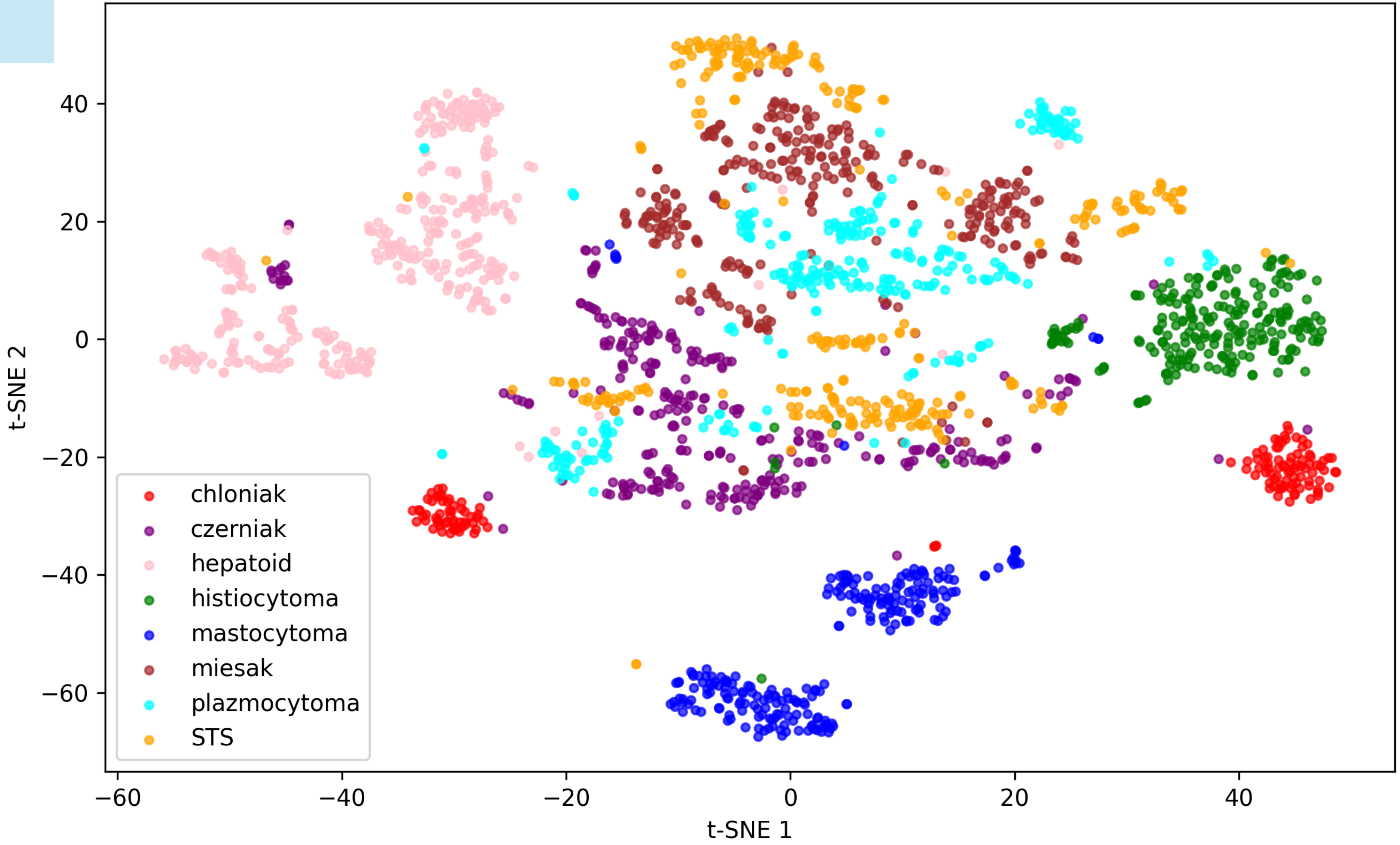
- 68% of errors are undetected non-diagnostic cells
- Some cells were missed by the specialist
- The rest of the errors can easily be eliminated (only one type of cancer can be present at a time)

IoU: 0.75, score: 0.75

Predictions \ Ground truth labels	Neoplastic mast cell	Cut cell	Unrecognized cell	Macrophage	Damaged cell	Eosinophil	Neutrophil	Histiocytoma	Lymphocyte	Lymphoma	Giant cell	not present
Neoplastic mast cell	2429	11	57	0	19	0	0	0	0	0	0	46
Cut cell	5	985	3	0	6	2	2	4	1	3	0	325
Unrecognized cell	4	3	513	11	62	1	4	25	22	3	0	142
Macrophage	0	0	31	171	2	0	0	2	0	0	10	9
Damaged cell	15	9	52	1	974	0	14	1	2	11	0	232
Eosinophil	0	0	3	0	0	42	0	0	0	0	0	5
Neutrophil	0	3	84	0	25	2	1436	0	0	0	0	90
Histiocytoma	0	11	73	1	1	0	0	752	0	0	0	54
Lymphocyte	0	0	1	0	0	0	0	0	17	0	0	0
Lymphoma	0	2	52	1	20	0	0	0	0	1170	0	39
Giant cell	0	0	0	0	0	0	0	0	0	0	2	0
not detected	179	432	1161	68	802	33	448	180	129	66	2	0

Predictions / Ground truth labels

T-SNE Data Visualization



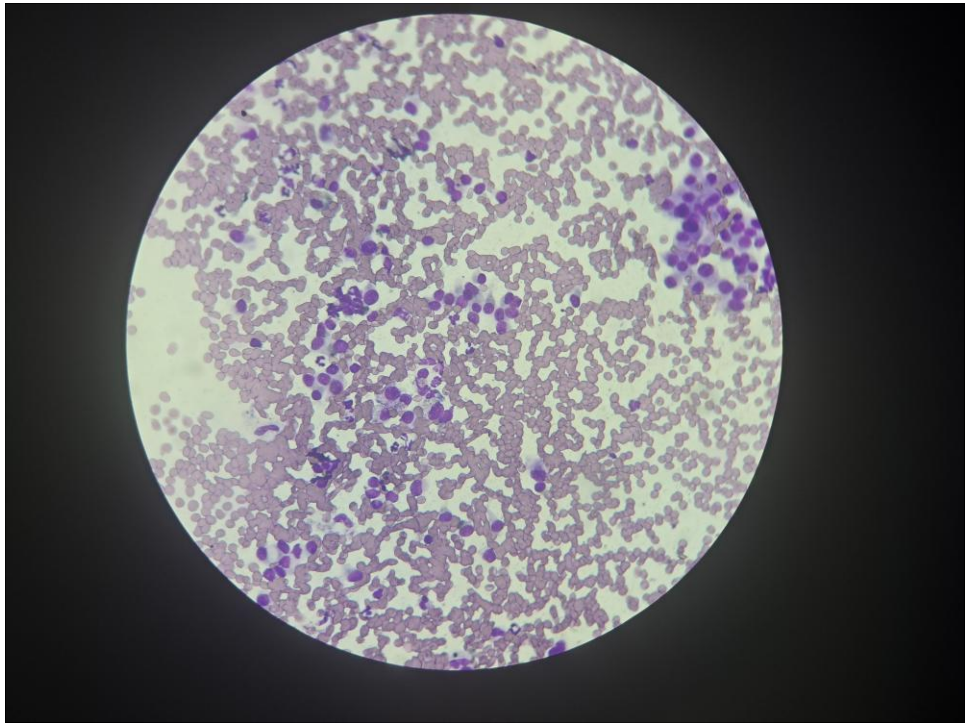
ViT-L Model
https://arxiv.org/abs/2010.11929
Output

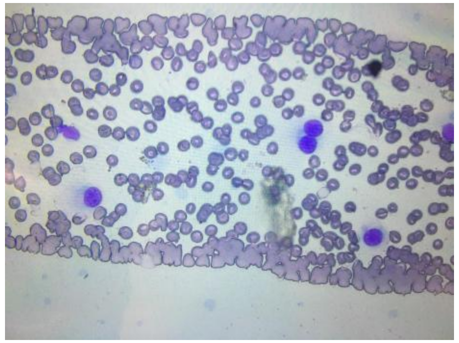
Weights:
(DINOv2
https://arxiv.org/abs/2304.07193

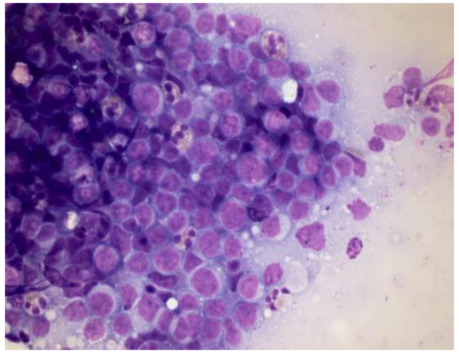
/

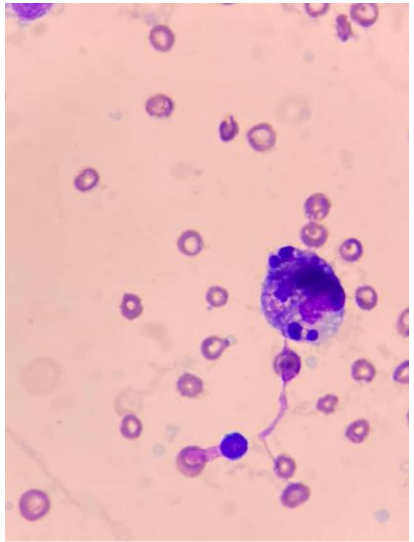
DinoBloom)
https://dl.acm.org/doi/10.1007/978-3-031-72390-2_49

Third-Party Data









- Testing on images submitted by a potential user

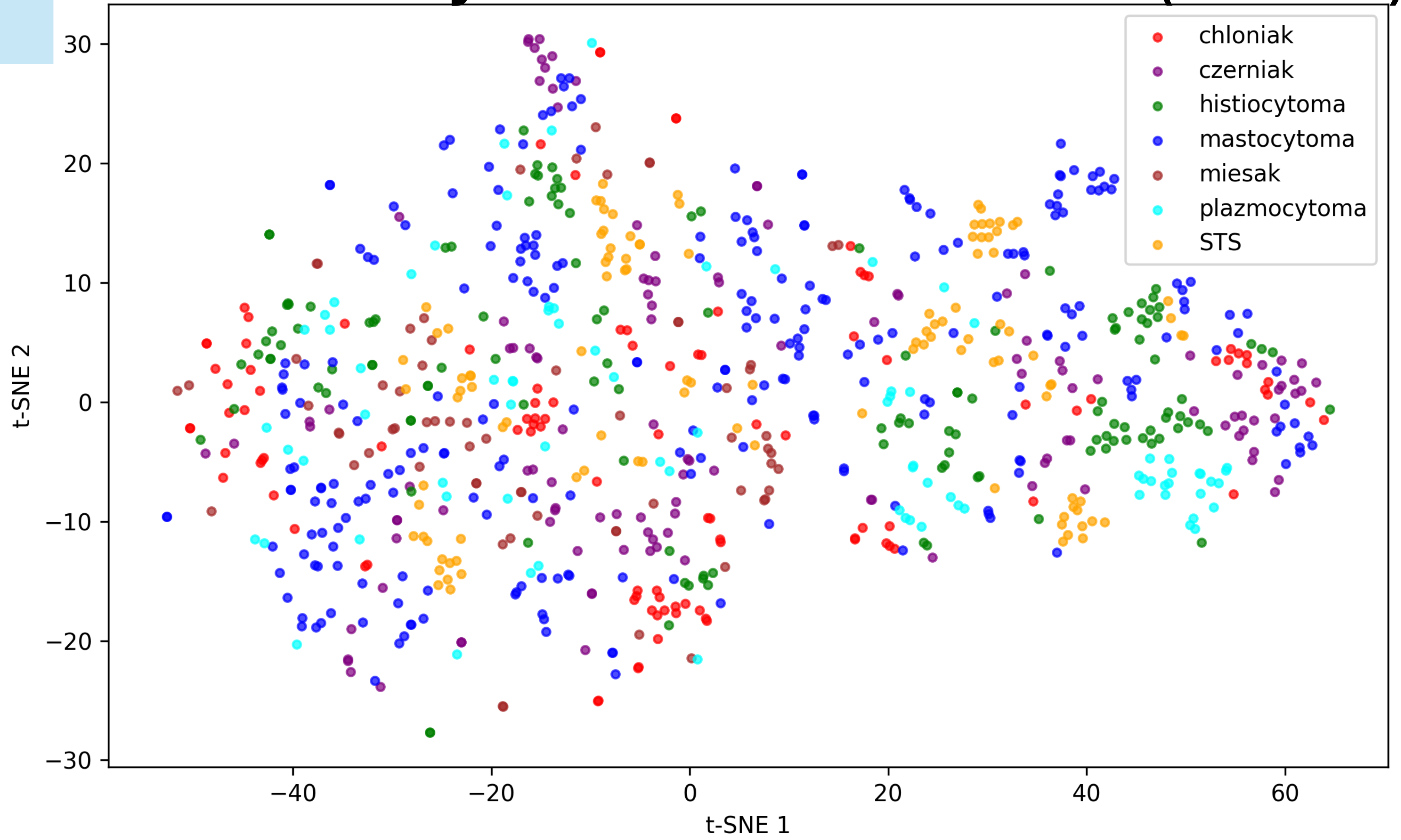
Cell Detection on Third-Party Data

Cancer Cell Recall:

	Training data	New images	Phone images	Third-Party data
lymphoma	93%	53%	26%	1%
MCT	92%	96%	38%	35%
melanoma	83%	58%	56%	32%
average	**90%**	**70%**	**40%**	**23%**

Significant drop in model performance.

Third-Party Data Visualization (t-SNE)



ViT-L Model

https://arxiv.org/abs/2010.11929

Output

Weights:

(DINOv2

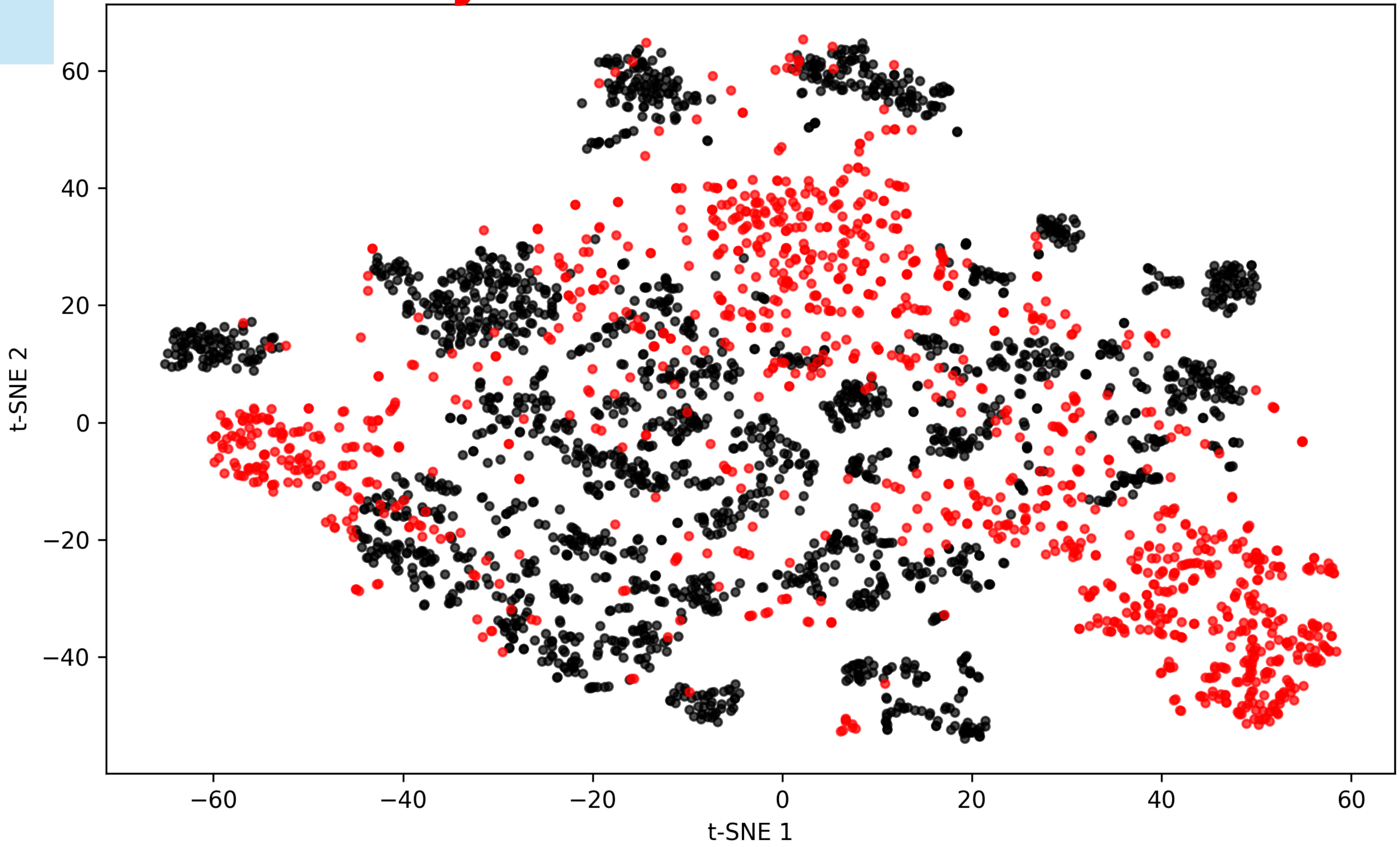
https://arxiv.org/abs/2304.07193

/

DinoBloom)

https://dl.acm.org/doi/10.1007/978-3-031-72390-2_49

Third-Party Data vs Internal Data



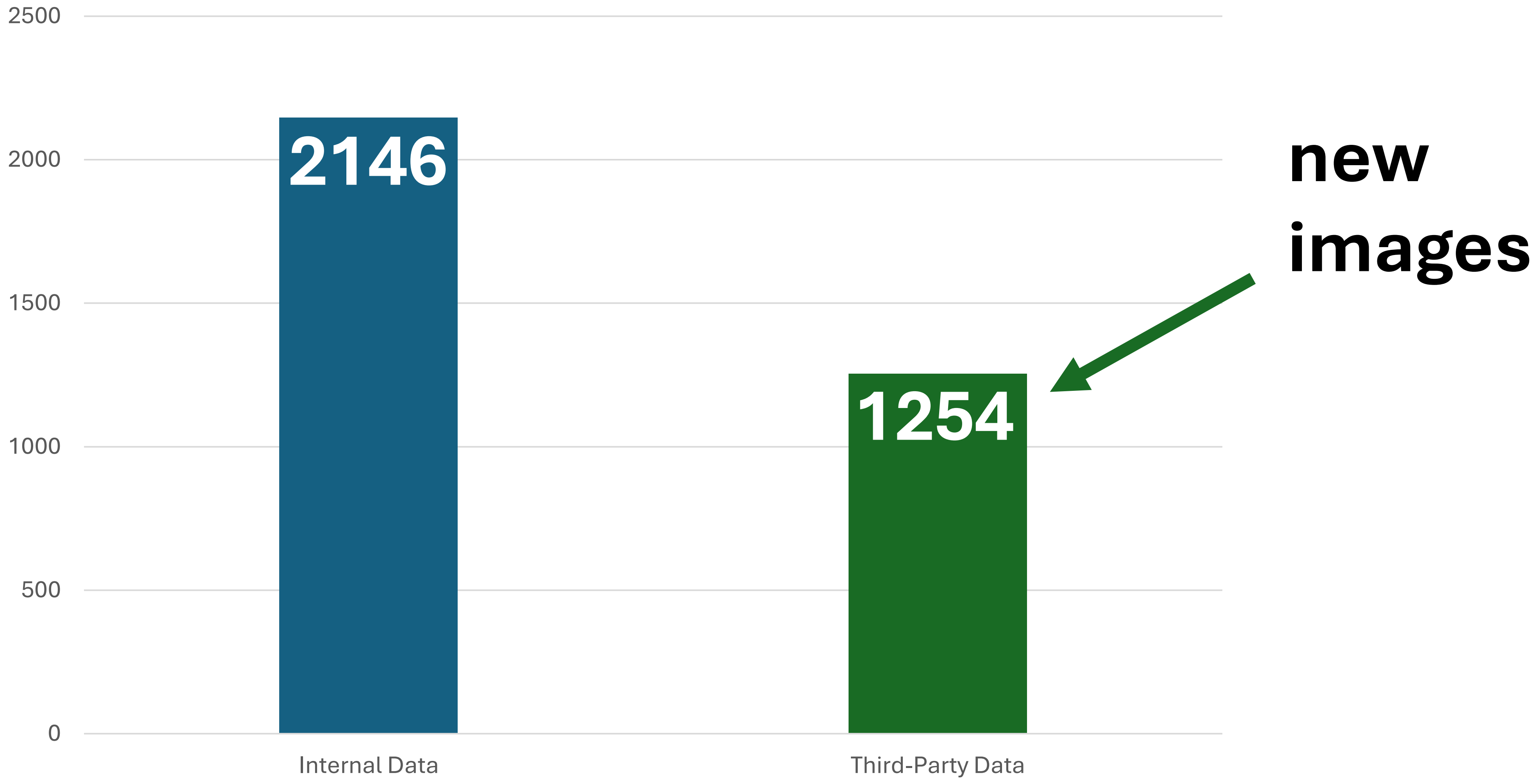
ViT-L Model
https://arxiv.org/abs/2010.11929
Output

Weights:
(DINOv2
https://arxiv.org/abs/2304.07193

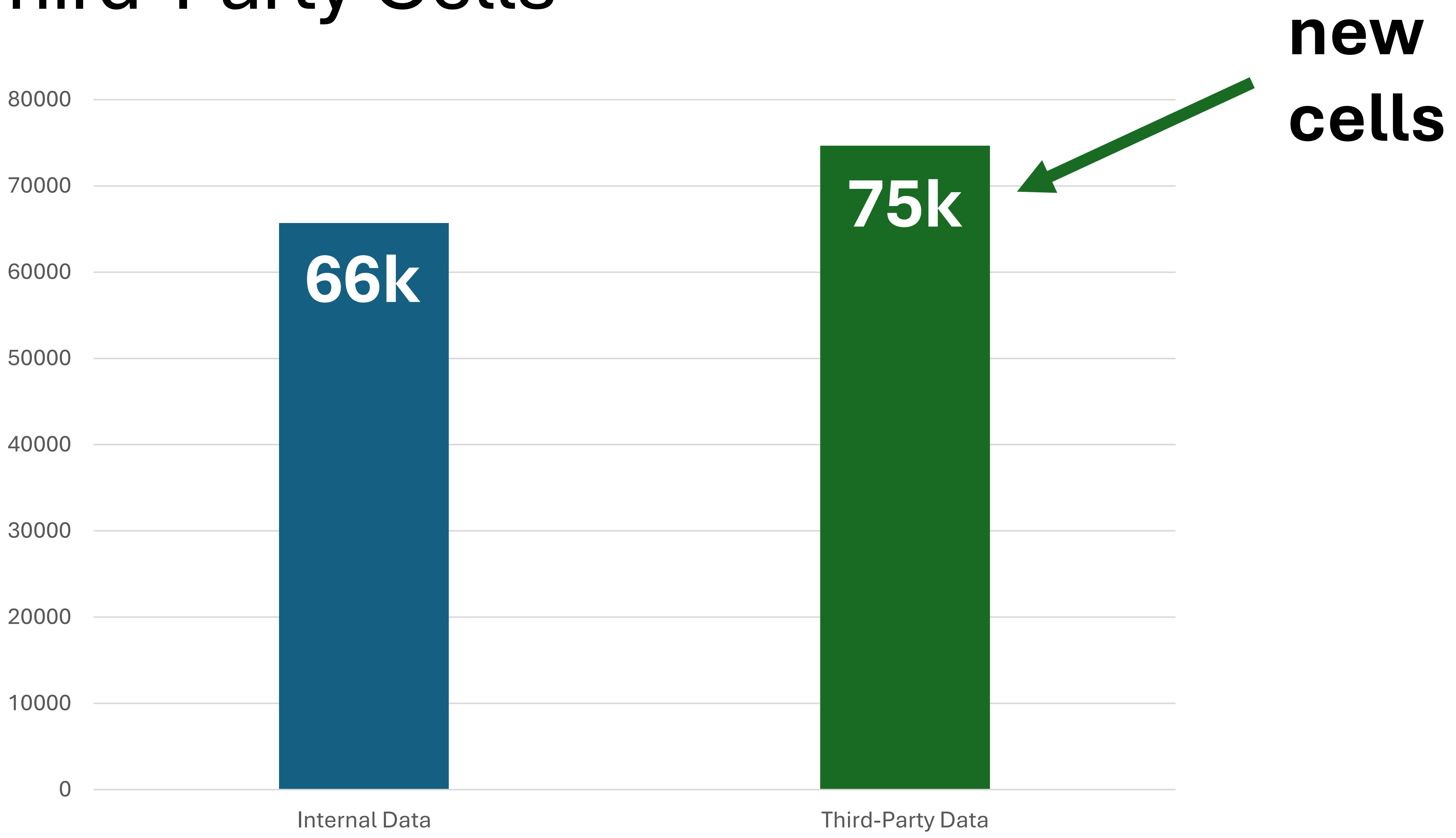
/

DinoBloom)
https://dl.acm.org/doi/10.1007/978-3-031-72390-2_49

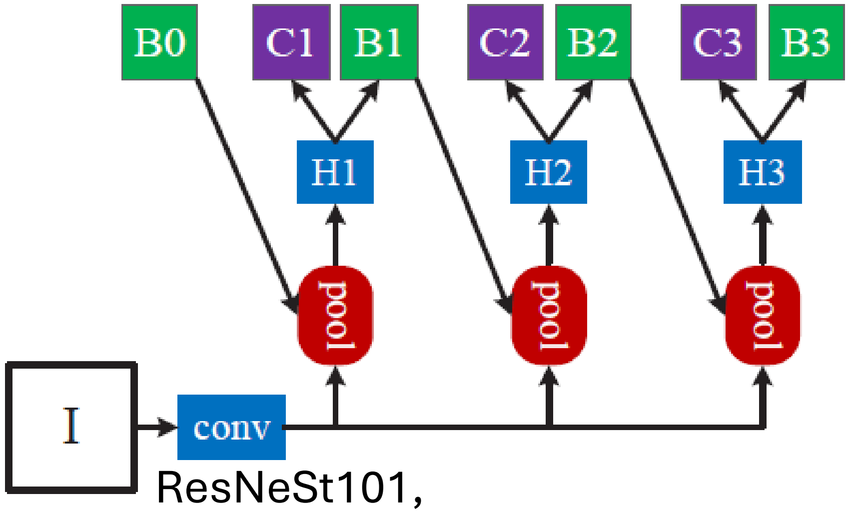
New Third-Party Images



New Third-Party Cells

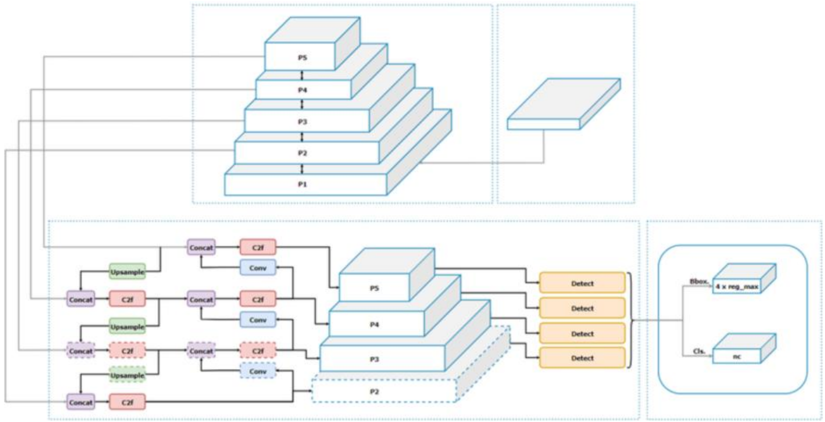


Change to Detection with Smaller Models



ResNeSt101,
SwinTransformer
(CBNetV2 backbone)

16x smaller



http://dx.doi.org/10.1109/ACCESS.2023.3293056

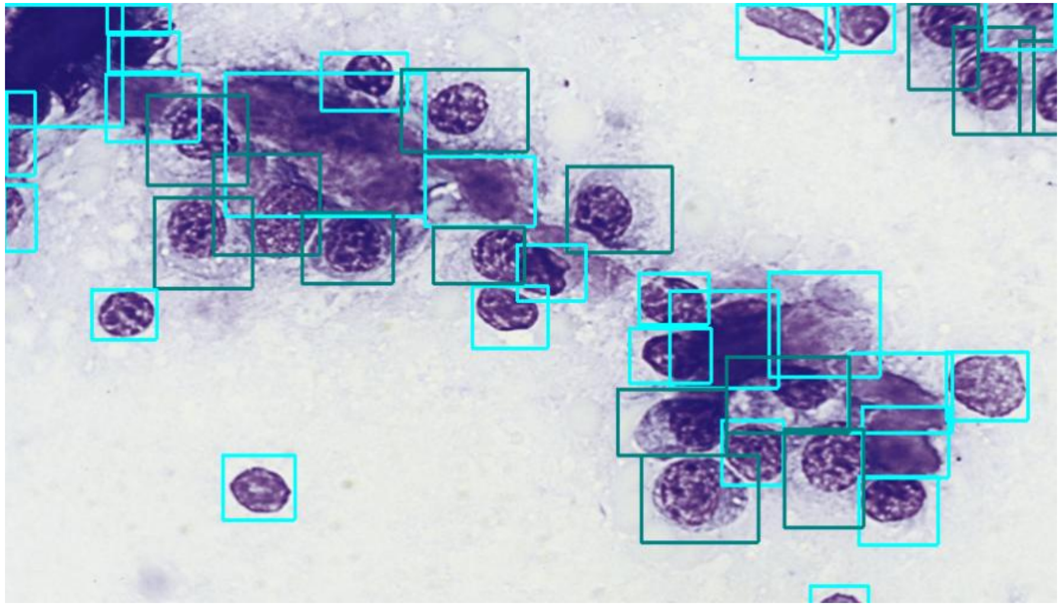
Cascade Mask R-CNN
416M parameters
~1.7GB

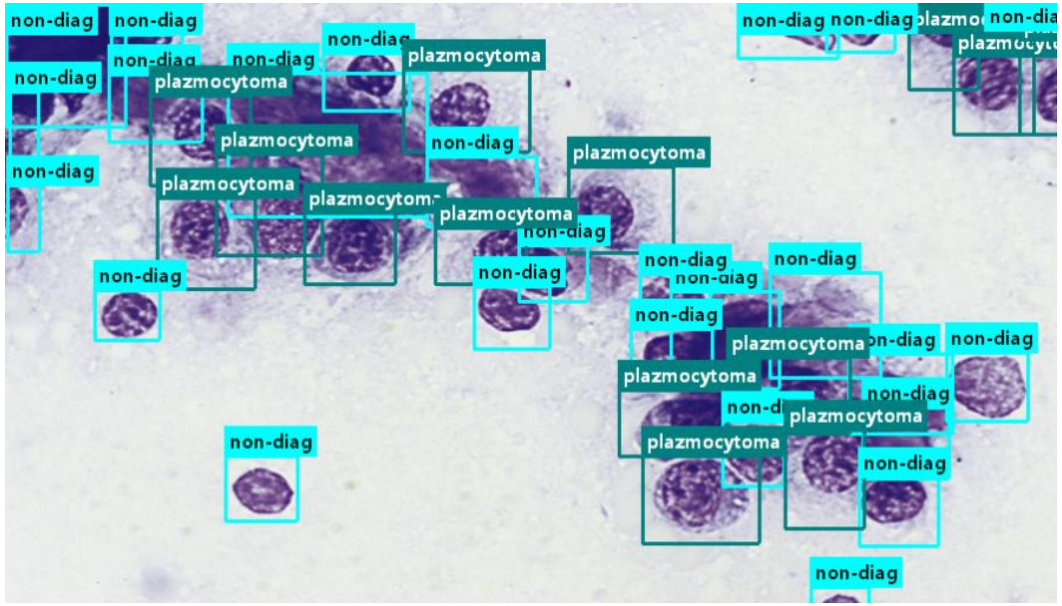
Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 6154-6162, doi: 10.1109/CVPR.2018.00644.

YOLOv8
26M parameters
104MB

Jocher, G., Qiu, J., & Chaurasia, A. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

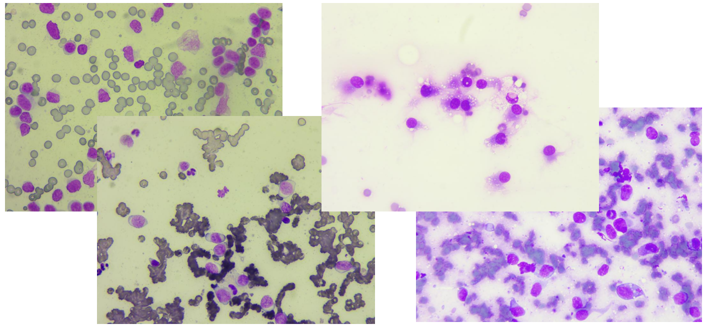
Labelling Cells for Detection





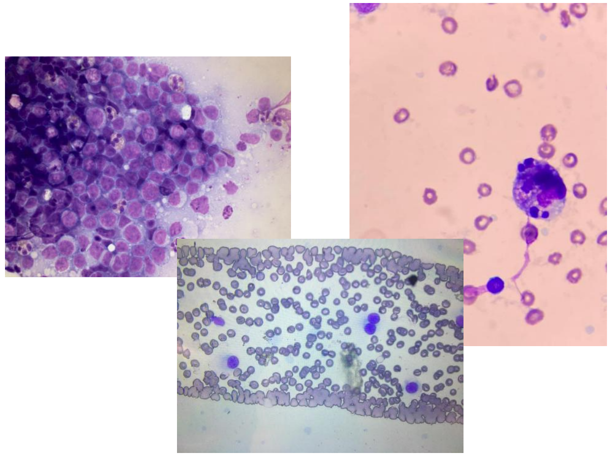
- Cells were detected by a previously trained model
- Cancer cell type known from the source

Improvement (mAP)



0.78 → **0.88**

Internal Data



0.26 → **0.54**

Third-Party Data

Conclusions & Future Research

- Diverse and high-quality data is crucial
- Biases in the internal data need to be further explored
- Bigger innovative models can wait until the dataset issues are resolved

Thank you for your attention!

Acknowledgements. The numerical experiment was possible through computing allocation on the Athena system at ACC Cyfronet AGH under the grants plglaoisi24, plgdyplomanci6.