

Preserving Marker Specificity with Lightweight Channel-Independent Representation Learning

Simon Gutwein^{1,2,4}

Arthur Longuefosse³

Jun Seita³

Sabine Taschner-Mandl¹

Roxane Licandro^{2,4,5}

SIMON.GUTWEIN@CCRI.AT

ARTHUR.LONGUEFOSSE@RIKEN.JP

JUN.SEITA@RIKEN.JP

SABINE.TASCHNER@CCRI.AT

ROXANE.LICANDRO@MEDUNIWIEN.AC.AT

¹ *St. Anna Children's Cancer Research Institute, Vienna, Austria*

² *Medical University of Vienna, Biomedical Imaging and Image-guided Therapy, Computational Imaging Research, ELIA Group, Vienna, Austria*

³ *RIKEN Center for Integrative Medical Sciences, Medical Data Deep Learning Team, Tokyo, Japan*

⁴ *Medical University of Vienna, Comprehensive Center for AI in Medicine, Vienna, Austria*

⁵ *Medical University of Vienna, Christian Doppler Lab for Mathematical Modelling and Simulation of Next-Generation Medical Ultrasound Devices, Vienna, Austria*

Abstract

Multiplex tissue imaging measures dozens of protein markers per cell, yet most deep learning models still apply early channel fusion, a design choice that erroneously imposes strong shared structure across weakly correlated channels. We present a systematic study of channel fusion as an architectural design choice for self-supervised representation learning. Using a Hodgkin lymphoma CODEX dataset with 145,000 cells and 49 markers, we demonstrate that preserving channel isolation yields substantially stronger representations than early fusion. Notably, a shallow channel-isolated architecture with only 5.5K parameters matches state-of-the-art models. We further demonstrate that channel-isolated embeddings enable interpretable phenotyping via marker attribution. Blind expert validation confirms that the proposed method produces annotations that respect biological constraints better than conventional pipelines (82.1% vs. 54.1% agreement). Our results show that delayed channel fusion can substitute for model scale in multiplex imaging, yielding interpretable representations that align more closely with biological ground truth and enable reliable expert-validated phenotyping. Code is available at <https://github.com/SimonBon/CIM-S>.

Keywords: Multiplex Imaging, Representation Learning, Cell Phenotyping

1. Introduction

Biological Multiplex Imaging (MI) technologies quantify multiple protein markers at sub-cellular resolution, providing high-dimensional data on single-cell states and tissue organization (Goltsev et al., 2018; Giesen et al., 2014). Widely used analysis pipelines rely on cell segmentation (Windhager et al., 2023) followed by aggregation of per-cell intensities. This approach propagates segmentation errors into downstream analyses (Bruhns et al., 2025) and restricts representation learning to predefined features. Convolutional Neural Networks (CNNs) offer an alternative by learning directly from raw images; however, standard architectures like ResNet (He et al., 2016) implement early-fusion, linearly combining all input channels in the first layer (Wang et al., 2024).

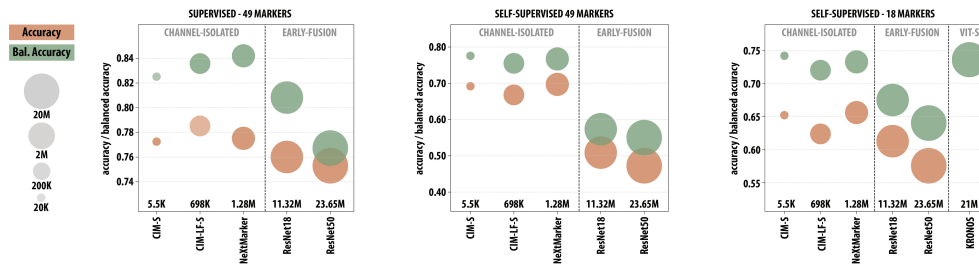


Figure 1: Representation-quality benchmark using self-supervised learning. Bubble size indicates parameter count. **Left:** Evaluation on all 49 markers. **Right:** Evaluation on a reduced 18-marker panel, including comparison with the KRONOS foundation model.

While early fusion is suited for highly correlated RGB images, MI channels represent independent molecular markers. Premature fusion can induce modality collapse (Chaudhuri et al., 2025), where weaker predictive signals are masked and essential markers cease to contribute to learning. We hypothesize that preserving marker independence in early layers is a critical inductive bias for MI. To test this, we systematically evaluate channel-isolated architectures, such as NeXtMarker (Gutwein et al., 2025), against early-fusion baselines and the KRONOS foundation model (Shaban et al., 2025) under self-supervised learning (SSL). We then demonstrate an interpretable, segmentation-free phenotyping pipeline using Layer-wise Relevance Propagation (LRP) (Bach et al., 2015) to automatically extract accurate cell labels.

2. Methods and Experiments

We utilize a publicly available CODEX dataset of classical Hodgkin lymphoma (cHL) (Shaban et al., 2024), containing approximately 145,000 cells across 49 protein markers. The baseline dataset includes annotations generated via DeepCell segmentation (Greenwald et al., 2022) and expert curation.

Proposed Architectures and SSL Benchmark: To isolate the effect of fusion strategy, we propose a lightweight Channel-Isolated Model (**CIM-S**). CIM-S (≈ 5.5 K params) uses grouped convolutions to preserve channel separation; **CIM-LF-S** adds late fusion (LF). We benchmark against ResNets (**ResNet18/50**), which serve as early-fusion baselines. Models were pretrained using the contrastive framework SimCLR (Chen et al., 2020). Subsequently, the encoders were frozen and evaluated via linear probing for cell classification.

Linear Probing Results: Early-fusion CNNs struggle to retain marker-specific information (Figure 1), with balanced accuracies of 55–58%. CIM-S achieves 77.6%, slightly outperforming NeXtMarker (76.7%) and the late-fusion variant CIM-LF-S (75.5%). When evaluated on a reduced 18-marker panel, the compact 5.5K-parameter CIM-S slightly outperforms the 21M-parameter KRONOS foundation model (74.2% vs. 73.6%). This demonstrates that architectural inductive bias can effectively substitute for massive model scale.

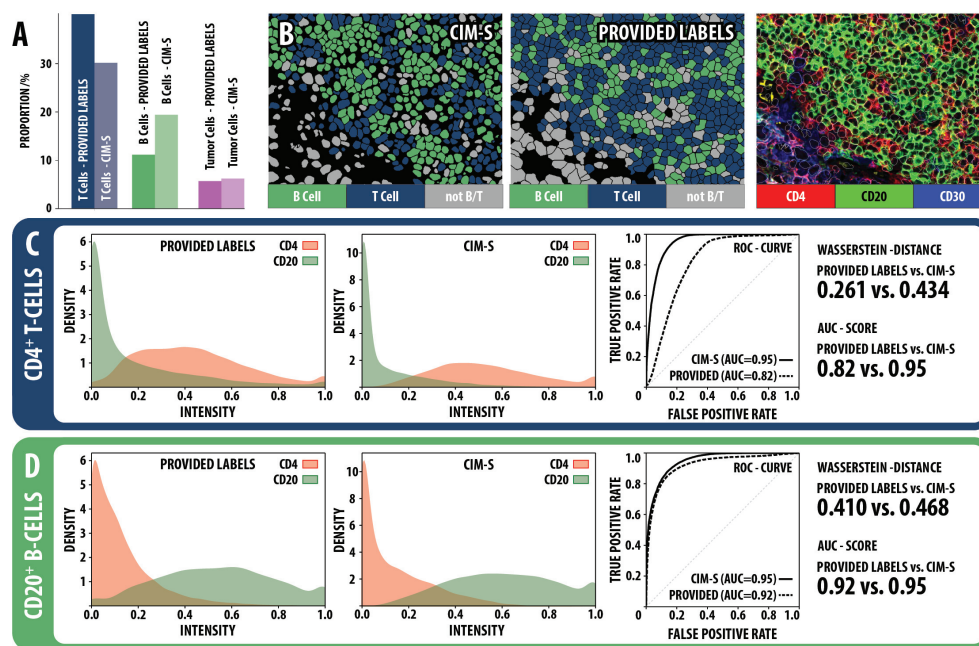


Figure 2: Quantitative comparison between CIM-S labels and provided baseline labels. **A**: Inferred cell-type proportions. **C/D**: CD4/CD20 separability within T-cell and B-cell populations, quantified by Wasserstein distance and ROC/AUC.

Label-Free Phenotyping: Frozen CIM-S representations enable segmentation-free phenotyping via LRP: spatial relevance maps are aggregated to channel-wise scores and multiplied with a marker-phenotype matrix (e.g., CD4→T-cells, CD20→B-cells).

Comparing our LRP-derived labels to the provided dataset labels revealed a strong shift: CIM-S identified 72.7% more B cells and 24.8% fewer T cells. To objectively evaluate this, we analyzed CD4 (T-cell) and CD20 (B-cell) signals, which are biologically mutual exclusive. CIM-S+LRP assignments achieved substantially higher CD4/CD20 separability within the T-cell population (WD, AUC in Figure 2) than the baseline provided cell type labels (Fig. 2). Blind evaluation with an external immunology expert on ~ 100 conflicting cells (no access to either label set) confirmed CIM-S superiority: 82.1% agreement vs. 54.1% for baseline labels, indicating systematic errors in conventional segmentation-clustering pipelines.

3. Conclusion

Our systematic evaluation demonstrates that preserving channel isolation in early layers prevents modality collapse in MI, allowing a 5.5K-parameter model to match a 21M-parameter foundation model. Furthermore, coupling channel-isolated embeddings with LRP enables interpretable, segmentation-free phenotyping. Blind expert validation confirmed that our pipeline successfully corrects systematic artifacts present in established provided labels, overcoming the "gold standard paradox" in digital pathology.

Acknowledgements

This research was funded by the Austrian Science Fund (FWF): the Emerging Fields grant DART2OS – Devising Advanced TCR-T cells to eradicate OsteoSarcoma (DOI: 10.55776/EFP45) and the Stand-Alone Project MAPMET – Mapping metastatic cancer by multi-modal imaging (DOI: 10.55776/P35841), both led by Sabine Taschner-Mandl (CCRI).

References

- S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R. Müller, and W. Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, 10(7), 7 2015. ISSN 19326203. doi: 10.1371/journal.pone.0130140.
- M. Bruhns, J. T. Schleicher, M. Wirth, et al. Effects of segmentation errors on downstream-analysis in highly-multiplexed tissue imaging. *PLoS Computational Biology*, 21(9): e1013350, 9 2025. ISSN 15537358. doi: 10.1371/journal.pcbi.1013350.
- A. Chaudhuri, A. Dutta, T. Bui, et al. A closer look at multimodal representation collapse. 8 2025. URL <http://arxiv.org/abs/2505.22483>.
- T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. 2 2020. URL <http://arxiv.org/abs/2002.05709>.
- C. Giesen, H. A. O. Wang, D. Schapiro, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods*, 11(4):417–422, 2014.
- Y. Goltsev, N. Samusik, J. Kennedy-Darling, et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, 174(4):968–981, 2018.
- N. F. Greenwald, G. Miller, E. Moen, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, 40(4):555–565, 4 2022. ISSN 15461696. doi: 10.1038/s41587-021-01094-0.
- S. Gutwein, D. Lazic, T. Walter, S. Taschner-Mandl, and R. Licandro. NeXtMarker: Contrastive learning for marker-level interpretability in single-cell multiplex imaging. In *MICCAI 2025 Workshop COMPAYL*. OpenReview, 7 2025. URL <https://openreview.net/forum?id=JYCUsFyMnj>.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- M. Shaban, Y. Bai, H. Qiu, et al. MAPS: pathologist-level cell type annotation from tissue images through machine learning. *Nature Communications*, 15(1), 12 2024. ISSN 20411723. doi: 10.1038/s41467-023-44188-w.
- M. Shaban, Y. Chang, H. Qiu, et al. A foundation model for spatial proteomics. 6 2025. URL <http://arxiv.org/abs/2506.03373>.

- X. Wang, R. Dilip, A. R. Iqbal, et al. Generalized cell phenotyping for spatial proteomics with language-informed vision models, 11 2024. URL <http://biorxiv.org/lookup/doi/10.1101/2024.11.02.621624>.
- J. Windhager, V. R. T. Zanotelli, D. Schulz, et al. An end-to-end workflow for multiplexed image processing and analysis. *Nature Protocols*, 18(11):3565–3613, 11 2023. ISSN 17502799. doi: 10.1038/s41596-023-00881-0.