# Continuous Magnification Training Improves Embedding Quality in Histopathological Self-Supervised Learning

**Alexander Möllers** A.MOELLERS@TU-BERLIN.DE
*TU Berlin, BIFOLD*, Aignostics - Berlin, Germany*

**Timo Milbich**
*Aignostics - Berlin, Germany*

**Maximilian Alber**
*Aignostics, Institute of Pathology, Charité - Universitätsmedizin Berlin, Germany*

**Lukas Ruff**
*Aignostics - Berlin, Germany*

## Abstract

Current histopathological foundation models are trained on discrete standard microscope magnifications (0.25, 0.5, 1.0, 2.0 microns per pixel). We use the unsupervised RankMe metric to show that this can affect embedding space quality at magnifications outside their training distribution, with rank scores dropping at intermediate scales. We introduce continuous magnification training, where patches are sampled from a continuous distribution during training, and show that this eliminates the irregularities in the embedding space.

**Keywords:** Histopathological foundation models, Self-supervised learning, Digital pathology, Continuous magnification training, Magnification Robustness

**Data and Code Availability** The models were trained on a large corpus of WSIs from TCGA and Charité - Universitätsmedizin Berlin that is in parts proprietary. We do not make code available in this version of the manuscript.

**Institutional Review Board (IRB)** Our research does not require IRB approval.

## 1. Introduction

In clinical practice, pathologists jump between continuous magnifications when examining tissue to incorporate cellular details and architectural patterns into their diagnostic assessments (Andrew and Anant, 2016). As this workflow has become increasingly digitized, self-supervised foundation models have been developed that power many analytical and diagnostic tools and lead to reliable performance across staining protocols and institutions (Chen et al., 2024; Zimmermann et al., 2024b; Alber et al., 2025). However, unlike pathologists, these models are not trained on the continuous magnification spectrum, but on image patches from one or more of the standard scanner magnifications. These are 0.25, 0.5, 1, and 2 microns per pixel (mpp).

This coarse discretization raises a critical question: *Does training on fixed magnifications create blind spots in the representation space?* Since existing benchmarks only evaluate at these same discrete scales, any degradation at intermediate magnifications would go undetected. To investigate this, we propose to use the unsupervised RankMe metric (Garrido et al., 2023), a strong indicator for downstream task performance (e.g., Ericsson et al., 2023; Jaume et al., 2024; Aben et al., 2024), to profile the quality of a model's embedding space across continuous magnifications.

Our controlled experiments reveal that current training practices create systematic degradation at intermediate magnifications, with embedding space quality dropping between standard training scales. This is particularly relevant as pathology AI is increasingly deployed in interactive diagnostic tools or multi-modal chatbots where pathologists are not bound to specific magnifications (Lu et al., 2024; Albastaki et al., 2025).

We solve this problem by extracting larger source patches and dynamically resizing them to a target

* Berlin Institute for the Foundations of Learning and Data

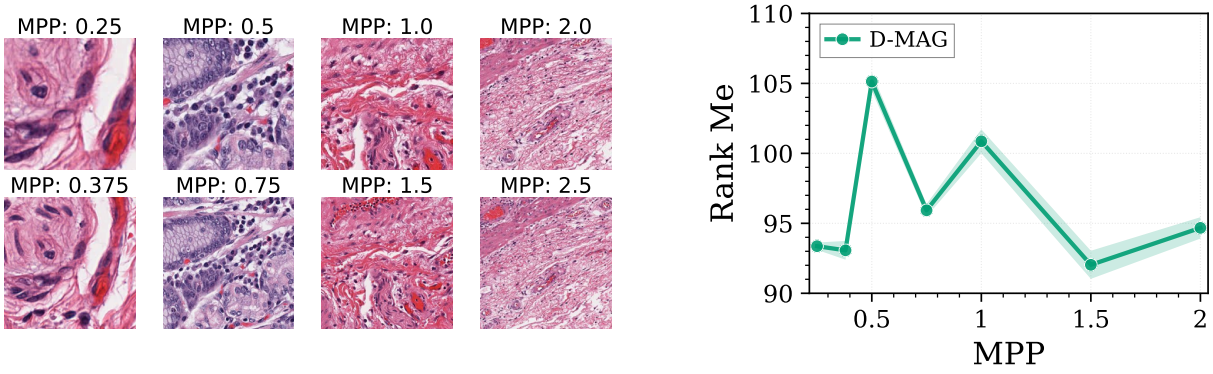


Figure 1: **Left** Top row displays tissue patches of microscopic images at the standard magnifications. Bottom row shows the same patches, but zoomed out slightly to intermediate magnifications. Clinically reliable FMs should be robust to slight changes in magnification at deployment time. **Right** Average representation quality of three multi-scale model trained with uniform sampling on the discrete standard magnifications (D-MAG). The representation quality of the models deteriorates at intermediate magnifications (0.375, 0.75, 1.5 mpp).

magnification drawn from a continuous distribution, effectively interpolating between standard scales during training. Rather than naive uniform sampling, we model magnification sampling as a domain adaptation problem and derive a distribution that maximizes the worst-case representation quality across magnifications. We use the proposed RankMe profiling to show that this leads to a more uniform embedding space quality. We summarize the contributions as follows:

- **Systematic profiling of magnification blind spots**: We propose to use the unsupervised RankMe metric to profile performance across magnifications in histopathological foundation models. We apply it in controlled experiments and use it to demonstrate that current discrete training strategies are suboptimal and lead to dimensional collapse at intermediate scales.
- **A practical solution through continuous sampling**: We propose continuous magnification training and show that this alleviates the dimensional collapse at intermediate magnifications.
- **A principled framework for optimal sampling**: We frame magnification sampling as a domain adaptation problem and derive a sampling distribution that maximizes the worst-case representation quality across magnifications.

## 2. The RankMe Metric for Profiling Representation Quality

We hypothesize that representation quality of a model varies across the magnification spectrum and deteriorates between training scales. To investigate this, we propose to use the RankMe metric (Garrido et al., 2023) as a magnification-agnostic profiling tool. RankMe quantifies the effective rank of a model's embedding space. Higher values indicate representations that span a larger subspace and encode richer information while lower values suggest dimensional collapse. We choose RankMe over alternative metrics as (1) it provides label-free assessment of representation quality at any magnification and (2) it has been validated to correlate strongly with downstream task performance across multiple digital pathology and domain adaptation benchmarks (Ericsson et al., 2023; Jaume et al., 2024; Aben et al., 2024). By computing RankMe for embeddings at standard (0.25, 0.5, 1.0, 2.0 mpp) and intermediate (0.375, 0.75, 1.5 mpp) magnifications, we can profile how representation quality varies across scales. A definition of RankMe is in Appendix B.

### 2.1. Embedding Quality Decreases in Blindspots

We now apply our RankMe-based profiling approach to investigate how performance varies across the con-

tinuous magnification spectrum. To do this, we conduct controlled experiments using the DINOv2 framework and train small vision transformers using two strategies: (1) single-scale models exposed to only one magnification, and (2) discrete multi-scale models (D-MAG) following current best practices with uniform sampling across standard magnifications (0.25, 0.5, 1.0, 2.0 mpp). We use standard training practices and list details in Appendix C.

For evaluation, we extract 10,000 source patches (392×392 pixels) at each standard magnification from held-out slides from TCGA. Through crop-and-resize operations, we then generate additional test patches at intermediate magnifications (0.375, 0.75, 1.5 mpp). By plotting the mpp of the patches on the x-axis and the rank of the corresponding embeddings on the y-axis we obtain a performance profile for each model.

Analyzing the performance profiles, we see that single-scale models exhibit sharp performance degradation when evaluated on magnifications distant from their training scale, with rank scores dropping by up to 40% at the extremes (Figure 2). Multi-scale models exhibit a more robust profile overall but reveal a distinctive "sawtooth" pattern (Figure 1). While these models avoid extreme degradation at any tested magnification, they show notable dips at scales absent from their training data. This pattern indicates that current discrete sampling strategies create blind spots in the representation space, even in models designed to be magnification-robust.

## 3. Implementing and Optimizing Continuous Magnification Sampling

Our experiments in the previous section demonstrated a systematic degradation of representation quality at magnifications that are not included in the training data. To alleviate this, we propose continuous magnification training through dynamic patch generation. By extracting larger source patches from existing WSIs and applying controlled crop-and-resize operations, we can synthesize training patches at arbitrary target magnifications. Formally, to create a patch at target magnification $t$ from a source patch at magnification $s$, we crop a region of size:

$$cs_{source} = cs_{target} \times \frac{t}{s}$$

and resize it to the desired patch size $cs_{target}$. During training, for each incoming patch at a standard magnification, we sample the target magnification from a continuous distribution $p$.

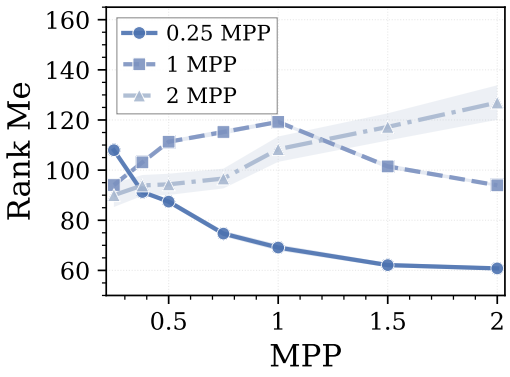


Figure 2: Representation quality across magnifications for single-scale models. Each line represents the average over 3 seeds and the legend indicates the magnification the models were trained on.

### 3.1. Magnification Sampling as a Domain Adaptation Problem

Rather than relying on heuristics to pick a sampling distribution $p$, we model how different choices affect downstream performance and optimize the resulting expression. To do this, we formalize the problem setting as a multi-source domain adaptation problem. Each magnification represents a source domain, and our target is the representation quality in a predefined continuous magnification range (e.g. 0.25-2.0 mpp). We then make two assumptions:

- **Domain Transfer Assumption:** We assume that representations learned at one magnification transfer to nearby magnifications with decreasing effectiveness as the magnification distance increases. We model this via a similarity kernel K(x,y), where x, y are the mpp values of the patches.

- **Coverage-based Proxy:** We expect the representation quality $I(y)$ at a specific magnification $y$ to improve with the number of similar training samples that the model has seen during training.

We can then model the expected representation quality at magnification $y$ as proportional to the ac-

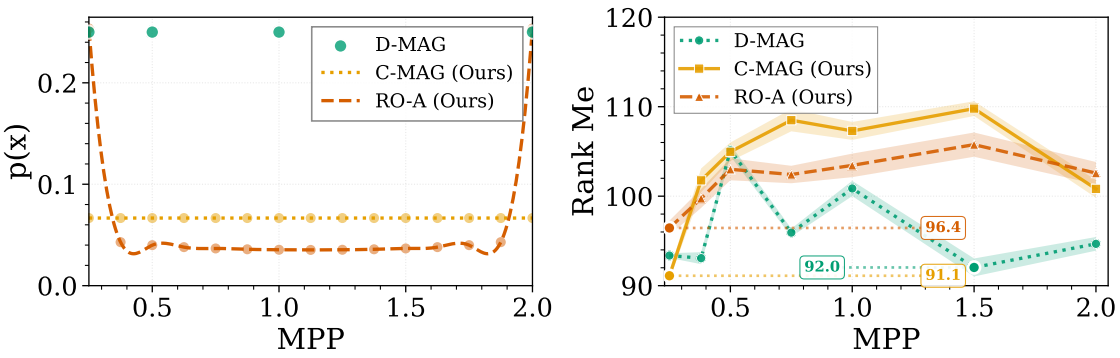


Figure 3: **Left**: Comparison of magnification sampling distributions. The min-max approach based on the absolute distance kernel (RO-A) assign higher probability mass to boundary magnifications compared to continuous uniform sampling (C-MAG). **Right**: Representation quality achieved with the different sampling methods. Both continuous sampling methods (C-MAG, RO-A) do not deteriorate at intermediate scales. RO-A achieves the highest worst-case representation quality and improves it over C-MAG.

cumulated training signal from all source magnifications:

$$I(y) = \int_{0.25}^{2} p(x)K(x,y)\,dx \tag{1}$$

We empirically validate the modeling assumptions made in this section by investigating the embedding space of trained models and showing how similar magnifications are close while more distant ones further apart (Appendix D).

### 3.2. Maximizing Worst-Case Performance for Robust Medical Foundation Models

To show how our framework can be used to develop principled sampling strategies we pick the absolute distance kernel $K(x,y) = \frac{1}{1+|x-y|}$ and formulate the sampling distribution selection as a max-min optimization problem:

$$p^* = \arg\max_{p} \min_{y\in[0.25,2]} I(y) \tag{2}$$

where we find $p^*$ that leads to the best worst-case representation quality. By framing the optimization problem in this way, we obtain a reliable clinical model that avoids failure modes at the boundaries and has a robust worst-case performance. To sample from the distribution, we discretize the problem and solve it using standard optimization techniques.

## 4. Experiments & Results

We train small vision transformers on 200,000 whole-slide images (WSIs) from The Cancer Genome Atlas (TCGA) and Charité - Universitätsmedizin Berlin (Details in Appendix C) . We compare three sampling strategies (1) D-MAG: discrete uniform sampling at standard magnifications (0.25, 0.5, 1.0, 2.0 mpp), (2) C-MAG: continuous uniform sampling across [0.25, 2.0] mpp, and (3) RO-A: robust optimization with absolute distance kernel as described in sec 3.2. We visualize the results in Figure 3.

Both our methods, C-MAG and RO-A, increase the representation space quality and eliminate the systematic degradation caused by standard discrete sampling. Interestingly, the RO-A optimized distributions oversamples the boundary magnifications. We can understand this intuitively as points closer to the magnification boundaries only have "close" samples in one direction while interior points profit from similar samples in both directions. Thus, naive continuous uniform sampling does in fact not create uniformly good representations, but leads to systematic weaknesses at the boundaries. RO-A alleviates this.

## 5. Discussion & Future Work

Our results show that the discrete sampling strategies of current training practices in histopathological self-supervised learning lead to a noticeably irregular embedding space. With this, we hope to inspire discus-

sions on the disconnect between the discrete common practice and the continuous reality. We look forward to works that investigate whether these phenomena persist at larger training scales and their potential implications for downstream clinical applications.

## Acknowledgments

The results in this paper are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

## References

Nanne Aben, Edwin D. de Jong, Ioannis Gatopoulos, Nicolas Känzig, Mikhail Karasikov, Axel Lagré, Roman Moser, Joost van Doorn, and Fei Tang. Towards Large-Scale Training of Pathology Foundation Models, March 2024. URL http://arxiv.org/abs/2404.15217. arXiv:2404.15217 [cs].

Shahad Albastaki, Anabia Sohail, Iyyakutti Iyappan Ganapathi, Basit Alawode, Asim Khan, Sajid Javed, Naoufel Werghi, Mohammed Bennamoun, and Arif Mahmood. Multi-resolution pathology-language pre-training model with text-guided visual representation. *arXiv preprint arXiv:2504.18856*, 2025.

Maximilian Alber, Stephan Tietz, Jonas Dippel, Timo Milbich, Timothée Lesort, Panos Korfiatis, Moritz Krügener, Beatriz Perez Cancer, Neelay Shah, Alexander Möllers, Philipp Seegerer, Alexandra Carpen-Amarie, Kai Standvoss, Gabriel Dernbach, Edwin de Jong, Simon Schallenberg, Andreas Kunft, Helmut Hoffer von Ankershoffen, Gavin Schaeferle, Patrick Duffy, Matt Redlon, Philipp Jurmeister, David Horst, Lukas Ruff, Klaus-Robert Müller, Frederick Klauschen, and Andrew Norgan. Atlas: A Novel Pathology Foundation Model by Mayo Clinic, Charité, and Aignostics, January 2025. URL http://arxiv.org/abs/2501.05409. arXiv:2501.05409 [cs].

Janowczyk Andrew and Madabhushi Anant. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathology Informatics*, 29(7), 2016.

Zak Buzzard, Konstantin Hemker, Nikola Simidjievski, and Mateja Jamnik. Paths: A hierarchical transformer for efficient whole slide image analysis. *arXiv preprint arXiv:2411.18225*, 2024.

Gianluca Carloni, Biagio Brattoli, Seongho Keum, Jongchan Park, Taebum Lee, Chang Ho Ahn, and Sergio Pereira. Pathology foundation models are scanner sensitive: Benchmark and mitigation with contrastive scangen loss. *arXiv preprint arXiv:2507.22092*, 2025.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021. URL https://arxiv.org/abs/2104.14294.

Binghao Chai, Jianan Chen, Paul Cool, Fatine Oumlil, Anna Tollitt, David F Steiner, Tapabrata Chakraborti, and Adrienne M Flanagan. Impact of variation in tissue staining and scanning devices on performance of pan-cancer ai models: a study of sarcoma and their mimics. *bioRxiv*, pages 2025–08, 2025.

Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood. Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16123–16134, New Orleans, LA, USA, June 2022. IEEE. ISBN 978-1-66546-946-3. doi: 10.1109/CVPR52688.2022.01567. URL https://ieeexplore.ieee.org/document/9880275/.

Richard J. Chen, Tong Ding, Ming Y. Lu, Drew F. K. Williamson, Guillaume Jaume, Andrew H. Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, Mane Williams, Lukas Oldenburg, Luca L. Weishaupt, Judy J. Wang, Anurag Vaidya, Long Phi Le, Georg Gerber, Sharifa Sahai, Walt Williams, and Faisal Mahmood. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, March 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-02857-3. URL https://www.nature.com/articles/s41591-024-02857-3. Publisher: Nature Publishing Group.

Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine Learning with*

*Applications*, 7:100198, March 2022. ISSN 2666-8270. doi: 10.1016/j.mlwa.2021.100198. URL https://www.sciencedirect.com/science/article/pii/S2666827021000992.

Ruining Deng, Can Cui, Lucas W. Remedios, Shunxing Bao, R. Michael Womick, Sophie Chiron, Jia Li, Joseph T. Roland, Ken S. Lau, Qi Liu, Keith T. Wilson, Yaohong Wang, Lori A. Coburn, Bennett A. Landman, and Yuankai Huo. Cross-scale multi-instance learning for pathological image diagnosis, 2024. URL https://arxiv.org/abs/2304.00216.

Jonas Dippel, Barbara Feulner, Tobias Winterhoff, Timo Milbich, Stephan Tietz, Simon Schallenberg, Gabriel Dernbach, Andreas Kunft, Simon Heinke, Marie-Lisa Eich, Julika Ribbat-Idel, Rosemarie Krupar, Philipp Anders, Niklas Prenißl, Philipp Jurmeister, David Horst, Lukas Ruff, Klaus-Robert Müller, Frederick Klauschen, and Maximilian Alber. RudolfV: A Foundation Model by Pathologists for Pathologists, June 2024. URL http://arxiv.org/abs/2401.04079. arXiv:2401.04079 [eess].

Linus Ericsson, Da Li, and Timothy Hospedales. Better practices for domain adaptation. In Aleksandra Faust, Roman Garnett, Colin White, Frank Hutter, and Jacob R. Gardner, editors, *Proceedings of the Second International Conference on Automated Machine Learning*, volume 224 of *Proceedings of Machine Learning Research*, pages 4/1–25. PMLR, 12–15 Nov 2023. URL https://proceedings.mlr.press/v224/ericsson23a.html.

Quentin Garrido, Randall Balestriero, Laurent Najman, and Yann LeCun. Rankme: assessing the downstream performance of pretrained self-supervised representations by their rank, 2023.

Guillaume Jaume, Paul Doucet, Andrew H. Song, Ming Y. Lu, Cristina Almagro-Pérez, Sophia J. Wagner, Anurag J. Vaidya, Richard J. Chen, Drew F. K. Williamson, Ahrong Kim, and Faisal Mahmood. Hest-1k: A dataset for spatial transcriptomics and histology image analysis, 2024. URL https://arxiv.org/abs/2406.16192.

Dinkar Juyal, Harshith Padigela, Chintan Shah, Daniel Shenker, Natalia Harguindeguy, Yi Liu, Blake Martin, Yibo Zhang, Michael Nercessian, Miles Markey, Isaac Finberg, Kelsey Luu, Daniel Borders, Syed Ashar Javed, Emma Krause, Raymond Biju, Aashish Sood, Allen Ma, Jackson Nyman, John Shamshoian, Guillaume Chhor, Darpan Sanghavi, Marc Thibault, Limin Yu, Fedaa Najdawi, Jennifer A. Hipp, Darren Fahy, Benjamin Glass, Eric Walk, John Abel, Harsha Pokkalla, Andrew H. Beck, and Sean Grullon. PLUTO: Pathology-Universal Transformer, May 2024. URL http://arxiv.org/abs/2405.07905. arXiv:2405.07905 [eess].

Mingu Kang, Heon Song, Seonwook Park, Donggeun Yoo, and Sérgio Pereira. Benchmarking Self-Supervised Learning on Diverse Pathology Datasets, April 2023. URL http://arxiv.org/abs/2212.04690. arXiv:2212.04690 [cs].

Mikhail Karasikov, Joost van Doorn, Nicolas Känzig, Melis Erdal Cesur, Hugo Mark Horlings, Robert Berke, Fei Tang, and Sebastian Otálora. Training state-of-the-art pathology foundation models with orders of magnitude less data, 2025. URL https://arxiv.org/abs/2504.05186.

Jonah Kömen, Edwin D de Jong, Julius Hense, Hannah Marienwald, Jonas Dippel, Philip Naumann, Eric Marcus, Lukas Ruff, Maximilian Alber, Jonas Teuwen, et al. Towards robust foundation models for digital pathology. *arXiv preprint arXiv:2507.17845*, 2025.

Ming Y. Lu, Bowen Chen, Drew F.K. Williamson, Richard J. Chen, Melissa Zhao, Aaron K. Chow, Kenji Ikemura, Ahrong Kim, Dimitra Pouli, Ankush Patel, Amr Soliman, Chengkuan Chen, Tong Ding, Judy J. Wang, Georg Gerber, Ivy Liang, Long Phi Le, Anil V. Parwani, Luca L. Weishaupt, and Faisal Mahmood. A multimodal generative ai copilot for human pathology, October 2024. ISSN 0028-0836.

Faisal Mahmood. A benchmarking crisis in biomedical machine learning. *Nature Medicine*, 31(4): 1060–1060, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu,

Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2024. URL https://arxiv.org/abs/2304.07193.

Maral Rasoolijaberi, Morteza Babaei, Abtin Riasatian, Sobhan Hemati, Parsa Ashrafi, Ricardo Gonzalez, and Hamid R. Tizhoosh. Multi-magnification image search in digital pathology. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4611–4622, 2022. doi: 10.1109/JBHI.2022.3181531.

Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, and Hervé Jégou. Spreading vectors for similarity search. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=SkGuG2R5tm.

Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016. doi: 10.1109/TBME.2015.2496264.

Anurag Vaidya, Andrew Zhang, Guillaume Jaume, Andrew H Song, Tong Ding, Sophia J Wagner, Ming Y Lu, Paul Doucet, Harry Robertson, Cristina Almagro-Perez, et al. Molecular-driven foundation model for oncologic pathology. *arXiv preprint arXiv:2501.16652*, 2025.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, Ellen Yang, Philippe Mathieu, Alexander van Eck, Donghun Lee, Julian Viret, Eric Robert, Yi Kan Wang, Jeremy D. Kunz, Matthew C. H. Lee, Jan H. Bernhard, Ran A. Godrich, Gerard Oakley, Ewan Millar, Matthew Hanna, Hannah Wen, Juan A. Retamero, William A. Moye, Razik Yousfi, Christopher Kanan, David S. Klimstra, Brandon Rothrock, Siqi Liu, and Thomas J. Fuchs. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*, 30(10):2924–2935, October 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-03141-0. URL https://www.nature.com/articles/s41591-024-03141-0. Publisher: Nature Publishing Group.

Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer. *arXiv preprint arXiv:2111.07832*, 2021.

Eric Zimmermann, Neil Tenenholtz, James Hall, George Shaikovski, Michal Zelechowski, Adam Casson, Fausto Milletari, Julian Viret, Eugene Vorontsov, Siqi Liu, and Kristen Severson. Adapting Self-Supervised Learning for Computational Pathology, May 2024a. URL http://arxiv.org/abs/2405.01688. arXiv:2405.01688 [cs].

Eric Zimmermann, Eugene Vorontsov, Julian Viret, Adam Casson, Michal Zelechowski, George Shaikovski, Neil Tenenholtz, James Hall, David Klimstra, Razik Yousfi, Thomas Fuchs, Nicolo Fusi, Siqi Liu, and Kristen Severson. Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology, November 2024b. URL http://arxiv.org/abs/2408.00738. arXiv:2408.00738 [cs].

## Appendix A. Related Work

**Vision foundation models in digital pathology** Microscopic visual data in digital pathology comes in the form of gigapixel whole slide images (WSIs) that show human tissue in fine-grained detail. Due to the immense size of these images, many existing vision foundation models in the field cut out small patches of tissue from the WSI for training (e.g., Chen et al., 2024; Dippel et al., 2024; Zimmermann et al., 2024b; Vorontsov et al., 2024; Alber et al., 2025). The resulting patch-level embeddings encode local information and several embeddings from different locations on the same slide can be combined to inform specific downstream tasks. Most of the state-of-the-art models for this patch-level encoding rely on the DINOv2 framework for training Oquab et al. (2024). Alternative approaches exist that directly optimize slide-level objectives, but the resulting embeddings do usually not lend themselves directly to the analysis of local features (e.g., Chen et al., 2022; Buzzard et al., 2024). Models have been probed on scanner or hospital robustness (Chai et al., 2025; Kömen et al., 2025; Carloni et al., 2025), but no work has investigated magnification robustness in detail.

**Multi-scale self-supervised training in digital pathology** Many current pathology foundation models include patches from different resolutions into their training. Often, uniform sampling is performed over the standard scanner magnifications and it is frequently observed that this increases performance across benchmarks (e.g., Ciga et al., 2022; Kang et al., 2023; Aben et al., 2024; Karasikov et al., 2025). Other approaches to multi-scale modeling include, Zimmermann et al. (2024a) who propose positional encodings that rely on the magnification difference of two patches. Furthermore, Juyal et al. (2024) add a masked autoencoder loss that reconstructs patch regions of varying sizes in order to obtain informative representations of biological features at different scales. In the vision-language domain Albastaki et al. (2025) investigate including different standard magnifications into training in order to improve the responses of language models on different scales. Importantly, the choice and the effect of the sampling distribution of the multi-scale data has not been thoroughly investigated.

**Multi-scale Benchmarks and Applications in Digital Pathology** Many downstream applications in digital pathology require analysis across multiple scales, such as multi-magnification image search (Rasoolijaberi et al., 2022) and cross-scale multiple instance learning for cancer classification (Deng et al., 2024). To evaluate current state-of-the-art foundation models for these applications, several multi-magnification benchmarks have been used in the literature. These include BreakHis (0.01, 0.05, 0.1, 0.25 mpp), TCGA Uniform (0.5, 1.0 mpp), and the proprietary PanMSK dataset (0.5, 1.0, 2.0 mpp) (Vaidya et al., 2025; Spanhol et al., 2016; Zimmermann et al., 2024b). However, no single analysis covers all standard scanner magnifications (2.0, 1.0, 0.5 , 0.25 mpp) and existing benchmarks only evaluate at standard discrete magnifications. Furthermore, many existing benchmarks are saturated and are hardly able to distinguish differences between models and evaluations rarely happen with standardized evaluation protocols (Mahmood, 2025).

## Appendix B. Rank Me

Given a set of N patches $\mathbf{X}_{\text{mpp}} = \{x_1, ..., x_N\}$ extracted at specific mpp, we obtain their embeddings $\mathbf{Z}_{\text{m,mpp}} = [z_1, ..., z_N]^T \in \mathbb{R}^{N \times K}$ by passing them through model m, where K is the embedding dimension. Then RankMe is defined as:

$$\text{RankMe}(\mathbf{Z}_{\text{m,mpp}}) = \exp\left(-\sum_{k=1}^{\min(N,K)} p_k \log p_k\right), \tag{3}$$

where

$$p_k = \frac{\sigma_k(\mathbf{Z}_{\text{m,mpp}})}{|\sigma(\mathbf{Z}_{\text{m,mpp}})|_1} + \epsilon, \tag{4}$$

where $\sigma_k(\mathbf{Z}_{m,\text{mpp}})$ denotes the $k$-th singular value of the embedding matrix $\mathbf{Z}_{m,\text{mpp}}$, $|\sigma(\mathbf{Z}_{m,\text{mpp}})|_1 = \sum_{i=1}^{\min(N,K)} \sigma_i(\mathbf{Z}_{m,\text{mpp}})$ is the sum of all singular values, and $\epsilon$ is a small constant for numerical stability.

## Appendix C. Training Details

For training, we adapt the DinoV2 framework Oquab et al. (2024) and train a student network $f_{\theta_s}(\mathbf{X})$ and a teacher network $f_{\theta_t}(\mathbf{X})$ for 60.000 iterations with a batch size of 320, with $\mathbf{X} \in R^{224x224x3}$. For training, we extract two global crops $\mathbf{X}_g \in R^{224x224x3}$ and eight local crops $\mathbf{X}_l \in R^{98x98x3}$ from a larger source patch $\mathbf{X}_s \in R^{256x256x3}$. Furthermore, we create masked versions of the global crops $\mathbf{X}_{g_m}$. The objective is then a combination of the Dino loss $\mathcal{L}_{Dino}$

(Caron et al., 2021), the Ibot loss $\mathcal{L}_{ibot}$ (Zhou et al., 2021) and the Koleo loss $\mathcal{L}_{Koleo}$ (Sablayrolles et al., 2019) that together encourage image-level distillation between the local and global crops and patch-level reconstruction between the masked and unmasked ones Oquab et al. (2024). As architecture we use small vision transformers (VitS) and use the teacher network to create the final embeddings. For $N$ patches this results in an embedding matrix $\mathbf{Z} \in R^{Nx384}$. We use the standard DinoV2 hyperparameters and train with a base learning rate of 0.001 and a weight decay cosine schedule from 0.04 to 0.2. The scale ranges for the resizing of global and local crops are set to [1, 0.35] and [0.35, 0.05] respectively.

As training data we use patch datasets at the four standard magnifications (0.25, 0.5, 1.0, and 2.0 mpp) from 200,000 whole-slide images (WSIs) from The Cancer Genome Atlas (TCGA) and Charité - Universitätsmedizin Berlin. This yields 774M, 280M, 69M, and 18M patches at the respective magnifications.

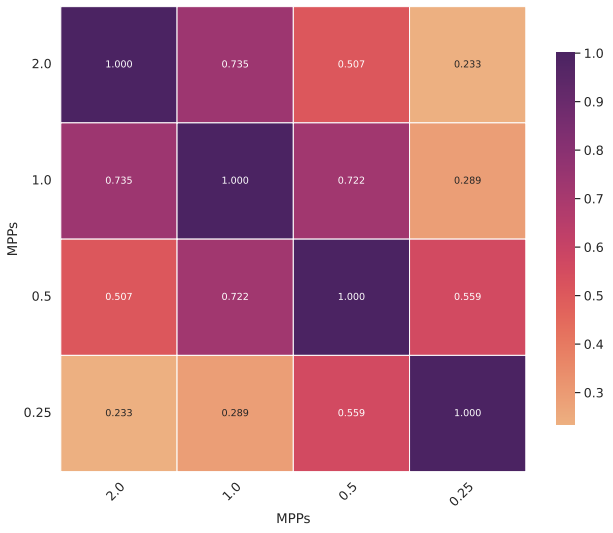


Figure 4: Cosine Similarities between embedding centroids for the standard magnifications in a multi-scale model. Larger magnification differences correspond to greater separation in the learned feature space.

## Appendix D. Embeddings Space Similarities

To show that different magnifications are more distant in a models embedding space, we compute the centroids for embeddings of patches from each of the standard magnification. From the heatmap in Figure 4 we see that distances between centroids increases with magnification distance.

## Appendix E. Comparison between different Kernels

In this section of the appendix we present how a different choice of Kernel affects the optimization problem in 3.2. We display the results in figure 5 and compare the following two Kernels:

**Absolute Distance Kernel** The absolute distance kernel $K_{\text{abs}}(x, y) = \frac{1}{1+|x-y|}$ encodes that the effect of sampling a patch at magnification $x$ on the representation quality at magnification $y$ decreases proportionally to the magnification distance $|x - y|$.

**Information-Based Kernel** The information-based kernel $K_{\text{area}}(x, y) = \left(\frac{\min(x,y)}{\max(x,y)}\right)^2$ models information transfer through field of view overlap. A 224×224 patch at magnification $x$ x mpp covers a tissue area of $(224 * x)^2$ square microns. When viewing this same tissue at magnification $y$, the kernel represents the fraction of overlapping field of view, squared to account for 2D area.

## Appendix F. Extension to larger models

To investigate qualitatively if the observed phenomena also extend to larger models, we train a Vit-L with batch size 960 for 62.500 iterations with discrete uniform sampling over the 4 standard magnifications. We display the results in figure 6 and observe a drop of embedding quality at intermediate magnifications similar to what we have seen in smaller models.

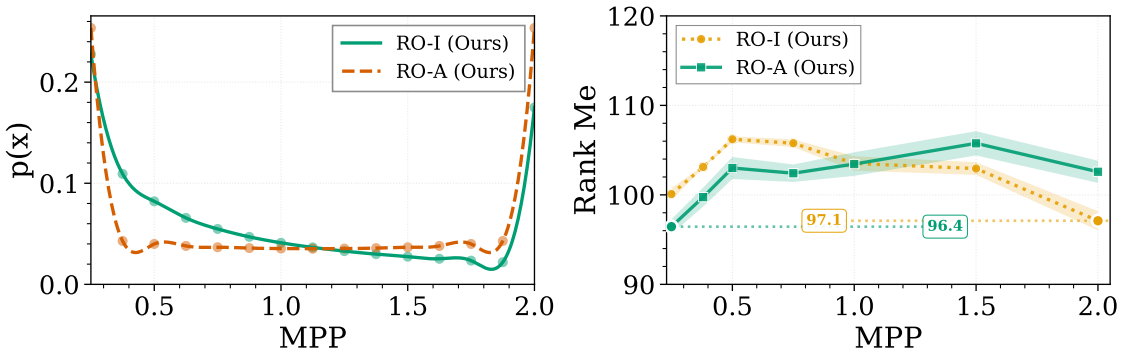


Figure 5: Optimized sampling distribution for two different kernels. RO-I is optimized with an Information-based Kernel and RO-A with an absolute distance kernel. Both kernels lead to an upsampling of the data points at the boundaries.

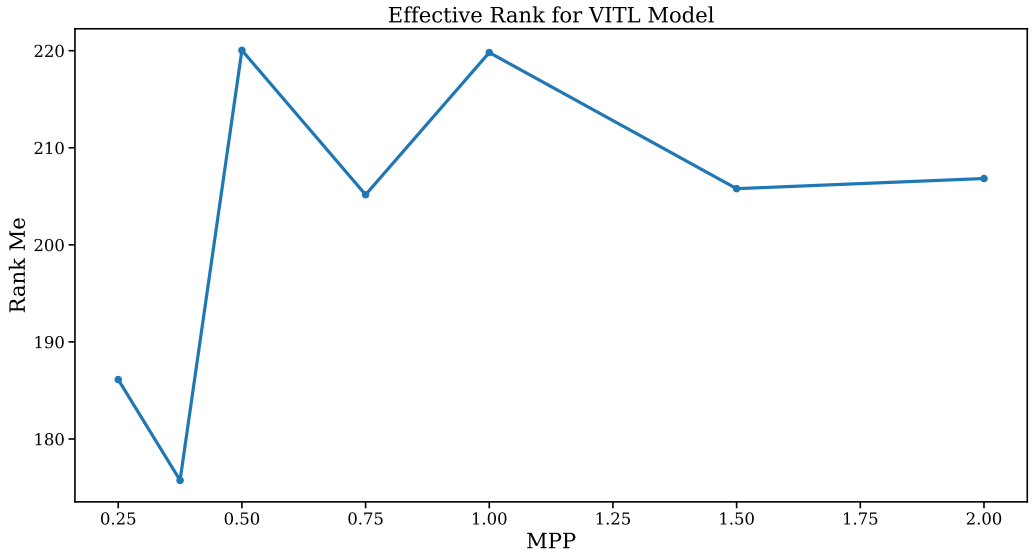


Figure 6: The Rank Me Metric for a Vit-L model trained with discrete uniform sampling. We observe a similar drop in embedding quality as for the smaller models.