Are the Latent Representations of Foundation Models for Pathology Invariant to Rotation?

Matouš Elphick[1,2,3] MATOUS.ELPHICK@CRICK.AC.UK
Guang Yang*[3] G.YANG@IMPERIAL.AC.UK
Samra Turajlic*[1,2] SAMRA.TURAJLIC@CRICK.AC.UK
[1] *The Francis Crick Institute,* [2] *The Institute of Cancer Research,* [3] *Imperial College London*

Editors: Accepted for publication at MIDL 2025

Abstract

Self-supervised foundation models for digital pathology encode small patches from H&E whole slide images into latent representations used for downstream tasks. However, the invariance of these representations to patch rotation remains unexplored. This study investigates the rotational invariance of latent representations across twelve foundation models by quantifying the alignment between non-rotated and rotated patches using mutual k-nearest neighbours and cosine distance. Models that incorporated rotation augmentation during self-supervised training exhibited significantly greater invariance to rotations. We hypothesise that the absence of rotational inductive bias in the transformer architecture necessitates rotation augmentation during training to achieve learned invariance. Code [1].

Keywords: computational pathology, oncology, foundation models, latent representations

1. Introduction

Digital pathology has enabled the use of deep learning on H&E whole slide images (WSIs). Foundation models (FMs), trained through self-supervised learning on large WSI datasets, generate compact latent representations which can be used downstream for tasks such as tumour detection and grading. However, as these models are adopted in practice, concerns have arisen about their robustness. While most studies on FMs for pathology evaluate their accuracy and ability to generalise across datasets, few have investigated their resilience to geometric transformations such as rotation. Ideally, such transformations should not affect the diagnostic features encoded in the latent representations, and understanding the impact of rotation is crucial, as invariance could enhance model reliability and robustness. However, the lack of rotational inductive bias in the transformer architecture suggests that achieving invariance may require rotation augmentation during training. This paper investigates the extent to which the latent representations of FMs remain invariant to rotation by quantifying the alignment between representations of images with and without rotation.

2. Methods

2.1. Dataset

For this study, the FMs were benchmarked using the TCGA-KIRC dataset (Akin et al., 2016), which contains 363 H&E WSIs. WSIs are read at 20× magnification and converted

* Joint senior authorship
1. https://github.com/MatousE/rot-invariance-analysis

to RGB and HSV color spaces. Non-background regions in the HSV images are isolated through segmentation, noise reduction, and contour detection. The five largest contours are selected, and 256×256 pixel patches are extracted from their bounding boxes, ensuring at least 75% of pixels are foreground.

2.2. Models

In this study, we evaluated twelve FMs that have used self-supervised training on WSI datasets including: Conch (Lu et al., 2024), Hibou (base and large) (Nechaev et al., 2024), Kaiko (base and large) (ai et al., 2024), PathDino (Alfasly et al., 2024), Phikon (Filiot et al., 2023), Phikon 2 (Filiot et al., 2024), Prov-GigaPath (Xu et al., 2024), UNI (Chen et al., 2024), Virchow (Vorontsov et al., 2024) and Virchow 2 (Zimmermann et al., 2024).

2.3. Rotations

We applied rotations from 0° to 360° at 15° intervals to WSI patches and applied the FM to the patch at each rotation and extracted the final latent representation. The 'control' condition, with no rotation, serves as the baseline for comparison.

2.4. Metric

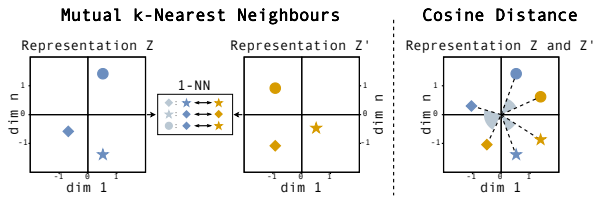

Figure 1: (Left) Representations are compared based on similarity of their k nearest neighbors, here k = 1. (Right) Distance is measured by the cosine of the angle between representations.

To quantify the alignment between control and rotated latent representations, we used two metrics: mutual k-nearest neighbours (m-kNN) (Fig. 1 Left), which evaluates the proximity of similar images in latent space, and cosine distance, which measures how consistently images are encoded in the same direction (Fig. 1 Right) (Klabunde et al., 2024; Huh et al., 2024). These metrics are justified as H&E WSI patches of a similar type should not only be close in latent space but also map in similar directions, regardless of rotation. Let $X = \{x_i \in \mathbb{R}^{H\times W\times 3}\}_{i=1}^{N}$ represent a set of image patches extracted from WSIs, where H and W are the patch dimensions, and 3 is the number of channels. A model f maps each patch to a latent representation, yielding $Z = \{z_i = f(x_i) \in \mathbb{R}^d\}_{i=1}^{N}$ where d is the representation dimension. Given a rotation $\mathcal{R}$ which is applied to the input patches, the resulting representations are $Z' = \{z'_i = f(\mathcal{R}(x_i)) \in \mathbb{R}^d\}_{i=1}^{N}$. Now let $\mathcal{N}_k(z_i)$ be the set of k-nearest neighbours of $z_i \in Z$ and $\mathcal{N}_k(z'_i)$ be the k-nearest neighbours of $z'_i \in Z'$, measured in the respective latent spaces with Euclidean distance. The m-kNN between Z and Z' is then defined:

$$m^k_{\text{NN}}(Z, Z') = \frac{1}{N}\sum_{i=1}^{N}\frac{|\mathcal{N}_k(z_i) \cap \mathcal{N}_k(z'_i)|}{k}. \tag{1}$$

The cosine distance between Z and Z' is:

$$\text{Cosine Distance}(Z, Z') = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{z_i \cdot z'_i}{\|z_i\|\|z'_i\|}\right). \tag{2}$$

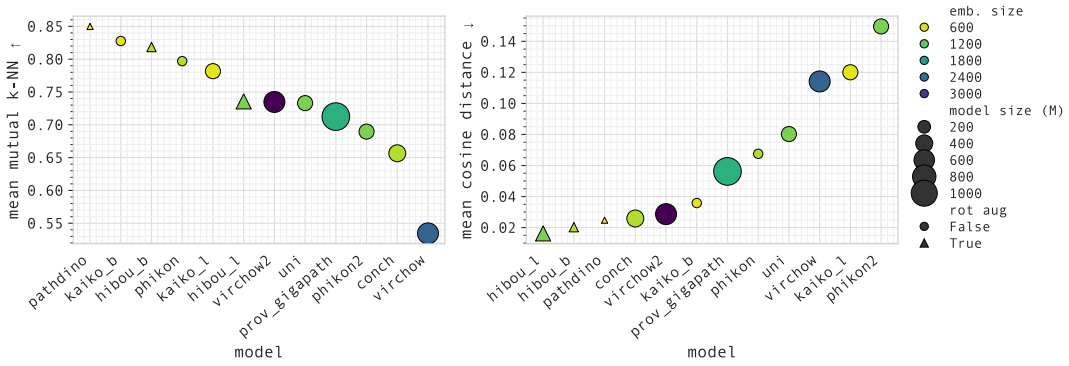

Figure 2: Graphs showing how model characteristics relate to representation invariance under rotation. (Left) Mean mutual k-nearest neighbour across rotations, with higher values indicating greater invariance. (Right) Mean cosine distance, where lower values indicate stronger alignment in latent space representations under rotation.

3. Results

In this study, FM invariance to H&E WSI patch rotation is evaluated using m-kNN and cosine distance. Latent representations were extracted for each patch at 15° increments across all models. m-kNN, with $k = 10$, and cosine distance were computed between non-rotated and rotated representations to quantify invariance. Fig. 2 shows the mean m-kNN and cosine distances across rotations. PathDino was the most invariant for m-kNN (0.85), while Virchow was the least (0.53). For cosine distance, Hibou-L showed the most invariance (0.016), whereas Phikon2 showed the least (0.145). Models were divided into two groups based on whether they used rotation augmentation during training. A t-test comparing these groups showed significant differences in both metrics: models trained with rotation augmentation achieved better scores for both cosine distance ($t = -8.88$, $p < 0.0001$) and m-kNN ($t = 6.91$, $p < 0.0001$). Fig. 3 presents a heatmap of m-kNN and cosine distances across models and rotations. Alignment was poorest at angles between cardinal rotations (45°, 135°, 225°, 315°), likely due to slight differences in the patch corners introduced by rotation.

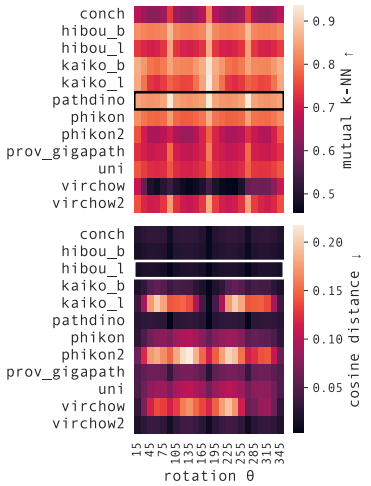

Figure 3: (Top) Mutual k-nearest neighbour across rotation angles, where higher values indicate greater invariance. (Bottom) Cosine distance across rotations, with lower values indicating greater latent representation alignment.

4. Conclusion

This study revealed that FMs for pathology exhibit varying degrees of rotational invariance, with models that use rotation augmentation during training being more invariant to rotation than those that do not. This result suggests that due to the transformers' lack of rotational inductive bias, rotation augmentation is necessary to achieve learned invariance. In future work, evaluation of rotation invariance could extend across a number of datasets and leverage several alignment metrics.

References

kaiko ai, 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].

Oguz Akin, Pierre Elnajjar, Matthew Heller, Rose Jarosz, Bradley J. Erickson, Shanah Kirk, Yueh Lee, Marston W. Linehan, Rabindra Gautam, Raghu Vikram, Kimberly M. Garcia, Charles Roche, Ermelinda Bonaccio, and Joe Filippini. The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma Collection (TCGA-KIRC), 2016. URL https://www.cancerimagingarchive.net/collection/tcga-kirc/.

Saghir Alfasly, Abubakr Shafique, Peyman Nejat, Jibran Khan, Areej Alsaafin, Ghazal Alabtah, and H. R. Tizhoosh. Rotation-Agnostic Image Representation Learning for Digital Pathology, March 2024. URL http://arxiv.org/abs/2311.08359. arXiv:2311.08359 [cs].

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 1078-8956, 1546-170X. doi: 10.1038/s41591-024-02857-3. URL https://www.nature.com/articles/s41591-024-02857-3.

Alexandre Filiot, Ridouane Ghermi, Antoine Olivier, Paul Jacob, Lucas Fidon, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling, July 2023. URL http://medrxiv.org/lookup/doi/10.1101/2023.07.21.23292757.

Alexandre Filiot, Paul Jacob, Alice Mac Kain, and Charlie Saillard. Phikon-v2, A large and public feature extractor for biomarker prediction, September 2024. URL http://arxiv.org/abs/2409.09173. arXiv:2409.09173 [eess].

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The Platonic Representation Hypothesis, July 2024. URL http://arxiv.org/abs/2405.07987. arXiv:2405.07987 [cs].

Max Klabunde, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Similarity of Neural Network Models: A Survey of Functional and Representational Measures, August 2024. URL http://arxiv.org/abs/2305.06329. arXiv:2305.06329 [cs].

Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, Anil V. Parwani, Andrew Zhang, and Faisal Mahmood. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, March 2024. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-024-02856-4. URL https://www.nature.com/articles/s41591-024-02856-4.

Dmitry Nechaev, Alexey Pchelnikov, and Ekaterina Ivanova. Hibou: A Family of Foundational Vision Transformers for Pathology, August 2024. URL http://arxiv.org/abs/2406.05074. arXiv:2406.05074 [eess].

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 1078-8956, 1546-170X. doi: 10.1038/s41591-024-03141-0. URL https://www.nature.com/articles/s41591-024-03141-0.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J. Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, and Hoifung Poon. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07441-w. URL https://www.nature.com/articles/s41586-024-07441-w.

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, August 2024. URL http://arxiv.org/abs/2408.00738. arXiv:2408.00738 [cs].