# Squeezing performance from pathology foundation models with chained hyperparameter searches

Joseph Cappadona[1], Ken Zeng[1], Carlos Fernandez-Granda[1,2], Jan Witowski[1], Yann LeCun[2,3], Krzysztof J. Geras[1,2]
[1]Ataraxis AI, [2]New York University, [3]Meta AI

ATARAXIS

NEURAL INFORMATION PROCESSING SYSTEMS

## Summary

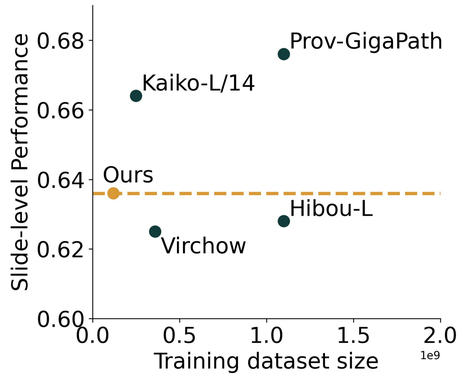
**SSL in digital pathology**: SSL is well-suited for digital pathology, leveraging an abundance of unlabeled slides and addressing the scarcity of labeled patient data.

**Study focus**: Explore the effect of hyperparameter tuning on the downstream performance of models trained with DINOv2, moving beyond the standardized hyperparameter schedules and augmentation policies used in most studies.

## Key Results

Chained hyperparameter searches can be an effective way to tune models in large hyperparameter search spaces.

**Figure 1:** We used this approach to train pathology foundation models with downstream performance on par with models trained on significantly more data.




## Experiment Setup

**Models:** We train two ViT sizes,

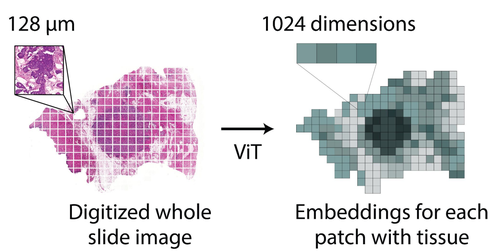
- ViT-Small (21M)
- ViT-Large (303M)

**SSL:** We tune hyperparameters related to

- learning rate
- weight decay
- teacher temperature
- EMA momentum
- augmentation policy

**Feature extraction:**

- Slide-level tasks are evaluated using mean-pooled patch embeddings, reduced to 50 dimensions with PCA.




## Main methodology: chaining hyperparameter searches

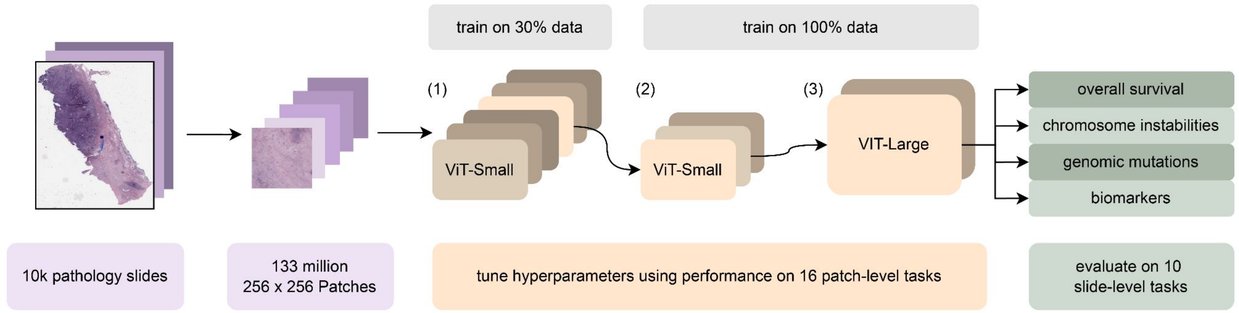



**Figure 2:** Successive hyperparameter searches are used to tune DINOv2 hyperparameters by optimizing performance on 16 patch-level tasks. In each successive step, either model size or dataset size is increased, and the hyperparameter search space is narrowed.

## Additional results

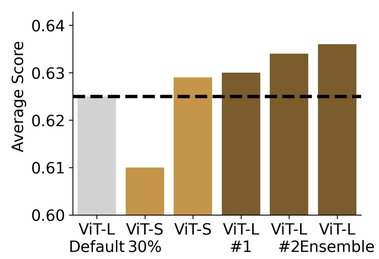



**Figure 3:** Hyperparameter tuning improves upon default DINOv2 hyperparameters.

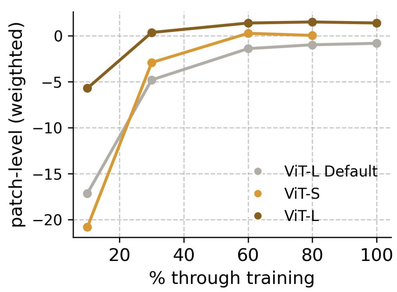



**Figure 4:** Patch-level task performance used for early stopping and hyperparameter selection.

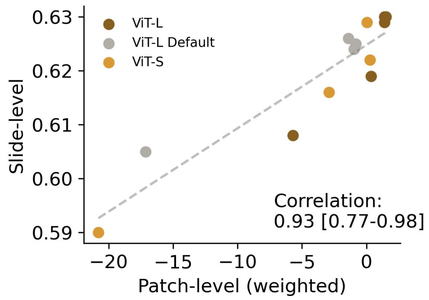



**Figure 5:** Patch-level performance is strongly correlated with slide-level performance.

## Discussion

- Our work highlights the importance of hyperparameter tuning in SSL.
- The best tuned ViT-S performed only marginally worse than the best tuned ViT-L.
- Patch-level task suites can be used in lieu of slide-level task suites to save computation.
- Hyperparameters transfer reasonably well across model and dataset size.

**Full paper:**