Paired-Acquisition Neural Factorization on the SCORPION Paired-Scanner Benchmark

Matthew Vaishnav

Independent Researcher

Kitchener-Waterloo, Canada

paired-acquisition-factorization-scorpion

Abstract

This package reports the SCORPION core study for Paired-Acquisition Neural Factorization. The benchmark contains 48 original human H&E slides, ten aligned tissue regions per slide, and one image from each of five scanners for every region, giving 480 biological regions and 2,400 paired-scanner images. A frozen neural factorization objective was trained on standardized frozen features to separate a biological branch from a compact acquisition branch. On DINOv2-Base features, Paired-Acquisition Neural Factorization reduced biological-branch scanner-probe accuracy from 0.7825 to 0.3989 relative to paired consistency, improved paired cosine agreement, and preserved near-saturated same-region retrieval. The same frozen objective was then transferred without backbone-specific tuning to Phikon and ImageNet ResNet50 features. Across those two unseen representation families, the slide-blocked scanner-probe difference was -0.4019 with 95% bootstrap interval $[-0.4145, -0.3895]$, favorable on all 48 slides. Scanner identity remained recoverable from the acquisition branch, while tissue retrieval stayed concentrated in the biological branch. The supported conclusion is bounded: this is representation-identifiability evidence on SCORPION, not clinical validation or proof of perfect disentanglement.

1 Purpose of This Package

This PDF is the public SCORPION child-package paper for **Paired-Acquisition Neural Factorization**. It contains the SCORPION method, protocol, results, artifacts, and claim boundary rather than the broader computational-pathology research-program paper. The repository contains the focused code, protocol notes, compact result artifacts, and paper source for the core paired-scanner study.

The scientific question is whether a biology-preserving paired acquisition structure can reduce scanner recoverability from a biological representation while retaining same-tissue identity. The study also asks whether the effect transfers beyond the DINOv2 development feature family to a pathology-native transformer and a conventional ImageNet residual network.

2 Dataset and Unit of Analysis

The study uses the public SCORPION benchmark [4]. Each original human H&E slide contributes ten aligned tissue regions. Each region is imaged by five scanners: AT2, GT450, DP200, P1000, and B300. This gives 48 original slides, 480 biological regions, and 2,400 images.

The original slide is the statistical unit. All regions and scanner views from one slide remain in the same fold. Optimization seeds are averaged within slide before bootstrap and sign-flip inference. Patches, scanner views, tissue regions, and repeated representation families are not treated as independent biological samples.

Three frozen feature families are evaluated:

- DINOv2-Base, the development representation family [3];
- Phikon, a pathology-native transformer feature family [1];
- ImageNet ResNet50, a conventional convolutional baseline [2].

3 Frozen Learning Problem

For a tissue region r and scanner s, let $x_{r,s} \in \mathbb{R}^D$ be the standardized frozen feature. The paired acquisition set

$$\mathcal{V}_r = \{x_{r,s_1}, \ldots, x_{r,s_m}\}$$

contains multiple scanner views of the same biological region. The biological branch should agree across members of $\mathcal{V}_r$, while the acquisition branch should retain scanner-associated variation. Under the simple additive model

$$x_{r,s} = Bz_{b,r} + Az_{a,s} + \varepsilon_{r,s},$$

a paired difference cancels the shared biological term:

$$x_{r,s'} - x_{r,s} = A(z_{a,s'} - z_{a,s}) + (\varepsilon_{r,s'} - \varepsilon_{r,s}).$$

The neural model does not require the observation map to be exactly linear; the equation states why paired observations provide information that ordinary scanner labels alone do not.

3.1 Architecture

For a standardized frozen feature $x \in \mathbb{R}^D$, Paired-Acquisition Neural Factorization uses two factor encoders, a joint decoder, and two scanner heads:

$$z_b = f_b(x) \in \mathbb{R}^{d_b}, \qquad d_b = 256, \tag{1}$$

$$z_a = f_a(x) \in \mathbb{R}^{d_a}, \qquad d_a = 64, \tag{2}$$

$$\widehat{x} = g([z_b; z_a]) \in \mathbb{R}^D, \tag{3}$$

$$\widehat{y}_b = q_b(\mathrm{GRL}_\gamma(z_b)), \tag{4}$$

$$\widehat{y}_a = q_a(z_a), \qquad \gamma = 1. \tag{5}$$

Both encoders are MLPs with hidden width 512:

$$h_c(x) = \mathrm{LN}\{\mathrm{GELU}(W_{c,1}x + b_{c,1})\}, \tag{6}$$

$$f_c(x) = W_{c,2}h_c(x) + b_{c,2}, \qquad c \in \{b, a\}. \tag{7}$$

The decoder has widths $(d_b + d_a) \to 512 \to D$. The biological scanner head has widths $256 \to 128 \to C$, the acquisition scanner head has widths $64 \to 64 \to C$, and $C = 5$ scanners. Gradient reversal is the identity in the forward pass and negates the encoder gradient:

$$\mathrm{GRL}_\gamma(z) = z, \qquad \frac{\partial\, \mathrm{GRL}_\gamma}{\partial z} = -\gamma I. \tag{8}$$

Thus q_b predicts scanner identity while f_b receives the opposite gradient; q_a and f_a minimize ordinary scanner cross-entropy. The architecture is therefore the factorization

$$x \mapsto (z_b, z_a) = (f_b(x), f_a(x)), \qquad \widehat{x} = g([z_b; z_a]), \tag{9}$$

where z_b is optimized for same-region agreement and scanner suppression, and z_a is optimized for scanner prediction and acquisition retention.

3.2 Objective

Let a minibatch be $\{(x_i, r_i, s_i)\}_{i=1}^n$, where r_i is tissue-region identity and $s_i \in \{1, \dots, C\}$ is scanner identity. Write $Z_b \in \mathbb{R}^{n \times d_b}$ and $Z_a \in \mathbb{R}^{n \times d_a}$ for the stacked factors, and let $u_i = z_{b,i}/\|z_{b,i}\|_2$. For anchor i, define $P(i) = \{p \neq i : r_p = r_i\}$ and

$$S_i = \sum_{j \neq i} \exp(u_i^\top u_j/\tau), \tag{10}$$

$$\ell_i = -\frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(u_i^\top u_p/\tau)}{S_i}, \tag{11}$$

$$\mathcal{L}_{\text{pair}} = \frac{1}{n} \sum_{i=1}^n \ell_i, \qquad \tau = 0.1. \tag{12}$$

The reconstruction loss is the mean squared error

$$\mathcal{L}_{\text{recon}} = \frac{1}{nD} \|\widehat{X} - X\|_F^2. \tag{13}$$

For $Z \in \mathbb{R}^{n \times d}$, define

$$v_j(Z) = \frac{1}{n} \sum_{i=1}^n (Z_{ij} - \bar{Z}_j)^2, \tag{14}$$

$$\ell_{\text{var}}(Z) = \frac{1}{d} \sum_{j=1}^d \left[1 - \sqrt{v_j(Z) + \varepsilon}\right]_+, \qquad \varepsilon = 10^{-4}, \tag{15}$$

$$\mathcal{L}_{\text{var},b} = \ell_{\text{var}}(Z_b), \qquad \mathcal{L}_{\text{var},a} = \ell_{\text{var}}(Z_a). \tag{16}$$

Let $Z_b^c = Z_b - \mathbf{1}\bar{z}_b^\top$ and

$$C_b = \frac{1}{n-1} (Z_b^c)^\top Z_b^c. \tag{17}$$

The within-biological covariance penalty is

$$\mathcal{L}_{\text{cov},b} = \frac{1}{d_b} \|C_b - \text{diag}(\text{diag}\, C_b)\|_F^2. \tag{18}$$

The scanner losses are forward cross-entropies:

$$\mathcal{L}_{\text{scan},b} = -\frac{1}{n} \sum_i \log[\text{softmax}\{q_b(\text{GRL}_\gamma(z_{b,i}))\}]_{s_i}, \tag{19}$$

$$\mathcal{L}_{\text{scan},a} = -\frac{1}{n} \sum_i \log[\text{softmax}\{q_a(z_{a,i})\}]_{s_i}. \tag{20}$$

Only the gradient entering f_b is reversed.

Let $Y_s \in \mathbb{R}^{n \times C}$ be the one-hot scanner matrix. Column-standardize Z_b and Y_s using the same $\varepsilon = 10^{-4}$ variance floor, producing $\widetilde{Z}_b$ and $\widetilde{Y}_s$. The direct scanner-dependence penalty is

$$\mathcal{L}_{\text{dep}} = \frac{1}{d_b C} \left\| \frac{1}{n-1} \widetilde{Z}_b^\top \widetilde{Y}_s \right\|_F^2. \tag{21}$$

For $Z_a^c = Z_a - \mathbf{1}\bar{z}_a^\top$, the cross-factor penalty is

$$\mathcal{L}_{\text{xcov}} = \frac{1}{d_b d_a} \left\| \frac{1}{n-1} (Z_b^c)^\top Z_a^c \right\|_F^2. \tag{22}$$

The frozen Paired-Acquisition Neural Factorization objective used in the five-fold and cross-backbone

experiments is

$$\mathcal{L} = \mathcal{L}_{\text{pair}} + \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{var},b} + 0.25\mathcal{L}_{\text{var},a} \tag{23}$$

$$+ 0.01\mathcal{L}_{\text{cov},b} + 0.5\mathcal{L}_{\text{scan},b} + 0.5\mathcal{L}_{\text{scan},a} \tag{24}$$

$$+ 20\mathcal{L}_{\text{dep}} + 0.05\mathcal{L}_{\text{xcov}}. \tag{25}$$

The paired-consistency baseline is

$$\mathcal{L}_{\text{base}} = \mathcal{L}_{\text{pair}} + \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{var},b} + 0.01\mathcal{L}_{\text{cov},b}. \tag{26}$$

It shares the main retention and collapse controls but has no acquisition branch, scanner classifiers, direct scanner-dependence penalty, or cross-factor penalty.

4 Protocol

The DINOv2 objective was frozen after bounded development. The cross-backbone transfer protocol was written before inspecting Phikon or ResNet50 outcomes. No backbone-specific loss weight, architecture width, optimizer setting, fold definition, training duration, checkpoint rule, or success threshold was changed for the transfer backbones.

The frozen settings were:

Quantity	Value
Biological dimension	256
Acquisition dimension	64
Hidden dimension	512
Temperature	0.1
Scanner adversary weight	0.5
Scanner acquisition weight	0.5
Direct scanner-dependence weight	20.0
Biological/acquisition covariance weight	0.05
Reconstruction weight	1.0
Variance weight	1.0
Biological covariance weight	0.01
Epochs	75
Region batch size	32
AdamW learning rate	3×10^{-4}
AdamW weight decay	10^{-4}

The comparison model is paired consistency: it shares the pair, reconstruction, variance, and biological covariance controls but omits the acquisition branch, scanner classifiers, direct scanner-dependence penalty, and cross-factor penalty.

Each backbone uses five rotating original-slide test folds. Transfer backbones use shared seeds 701–705. A transfer backbone passes only if all predefined criteria are met: a scanner-probe reduction of at least 0.15 absolute, a scanner-probe interval entirely below zero, retrieval non-inferiority for mean and worst-pair retrieval, paired-cosine intervals entirely above zero, and nonzero variance in every biological dimension.

5 Results

5.1 Frozen Feature Audit

Before factorization, scanner identity was strongly recoverable from all three frozen feature families even when same-tissue retrieval was high. Phikon had the highest baseline scanner-probe score.

Table 1: Frozen representation audit before Paired-Acquisition Neural Factorization training.

Encoder	Cosine	Retrieval	Scanner
ResNet50	0.8849	0.9550	0.7560
DINOv2-B	0.9851	0.9995	0.8400
Phikon	0.8787	0.9825	0.9640

5.2 Backbone-Specific Means

On DINOv2-Base, the frozen five-fold comparison reduced biological scanner-probe accuracy from 0.7825 to 0.3989, increased mean paired cosine from 0.8476 to 0.8789, and preserved mean same-region retrieval near 0.9998.

Table 2: Backbone-specific descriptive means. Bold entries mark the favorable method for scanner-probe reduction or retention metrics.

Backbone	Method	Mean cosine	Worst cosine	Mean retrieval	Worst retrieval	Scanner probe
DINOv2-B	Paired consistency	0.8476	0.8202	**0.9999**	**0.9991**	0.7825
DINOv2-B	Factorization	**0.8789**	**0.8502**	0.9998	0.9987	**0.3989**
Phikon	Paired consistency	0.7740	0.7364	0.9991	0.9940	0.9543
Phikon	Factorization	**0.8645**	**0.8300**	**0.9997**	**0.9974**	**0.5200**
ResNet50	Paired consistency	0.6286	0.5713	0.9645	0.9356	0.6828
ResNet50	Factorization	**0.6544**	**0.5978**	**0.9726**	**0.9455**	**0.3145**

5.3 Slide-Blocked Contrasts

Differences are Paired-Acquisition Neural Factorization minus paired consistency. Negative scanner-probe values are favorable; positive cosine and retrieval values are favorable.

Table 3: Slide-blocked method contrasts by backbone.

Backbone	Scanner probe	Mean cosine	Worst cosine	Mean retrieval	Worst retrieval
DINOv2-Base	-0.3854	$+0.0310$	$+0.0311$	-0.00008	-0.00042
Phikon	-0.4353	$+0.0900$	$+0.0963$	$+0.00060$	$+0.00479$
ResNet50	-0.3685	$+0.0255$	$+0.0294$	$+0.00808$	$+0.01688$

All three scanner-probe confidence intervals were entirely below zero. All mean- and worst-cosine confidence intervals were entirely above zero. Every backbone satisfied both predefined retrieval non-inferiority margins and every biological dimension retained nonzero test-fold variance.

5.4 Transfer-Backbone Repeated-Measures Summary

Phikon and ResNet50 method differences were averaged within each of the same 48 original slides. This is a repeated-measures summary over 48 slides, not evidence from 96 independent biological units.

Table 4: Repeated-measures transfer summary over Phikon and ResNet50.

Metric	Mean difference	95% slide-bootstrap interval	Favorable slides
Scanner-probe accuracy	-0.401875	$[-0.414458, -0.389458]$	48/48
Mean paired cosine	$+0.057762$	$[+0.053949, +0.061526]$	48/48
Worst paired cosine	$+0.062832$	$[+0.058452, +0.067158]$	48/48
Mean top-1 retrieval	$+0.004344$	$[+0.003313, +0.005427]$	41/48
Worst top-1 retrieval	$+0.010833$	$[+0.007083, +0.014583]$	36/48

All Monte Carlo sign-flip tests were significant at or below 0.000008 with 250,000 draws. Both transfer backbones independently passed all seven predefined criteria, so the cross-backbone transfer claim passed.

5.5 Factorization Audit

The acquisition branch retained scanner information while losing most same-tissue retrieval behavior. This is the expected signature of factor separation, not simply making every representation uninformative.

Table 5: Acquisition-branch audit.

Backbone	Scanner	Retrieval	Rank	XCov RMS
DINOv2-B	0.8602	0.1042	5.4446	0.09280
Phikon	0.9711	0.0739	8.6024	0.08887
ResNet50	0.7845	0.1705	9.2414	0.08957

6 Reproducibility Artifacts

The child repository includes:

- core projection code under `src/models`;
- SCORPION experiment runners under `experiments/scorpion`;
- feature extraction and analysis scripts under `scripts/scorpion`;
- preregistered cross-backbone protocol in `docs/crossbackbone_protocol.md`;
- result summary in `docs/results.md`;
- compact result artifacts under `results/` when available locally;
- this paper source under `paper/arxiv`.

The repository intentionally excludes raw images, extracted patches, feature archives, model checkpoints, and large generated run directories. Completed Windows CUDA runs emitted PyTorch warnings because deterministic algorithms were enabled without setting `CUBLAS_WORKSPACE_CONFIG`; the multi-seed slide-blocked effects are large and directionally consistent, but the completed runs should not be described as guaranteed bitwise deterministic.

7 Claim Boundary

This package supports a bounded representation-identifiability claim: on SCORPION, a frozen Paired-Acquisition Neural Factorization objective reduced linearly recoverable scanner identity in the biological branch while preserving same-tissue retrieval and transferring across DINOv2-Base, Phikon, and ResNet50 feature families.

It does not establish external-dataset generalization, transfer to unseen laboratories, new tissues, staining procedures, scanner models outside SCORPION, clinical safety, diagnostic equivalence, improved patient outcomes, end-to-end whole-slide task performance, or perfect biological/acquisition disentanglement.

8 Conclusion

SCORPION provides paired views of the same human tissue under five scanners. That acquisition structure is informative: it permits a neural factorization in which tissue identity remains concentrated in the biological branch while scanner identity is substantially less recoverable from that branch and remains recoverable in a compact acquisition branch. The same frozen objective passed the preregistered transfer protocol on Phikon and ResNet50 without backbone-specific retuning, making the result stronger than a DINOv2-only finding while keeping the claim explicitly bounded to this paired-scanner benchmark.

References

[1] Alexandre Filiot, Roux Ghermi, Amandine Olivier, Paul Jacob, Lucas Fidon, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. Scaling self-supervised learning for histopathology with masked image modeling. *medRxiv*, 2023.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[3] 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 Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[4] Jeongun Ryu, Heon Song, Seungeun Lee, Soo Ick Cho, Jiwon Shin, Kyunghyun Paeng, and Sérgio Pereira. SCORPION: Addressing scanner-induced variability in histopathology. *arXiv preprint arXiv:2507.20907*, 2025.