PPFUSE: FUSE WHAT ALIGNS, PRESERVE WHAT DOES NOT

Anonymous authors
Paper under double-blind review

ABSTRACT

Frozen pathology foundation models provide complementary representations of the same tissue patch, motivating feature fusion for whole-slide image classification. Aligning a lower-dimensional auxiliary representation to an anchor subspace leaves the treatment of the anchor orthogonal complement unspecified. Full-space interpolation implicitly ties complement scaling to the reachable-space mixing weight, although the correspondence does not require this coupling. We introduce **PPFuse**, which fits a semi-orthogonal correspondence from paired training features without task labels and decouples these decisions. After scale-matched mixing in the reachable coordinates, PPFuse adopts a minimum-change principle relative to the anchor: for the fixed correspondence and fused coordinates, the unique closest reconstruction in Frobenius distance preserves the complement at its native scale. The resulting transform is closed-form and retains the anchor dimensionality. Across three multiclass whole-slide datasets and three multiple instance learning backbones, PPFuse achieves higher mean balanced accuracy than both individual encoders in eight of nine settings and ranks first among the evaluated methods in seven. Across five few-shot sampling seeds, PPFuse exceeds both individual encoders in mean ROC-AUC across all nine backbone–shot settings. Component analyses show that the retained complement carries predictive signal and support native-scale preservation as a competitive fixed default across the evaluated datasets and backbones. Alignment controls further show that downstream adaptation narrows the performance gap between fitted and random alignments, highlighting the distinction between geometric alignment and predictive utility. Additional natural-image and image–text experiments demonstrate competitive performance against the evaluated fusion baselines while retaining the anchor dimensionality.

1 INTRODUCTION

Pathology foundation models (PFMs) learn from different datasets, objectives, and architectures, producing complementary representations of tissue (Neidlinger et al., 2026). In whole-slide image (WSI) classification, multiple frozen PFMs encode identical patches, and multiple instance learning (MIL) aggregates the resulting features into a slide prediction. Paired patch representations thus provide a natural setting for multi-view feature fusion (Xu et al., 2013).

Prediction ensembles combine slide-level outputs (Huang et al., 2026), while concatenation and learned fusion modules integrate features before or during downstream training (Gao et al., 2026; Xiao et al., 2026). A natural feature-fusion strategy is to align the auxiliary representation to the anchor. Under dimensional mismatch, the fitted map defines an auxiliary-reachable subspace and a coordinate correspondence within it. Alignment alone does not fully specify fusion: even after the fused reachable coordinates are chosen, the treatment of the anchor orthogonal complement remains unspecified. Full-space interpolation resolves this remaining freedom implicitly by tying complement scaling to the reachable-space mixing weight.

We make these two decisions explicit in **PPFuse** (Preserve-Private Fusion), separating reachable-space mixing from complement treatment. A semi-orthogonal Procrustes map defines the reachable subspace, within which PPFuse interpolates corresponding coordinates after scale matching. To resolve the remaining freedom without task-label supervision, we adopt a minimum-change principle

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

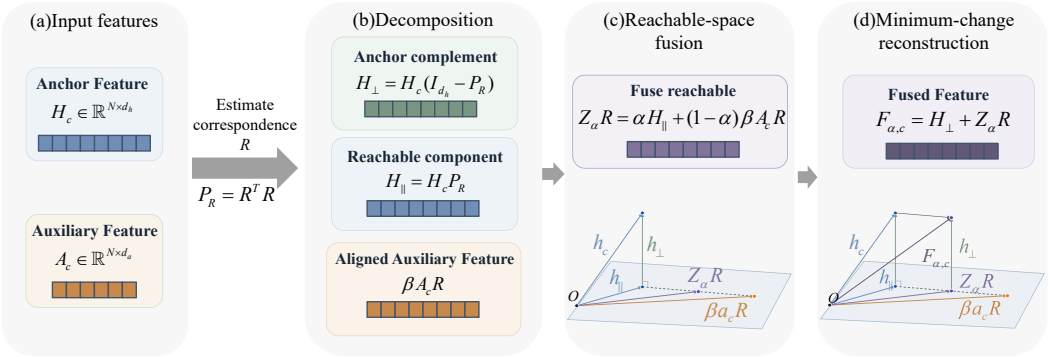


Figure 1: A semi-orthogonal map R embeds the auxiliary representation into an auxiliary-reachable subspace of the anchor space and decomposes the anchor into reachable and orthogonal components. PPFuse fuses the reachable anchor component with the scale-matched mapped auxiliary feature, then adds the unchanged anchor complement to obtain a centered fused representation with the anchor dimensionality. The anchor mean is restored after fusion.

relative to the anchor. For the fixed map and fused coordinates, the unique reconstruction minimizing the Frobenius distance to the anchor retains its orthogonal complement unchanged (Figure 1). Thus, the principle yields **fuse within the auxiliary-reachable subspace; preserve its orthogonal complement.** The transform is fitted in closed form from paired training representations and retains the anchor dimensionality.

Across multiclass and few-shot pathology MIL tasks, PPFuse achieves higher mean performance than both individual encoders in most evaluated settings. Additional evaluations use MLP classifiers on natural-image and image–text tasks.

Contributions.

- **Decoupling reachable-space mixing and complement treatment.** We identify the freedom left in complement treatment by a fitted correspondence and prescribed reachable coordinates. Full-space interpolation implicitly couples complement scaling to the mixing weight; PPFuse separates these two decisions.
- **A minimum-change reconstruction principle.** For a fixed correspondence and prescribed fused reachable coordinates, minimizing the Frobenius distance to the anchor uniquely yields complement preservation. This gives a closed-form fusion transform fitted without task labels, retaining the anchor dimensionality. We also establish its exact relation to full-space interpolation: for $\alpha > 0$, the two centered representations are related by an invertible rescaling of the anchor complement.
- **Controlled empirical analysis.** Complement-scale interventions on BRACS and PAM50 support native-scale preservation as a competitive fixed default under a common downstream training protocol, while alignment and adapter controls examine how downstream adaptation moderates the predictive benefit of geometric correspondence.

2 Related Work

2.1 Fusion of Frozen Representations

Multi-view learning integrates complementary observations of a sample (Xu et al., 2013). In computational pathology, frozen PFMs provide multiple representations of identical tissue patches. A benchmark of 19 histopathology foundation models reports task-dependent performance and complementary predictive signals, including improvements from cross-PFM prediction ensembles (Neidlinger et al., 2026). Existing integration strategies operate at several levels: LogitProd combines

slide-level predictions (Huang et al., 2026); Meta-Encoder learns feature-integration mechanisms (Gao et al., 2026); AdaFusion uses sample-conditioned weighting (Xiao et al., 2026); and COBRA learns a model-agnostic slide representation from several feature extractors (Lenz et al., 2025). Our setting differs from these learned or slide-level integration schemes: PPFuse performs closed-form fusion of paired frozen features before downstream learning, without task-label supervision for the fusion transform.

The same feature-level question also arises when the paired observations come from different modalities. Gated Multimodal Units learn supervised interactions between image and text features and introduced the poster–synopsis MM-IMDb benchmark (Arevalo et al., 2017). ImageBind learns a common embedding space through multimodal pretraining (Girdhar et al., 2023).

2.2 Representation Alignment and Functional Compatibility

A feature-level fusion rule needs coordinates in which its inputs can be combined. Canonical correlation analysis estimates correlated projections, while Orthogonal Procrustes fits a least-squares isometry between paired matrices; deep CCA learns nonlinear correlated mappings (Hotelling, 1936; Schoenemann, 1966; Andrew et al., 2013). PPFuse uses rectangular semi-orthogonal Procrustes to specify both an anchor subspace and the auxiliary coordinate correspondence within it.

SVCCA and CKA measure representation similarity (Raghu et al., 2017; Kornblith et al., 2019). Representation-similarity studies show that foundation-model similarities vary across datasets and training objectives (Ciernik et al., 2025); pathology-specific analyses report structural variation among PFMs (Mishra & Lotter, 2026). Model stitching evaluates functional compatibility through a learned interface between frozen networks (Lenc & Vedaldi, 2015; Bansal et al., 2021); learned projectors likewise connect frozen vision and language encoders (Maniparambil et al., 2025). For feature fusion, these perspectives motivate assessing both geometric correspondence and the predictive utility of the resulting representation.

2.3 Shared–Private Decomposition and Preservation

Once a correspondence has been established, fusion must determine how to combine the aligned components and how to treat the remaining information. Factorized orthogonal latent spaces, domain-separation networks, and modality-invariant/specific architectures learn shared and private factors through combinations of reconstruction, separation, orthogonality, and task objectives (Salzmann et al., 2010; Bousmalis et al., 2016; Hazarika et al., 2020). In PPFuse, the fitted semi-orthogonal map directly defines an anchor projection and its orthogonal residual. This geometric decomposition provides the reachable–complement structure on which PPFuse operates.

3 Method

We formulate PPFuse for paired frozen representations. The method fits an auxiliary-to-anchor map, decomposes the anchor into reachable and orthogonal components, and interpolates in the reachable coordinates. In pathology, each pair consists of two encodings of the same tissue patch; the formulation also accommodates paired image encodings and associated image–text inputs. The resulting features are supplied to the downstream predictor. Appendix A provides the derivations and fitting/fusion pseudocode.

3.1 Paired Representations and Alignment

Let $H \in \mathbb{R}^{N \times d_h}$ and $A \in \mathbb{R}^{N \times d_a}$, with $d_a \leq d_h$, denote paired frozen representations from the anchor and auxiliary encoders. Rows with the same index correspond to the same or associated observations, expressed in the independently learned coordinate systems of the two encoders.

For fitting, H and A contain paired training observations. Let $\mu_H \in \mathbb{R}^{d_h}$ and $\mu_A \in \mathbb{R}^{d_a}$ denote their global means, and define the centered feature matrices

$$H_c = H - \mathbf{1}\mu_H^\top, \qquad A_c = A - \mathbf{1}\mu_A^\top. \tag{1}$$

Following classical orthogonal Procrustes (Schoenemann, 1966), we solve a rectangular semi-orthogonal variant for $R \in \mathbb{R}^{d_a \times d_h}$:

$$R^\star \in \underset{RR^\top = I_{d_a}}{\arg\min} \ \|A_c R - H_c\|_F^2. \tag{2}$$

If $A_c^\top H_c = U\Sigma V^\top$ is a thin singular value decomposition, the closed-form minimizer is

$$R = UV^\top, \qquad RR^\top = I_{d_a}. \tag{3}$$

We omit the superscript below. The map R is an isometric embedding of auxiliary coordinates into the anchor space. It specifies both an anchor subspace and the coordinate correspondence used for fusion within that subspace. Appendix A.1 derives the solution and describes its non-uniqueness when the cross-covariance is rank-deficient.

3.2 Reachable Decomposition

The semi-orthogonal map induces a precise notion of which anchor directions are reachable from auxiliary coordinates. We define the *auxiliary-reachable subspace* as

$$\mathcal{S}_R := \mathrm{range}(R^\top) = \mathrm{row}(R) = \left\{ zR \mid z \in \mathbb{R}^{1 \times d_a} \right\} \subseteq \mathbb{R}^{d_h}. \tag{4}$$

Here, "reachable" denotes the algebraic image of auxiliary coordinates under R, with dimension d_a by construction. Every row of $A_c R$ lies in this subspace. Since $RR^\top = I_{d_a}$,

$$P_R = R^\top R = VV^\top \tag{5}$$

is the orthogonal projector from the anchor space onto $\mathcal{S}_R$ (see Appendix A.2).

The induced anchor coordinates are $Z_H = H_c R^\top \in \mathbb{R}^{N \times d_a}$. The projector decomposes the centered anchor into reachable and orthogonal components:

$$\begin{aligned} H_\parallel &= H_c P_R = (H_c R^\top)R = Z_H R, & H_\perp &= H_c(I_{d_h} - P_R) = H_c - Z_H R, \\ H_c &= H_\parallel + H_\perp, & H_\perp R^\top &= 0. \end{aligned} \tag{6}$$

Thus, $A_c R$ and $H_\parallel$ lie in $\mathcal{S}_R$, while $H_\perp$ lies in its orthogonal complement. In Preserve-Private Fusion, "private" denotes this geometric residual relative to the fitted map. When $d_a = d_h$, $P_R = I_{d_h}$ and $H_\perp = 0$; the fitted map still determines the coordinate correspondence within the full anchor space.

3.3 Scale-Matched Fusion and Preservation

The decomposition places Z_H and A_c in corresponding d_a-dimensional coordinates. With (μ_H, μ_A, R) fixed after alignment fitting, we compute the scale factor from all paired representations in the training partition. Let H_c^{tr} and A_c^{tr} denote these representations centered by the fitted means, and let $Z_H^{\mathrm{tr}} = H_c^{\mathrm{tr}} R^\top$. We use

$$\beta = \frac{\|Z_H^{\mathrm{tr}}\|_F}{\|A_c^{\mathrm{tr}}\|_F}, \qquad \|A_c^{\mathrm{tr}}\|_F > 0. \tag{7}$$

This gives $\|\beta A_c^{\mathrm{tr}}\|_F = \|Z_H^{\mathrm{tr}}\|_F$. For $\alpha \in [0, 1]$, the fused coordinates are

$$Z_\alpha = \alpha Z_H + (1 - \alpha)\beta A_c. \tag{8}$$

Complement treatment remains unspecified. Fixing R and Z_α constrains a centered reconstruction G through $GR^\top = Z_\alpha$, but leaves $G(I_{d_h} - P_R)$ free. To expose the choice made by full-space interpolation, consider the one-parameter reconstruction family

$$F_{\alpha,\lambda} = \mathbf{1}\mu_H^\top + \lambda H_\perp + Z_\alpha R, \qquad \lambda \in \mathbb{R}. \tag{9}$$

Every member has the prescribed centered reachable coordinates Z_α. Full-space OP interpolation chooses $\lambda = \alpha$, since $F_{\alpha,\alpha} - \mathbf{1}\mu_H^\top = \alpha H_c + (1 - \alpha)\beta A_c R$. The correspondence does not require this coupling. This family illustrates the scaling choice; the correspondence constraint also permits complement components other than scalar multiples of $H_\perp$.

Preservation as minimum-change reconstruction. To resolve this freedom without task-label supervision, we adopt a minimum-change principle relative to the anchor. For fixed R and Z_α, we minimize over all feasible centered reconstructions:

$$\operatorname*{argmin}_{G\in\mathbb{R}^{N\times d_h}} \|G - H_c\|_F^2 \quad \text{subject to} \quad GR^\top = Z_\alpha. \tag{10}$$

The constraint fixes the reachable component to $Z_\alpha R$; minimizing the remaining change uniquely retains $H_\perp$ (Appendix A.3). Restoring the fixed training anchor mean defines PPFuse. Equivalently, the same operation is a correction to the raw anchor:

$$F_\alpha := F_{\alpha,1} = \mathbf{1}\mu_H^\top + H_\perp + Z_\alpha R, \tag{11}$$

$$= H + (1-\alpha)(\beta A_c - Z_H)R. \tag{12}$$

Thus, $F_{\alpha,c} := F_\alpha - \mathbf{1}\mu_H^\top$ is the unique solution to Eq. 10. The same decomposition establishes exact complement preservation for every $\alpha \in [0,1]$. Since $R(I_{d_h} - P_R) = 0$,

$$F_{\alpha,c}(I_{d_h} - P_R) = H_c(I_{d_h} - P_R) = H_\perp. \tag{13}$$

The endpoints of α have direct interpretations: $\alpha = 0$ replaces the anchor reachable component $Z_H R$ with the scale-matched aligned auxiliary component $\beta A_c R$, whereas $\alpha = 1$ recovers the raw anchor representation exactly. With equal encoder dimensions, the construction reduces to scale-matched interpolation in the full anchor space.

Preserving information versus preserving scale. The $\lambda = \alpha$ member of the reconstruction family has an exact relation to PPFuse. For $\alpha > 0$, centered full-space OP interpolation $G_\alpha := F_{\alpha,\alpha} - \mathbf{1}\mu_H^\top$ satisfies

$$G_\alpha = F_{\alpha,c}\big[P_R + \alpha(I - P_R)\big] = \alpha H_c + (1-\alpha)\beta A_c R, \qquad \alpha > 0. \tag{14}$$

The right multiplier is invertible, with inverse $P_R + \alpha^{-1}(I - P_R)$. Thus, G_α and PPFuse retain the same linear information for $\alpha > 0$, but G_α scales the anchor complement by α, whereas PPFuse preserves its original scale. At $\alpha = 0$, the multiplier is singular when the complement is nonempty, and full-space interpolation discards that component. Although positive rescalings retain the same linear information, their relative component scales can affect learning under a fixed downstream training procedure. Section 4.4 examines this distinction through controlled interventions on the complement scale.

Application to held-out observations. After fitting, (μ_H, μ_A, R, β) are held fixed. For any paired collection q containing N_q observations, PPFuse produces $F_\alpha^{(q)} \in \mathbb{R}^{N_q \times d_h}$, preserving both the number of observations and the anchor feature dimensionality. It can therefore replace the raw anchor representation without changing the downstream feature interface or architecture.

4 Experiments

4.1 Datasets and Experimental Setup

Datasets and representations: Table 1 summarizes four WSI cohorts and two additional non-pathology datasets. BRACS-7 (Brancati et al., 2022) uses the official split; CamelyonPlus (Ling et al., 2025), TCGA-BRCA (The Cancer Genome Atlas Network, 2012), and TCGA-NSCLC (Campbell et al., 2016) use patient-level five-fold splits. The TCGA-BRCA task predicts four PAM50 subtypes after excluding the Normal-like class, which contains eight samples. For all WSI tasks, frozen features are extracted from non-overlapping 256×256 patches at $20\times$ magnification. Matching patch coordinates across encoders yields paired 1536-dimensional UNI2-h (hereafter UNI2) and 768-dimensional CONCHv1.5 representations, used as anchor and auxiliary, respectively (Chen et al., 2024; Ding et al., 2025). The representation-configuration study also uses Virchow2 and its CLS and mean-patch summaries (Zimmermann et al., 2024). Appendix B gives the complete evaluation protocols and feature extraction details.

Fusion comparisons: We compare feature constructions using the same frozen encoder outputs and downstream MIL backbone within each setting. The main baselines are individual encoders, Raw

Table 1: Datasets used in the evaluation. Sample counts reflect the experimental data used in this work.

Dataset	Domain	Task	Classes / labels	Samples
BRACS	Pathology WSI	Fine-grained lesion classification	7	546 WSIs
CamelyonPlus	Pathology WSI	Lymph-node metastasis classification	4	1,350 WSIs
TCGA-BRCA	Pathology WSI	PAM50 subtype classification	4	480 WSIs
TCGA-NSCLC	Pathology WSI	LUAD vs. LUSC few-shot classification	2	936 WSIs
Stanford Cars	Natural image	Fine-grained image classification	196	16,185 images
MM-IMDb	Image–text	Multi-label genre classification	23	25,959 pairs

Concat, RMS Concat, PCA RMS Concat, MLP Fusion, and Ridge CCA Concat. Raw and RMS Concat compare concatenation before and after global per-view scale normalization; PCA RMS Concat additionally compresses the normalized concatenation to the anchor dimensionality. We evaluate each construction with ABMIL, DSMIL, and TransMIL (Ilse et al., 2018; Li et al., 2021; Shao et al., 2021). PPFuse fits (μ_H, μ_A, R) on at most 512 uniformly sampled paired patches per training WSI and computes β from all paired patches in the training partition. We fix $\alpha = 0.5$ a priori as an equal-weight default across all datasets and backbones, without validation-based tuning. Appendices C and F specify training and fusion controls.

Evaluation: Balanced accuracy (BA) is the primary multiclass WSI metric and ROC-AUC is the primary binary TCGA-NSCLC metric. We report means and standard deviations over five training seeds for BRACS-7 and five patient-level folds for the other WSI tasks. For paired BA comparisons, we report effect estimates with unadjusted 95% percentile intervals from 10,000 bootstrap replicates to characterize test-patient sampling uncertainty (Appendix D). BA and Macro-F1 are the primary metrics for Cars and MM-IMDb, respectively (Section 4.6).

4.2 Multi-PFM Fusion Across WSI Tasks

PPFuse achieves higher mean BA than both individual encoders in eight of nine dataset–backbone settings (Table 2). It ranks first in seven: all six BRACS-7 and PAM50 settings and CamelyonPlus with ABMIL. Raw Concat and MLP Fusion lead on CamelyonPlus with DSMIL and TransMIL, respectively.

Global scale normalization affects concatenation unevenly across datasets and backbones. PPFuse achieves higher mean BA than both RMS Concat and dimension-matched PCA RMS Concat in all nine settings. Table 7 reports paired effect estimates and bootstrap intervals.

4.3 Few-Shot Classification

TCGA-NSCLC few-shot classification uses nested subsets of 4, 8, and 16 patients per class within each fold, with the complete validation and test partitions held fixed. All WSIs from each selected training patient are used, and PPFuse fits its correspondence on these training WSIs. Across five independently generated few-shot sampling seeds, each evaluated over five patient-level folds, PPFuse achieves higher fold-averaged ROC-AUC than the stronger of UNI2 and CONCHv1.5 in 39 of 45 sampling-seed–backbone–shot combinations. After averaging over the five sampling seeds, PPFuse remains above both individual encoders in all nine backbone–shot settings and achieves the highest mean ROC-AUC among the evaluated methods in seven. Five exceptions occur with ABMIL and one with TransMIL. This shows that the advantage varies with the few-shot patient selection. Full results are reported in Appendix E.

Table 2: Balanced accuracy of individual representations and fusion methods. We report mean ± standard deviation over five training seeds for BRACS-7 and five patient-level folds for Camelyon-Plus and PAM50. Output dim denotes the patch feature dimension supplied to MIL. Highest means within each dataset and MIL backbone are shown in bold.

Method	Output dim	ABMIL	DSMIL	TransMIL
BRACS-7				
UNI2	1536	0.4786 ± 0.0407	0.4459 ± 0.0351	0.4151 ± 0.0389
CONCHv1.5	768	0.4393 ± 0.0156	0.3632 ± 0.0390	0.3987 ± 0.0393
Raw Concat	2304	0.4777 ± 0.0286	0.4691 ± 0.0355	0.4178 ± 0.0430
RMS Concat	2304	0.4631 ± 0.0273	0.4253 ± 0.0644	0.4160 ± 0.0432
PCA RMS Concat	1536	0.3999 ± 0.0367	0.2844 ± 0.0465	0.3413 ± 0.0236
MLP Fusion	1536	0.4838 ± 0.0105	0.4664 ± 0.0317	0.3874 ± 0.0603
Ridge CCA Concat	1536	0.3337 ± 0.0237	0.2513 ± 0.0235	0.2306 ± 0.0302
PPFuse	1536	$\mathbf{0.5062 \pm 0.0512}$	$\mathbf{0.4859 \pm 0.0334}$	$\mathbf{0.4393 \pm 0.0438}$
CamelyonPlus				
UNI2	1536	0.6436 ± 0.0221	0.6305 ± 0.0121	0.6630 ± 0.0341
CONCHv1.5	768	0.6516 ± 0.0406	0.6438 ± 0.0127	0.6436 ± 0.0287
Raw Concat	2304	0.6276 ± 0.0150	$\mathbf{0.6764 \pm 0.0276}$	0.6634 ± 0.0293
RMS Concat	2304	0.6467 ± 0.0499	0.6197 ± 0.0428	0.6593 ± 0.0185
PCA RMS Concat	1536	0.6332 ± 0.0361	0.5684 ± 0.0323	0.5721 ± 0.0697
MLP Fusion	1536	0.6450 ± 0.0296	0.6558 ± 0.0342	$\mathbf{0.7084 \pm 0.0243}$
Ridge CCA Concat	1536	0.5941 ± 0.0392	0.5432 ± 0.0258	0.4164 ± 0.0173
PPFuse	1536	$\mathbf{0.6539 \pm 0.0176}$	0.6582 ± 0.0167	0.6612 ± 0.0437
PAM50				
UNI2	1536	0.5593 ± 0.0177	0.5374 ± 0.0229	0.5157 ± 0.0341
CONCHv1.5	768	0.5472 ± 0.0178	0.4909 ± 0.0541	0.5337 ± 0.0473
Raw Concat	2304	0.5406 ± 0.0362	0.5239 ± 0.0364	0.5122 ± 0.0454
RMS Concat	2304	0.5152 ± 0.0234	0.5248 ± 0.0122	0.4944 ± 0.0210
PCA RMS Concat	1536	0.5266 ± 0.0372	0.4676 ± 0.0542	0.4771 ± 0.0437
MLP Fusion	1536	0.5699 ± 0.0331	0.5510 ± 0.0390	0.5233 ± 0.0582
Ridge CCA Concat	1536	0.4857 ± 0.0198	0.3284 ± 0.0417	0.3972 ± 0.0523
PPFuse	1536	$\mathbf{0.5747 \pm 0.0438}$	$\mathbf{0.5713 \pm 0.0401}$	$\mathbf{0.5585 \pm 0.0434}$

4.4 Mechanistic Analysis

We examine the predictive role of the retained complement, its scale, and the contribution of fitted correspondence under different downstream adapters. Controls use the UNI2–CONCHv1.5 pair on BRACS-7, with complement interventions repeated on PAM50.

Predictive role of the complement: We assess the complement both as a standalone input and through its removal from the fused representation. On BRACS-7, its scale-matched 768-dimensional coordinates alone give ABMIL 0.2520 ± 0.0503 BA, above the $1/7$ chance level (Appendix F.8). Removing it in the Reachable-only control lowers mean BA across all three BRACS backbones (Table 3). On PAM50, retaining the complement raises mean BA with DSMIL and TransMIL, including a 6.486-percentage-point gain with TransMIL, while ABMIL has a slightly lower mean BA. These results identify predictive signal in the complement and show that the benefit of retaining it varies with the dataset and backbone. Tables 11 and 12 report the paired differences and bootstrap intervals.

Role of complement scale: At fixed $\alpha = 0.5$, we vary λ in Eq. 9 under identical fitting and training settings: $\lambda = 0$ removes the complement, 0.5 gives full-space OP interpolation, and 1 gives PPFuse. Positive values change component scale while retaining the same linear information.

Table 3: BRACS-7 fusion-design mean BA.

Design	ABMIL	DSMIL	TransMIL
Raw Concat	0.4777	0.4691	0.4178
Raw-Linear	0.4627	0.3970	0.4124
MLP Fusion	0.4838	0.4664	0.3874
OP Concat	0.4724	0.4579	0.4206
Reachable-only	0.4406	0.4408	0.4214
Full-space OP	0.4780	0.4599	0.4377
PPFuse	**0.5062**	**0.4859**	**0.4393**

Table 4: BRACS-7 Random-Q downstream BA.

Orientation	ABMIL	DSMIL	TransMIL
PPFuse	**0.5062**	**0.4859**	**0.4393**
Random-Q	0.4829	0.4649	0.4200

Table 5: BRACS-7 held-out alignment geometry.

Diagnostic	Random-Q	OP
Same-patch cosine ↑	0.0019 ± 0.0057	$\mathbf{0.4905}$
Margin ↑	0.0006 ± 0.0027	$\mathbf{0.3327}$
R@1 ↑	$0.38 \pm 0.12\%$	$\mathbf{38.20\%}$
Residual ↓	1.5271 ± 0.0042	$\mathbf{1.1035}$

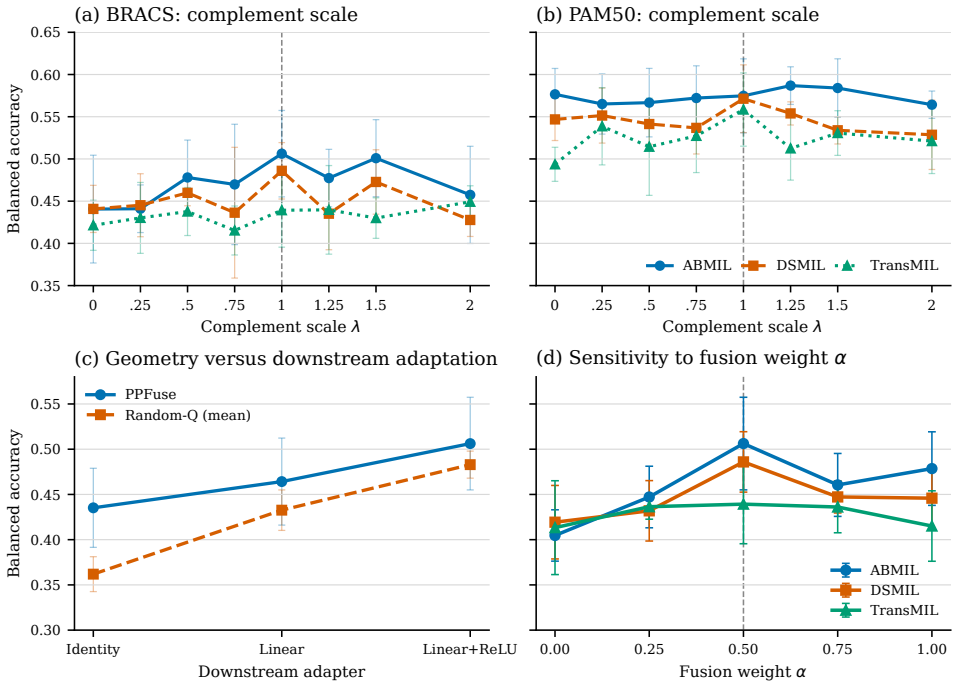


Figure 2: Mechanistic analyses. (a,b) Complement scaling on BRACS-7 and PAM50 ($\alpha = 0.5$; dashed line: PPFuse's $\lambda = 1$). (c) PPFuse versus the mean of five Random-Q orientations across BRACS ABMIL adapters. (d) BRACS fusion-weight sensitivity (dashed line: $\alpha = 0.5$). Markers show mean BA; error bars show one standard deviation over five runs, or across five orientation means for Random-Q in (c).

The fixed PPFuse choice $\lambda = 1$ achieves the highest mean BA in four of six dataset–backbone settings and the highest backbone-averaged BA on both datasets (Figure 2(a,b)). The exceptions favor $\lambda = 2$ for BRACS–TransMIL and 1.25 for PAM50–ABMIL. Thus, native-scale preservation is a competitive fixed default, with the highest-mean scale varying by dataset and backbone. Appendix F.7 gives the full sweep; paired comparisons with full-space OP appear in Tables 11 and 12.

Fusion weight: With the map fixed, α controls the balance of the two reachable representations. Figure 2(d) shows that $\alpha = 0.5$ gives the highest mean BA for ABMIL and DSMIL, while TransMIL

varies little over $\alpha \in \{0.25, 0.5, 0.75\}$. Performance decreases at $\alpha = 0$, where the auxiliary representation replaces the anchor reachable component; $\alpha = 1$ recovers raw UNI2.

Correspondence fitting: We compare OP with random semi-orthogonal orientations (Random-Q) that also preserve auxiliary inner products and distances. On held-out validation patches, same-patch R@1 is 38.20% with OP and $0.38 \pm 0.12\%$ with Random-Q; the chance level is 0.391% for 256 candidates per slide. The same-patch cosine and residual also indicate closer alignment with OP (Table 5; Appendix F.9). The Random-Q entries summarize five fixed orientations: mean $\pm$ standard deviation for geometry and mean BA for downstream results. This fitted correspondence also yields higher mean BA across all three MIL backbones (Table 4).

Interaction with downstream adaptation: Figure 2(c) compares PPFuse with the mean of five fixed Random-Q orientations under identity, linear, and Linear+ReLU adapters. Adaptation improves both curves and narrows their mean BA gap, while PPFuse remains higher, suggesting that downstream learning can partly compensate for weaker initial alignment. Per-orientation results are reported in Appendix F.9.

4.5 Fusion with Composite and Constituent Representations

On BRACS-7 and CamelyonPlus, we compare Virchow2's full descriptor with its CLS and mean-patch constituents. Full-descriptor fusion anchors on Virchow2; constituent fusion anchors on UNI2, so composition, anchor direction, and output dimension change jointly. PPFuse achieves higher mean BA than the corresponding raw constituent in all 12 constituent settings, and than the stronger of raw UNI2 and that constituent in nine. Full-descriptor results are mixed; Appendix G gives the paired deltas and absolute results.

4.6 Evaluation Beyond Pathology

Using frozen DINOv2-B and CLIP ViT-B/16 representations (Oquab et al., 2024; Radford et al., 2021), we evaluate same-image pairs on Stanford Cars (Krause et al., 2013) and poster–synopsis pairs on MM-IMDb (Arevalo et al., 2017). PPFuse ranks second among seven methods on both tasks—by mean BA on Cars and mean Macro-F1 on MM-IMDb—following RMS Concat, while retaining the 768-dimensional anchor interface rather than the 1280-dimensional concatenated representation. Appendix H reports complete results and protocols.

5 Conclusion

A fitted alignment defines reachable-space correspondence but leaves the treatment of the anchor orthogonal complement unspecified, while full-space interpolation implicitly ties complement scaling to the reachable-space mixing weight. PPFuse separates these decisions and adopts a minimum-change principle: for fixed correspondence and fused coordinates, the unique reconstruction minimizing the Frobenius distance to the anchor preserves its orthogonal complement. The method achieves higher mean performance than both individual encoders in most multiclass and few-shot WSI settings. Component diagnostics identify predictive signal in the retained complement, and adapter controls show that downstream learning can partly compensate for weaker initial alignment. Evaluations with MLP classifiers also demonstrate its applicability to paired image and image–text representations.

6 Limitations and Future Work

PPFuse fuses paired representations, with the higher-dimensional input serving as the anchor in our experiments. Selecting which representation to fuse remains open: for example, whether to pair UNI2 with the full Virchow2 descriptor, its CLS token, or its mean-patch summary. Our comparisons do not establish an automatic selection rule. Direct reuse of dataset-fitted fusion statistics across cohorts and joint fusion of more than two representations remain untested.

AI USE STATEMENT

We used generative AI tools to assist with methodological and experimental discussion, review of mathematical derivations and proofs, coding and debugging, and interpretation of experimental results. We also used these tools for literature search and summarization, manuscript drafting and editing, and feedback on scientific figure design. The authors independently reviewed and verified all AI-assisted mathematical claims, code, experimental analyses, citations, and manuscript content. All reported experimental results were obtained by running the computational experiments described in this paper. The authors take full responsibility for the final content of the paper.

ETHICS STATEMENT

This work uses existing publicly available research datasets and does not involve the collection of new human-subject data. The pathology experiments are conducted on de-identified benchmark cohorts, and all datasets are used in accordance with their respective access conditions and intended research purposes. PPFuse is a representation-fusion method evaluated for research purposes and is not intended for direct clinical decision-making. As with other computational pathology studies, performance may be affected by cohort, demographic, institutional, and acquisition biases; therefore, results on the evaluated benchmarks should not be interpreted as evidence of clinical validity without appropriate external and prospective validation.

REPRODUCIBILITY STATEMENT

Appendix A provides complete derivations and pseudocode for fitting and applying PPFuse. Appendix B documents dataset protocols and pathology feature extraction, and Appendix C specifies the MIL implementations, optimization settings, random seeds, and checkpoint-selection procedures. Appendices D and E detail the bootstrap and few-shot evaluation protocols, while Appendices F–H describe the fusion controls, representation-configuration experiments, and non-pathology evaluations. All fusion statistics are estimated using training data only and held fixed for validation and test data. Exact patient- and slide-level split manifests, TCGA-NSCLC few-shot patient-selection manifests, and source code and configuration files are provided in the supplementary material.

REFERENCES

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1247–1255. PMLR, 2013.

John Arevalo, Thamar Solorio, Manuel Montes-y-Gómez, and Fabio A. González. Gated multimodal units for information fusion. In *International Conference on Learning Representations (ICLR) Workshop Track Proceedings*, 2017.

Yamini Bansal, Preetum Nakkiran, and Boaz Barak. Revisiting model stitching to compare neural representations. In *Advances in Neural Information Processing Systems*, volume 34, pp. 225–236, 2021.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in Neural Information Processing Systems*, volume 29, pp. 343–351, 2016.

Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, Maria Gabrani, Florinda Feroce, and Maria Frucci. BRACS: A dataset for BReAst carcinoma subtyping in H&E histology images. *Database*, 2022:baac093, 2022. doi: 10.1093/database/baac093.

Joshua D. Campbell, Anton Alexandrov, Jaegil Kim, Jeremiah Wala, Alice H. Berger, Chandra Sekhar Pedamallu, Sachet A. Shukla, Guangwu Guo, Angela N. Brooks, Bradley A. Murray, Marcin Imielinski, Xin Hu, Shiyun Ling, Rehan Akbani, Mara Rosenberg, Carrie Cibulskis, Aruna Ramachandran, Eric A. Collisson, David J. Kwiatkowski, Michael S. Lawrence, John N. Weinstein, Roel G. W. Verhaak, Catherine J. Wu, Peter S. Hammerman, Andrew D. Cherniack,

Gad Getz, Cancer Genome Atlas Research Network, Maxim N. Artyomov, Robert Schreiber, Ramaswamy Govindan, and Matthew Meyerson. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature Genetics*, 48(6):607–616, 2016. doi: 10.1038/ng.3564.

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*, 30:850–862, 2024. doi: 10.1038/s41591-024-02857-3.

Laure Ciernik, Lorenz Linhardt, Marco Morik, Jonas Dippel, Simon Kornblith, and Lukas Muttenthaler. Objective drives the consistency of representational similarity across datasets. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 10920–10948. PMLR, 2025.

Tong Ding, Sophia J. Wagner, Andrew H. Song, Richard J. Chen, Ming Y. Lu, Andrew Zhang, Anurag J. Vaidya, Guillaume Jaume, Muhammad Shaban, Ahrong Kim, Drew F. K. Williamson, Harry Robertson, Bowen Chen, Cristina Almagro-Pérez, Paul Doucet, Sharifa Sahai, Chengkuan Chen, Christina S. Chen, Daisuke Komura, Akihiro Kawabe, Mieko Ochi, Shinya Sato, Tomoyuki Yokose, Yohei Miyagi, Shumpei Ishikawa, Georg Gerber, Tingying Peng, Long Phi Le, and Faisal Mahmood. A multimodal whole-slide foundation model for pathology. *Nature Medicine*, 31: 3749–3761, 2025. doi: 10.1038/s41591-025-03982-3.

Ruitian Gao, Zhaochang Yang, Xin Yuan, Yifan Wang, Yujia Xia, Yufei Zhang, Bingyan Zheng, Yuqiao Gong, Yiping Yue, Yansong Tu, and Zhangsheng Yu. Meta-encoder: a unified integration framework for multiple pathological foundation models in cancer detection. *Nature Communications*, 17:5738, 2026. doi: 10.1038/s41467-026-71558-x.

Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. ImageBind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15180–15190, 2023. doi: 10.1109/CVPR52729.2023.01457.

Devamanyu Hazarika, Roger Zimmermann, and Soujanya Poria. MISA: Modality-invariant and -specific representations for multimodal sentiment analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 1122–1131, 2020. doi: 10.1145/3394171.3413678.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3–4):321–377, 1936. doi: 10.1093/biomet/28.3-4.321.

Gexin Huang, Anqi Li, Yusheng Tan, Beidi Zhao, Gang Wang, Zu-Hua Gao, and Xiaoxiao Li. Plug-and-play logit fusion for heterogeneous pathology foundation models. *arXiv preprint arXiv:2604.07779*, 2026.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2127–2136. PMLR, 2018.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 2019.

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pp. 554–561, 2013.

Karel Lenc and Andrea Vedaldi. Understanding image representations by measuring their equivariance and equivalence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 991–999, 2015.

Tim Lenz, Peter Neidlinger, Marta Ligero, Georg Wölflein, Marko van Treeck, and Jakob Nikolas Kather. Unsupervised foundation model-agnostic slide-level representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30807–30817, 2025. doi: 10.1109/CVPR52734.2025.02869.

Bin Li, Yin Li, and Kevin W. Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14313–14323, 2021. doi: 10.1109/CVPR46437.2021.01409.

Xitong Ling, Yuanyuan Lei, Jiawen Li, Junru Cheng, Wenting Huang, Tian Guan, Jian Guan, and Yonghong He. Comprehensive benchmark dataset for pathological lymph node metastasis in breast cancer sections. *Scientific Data*, 12:1381, 2025. doi: 10.1038/s41597-025-05586-5.

Mayug Maniparambil, Raiymbek Akshulakov, Yasser Abdelaziz Dahou Djilali, Sanath Narayan, Ankit Singh, and Noel E. O'Connor. Harnessing frozen unimodal encoders for flexible multimodal alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29847–29857, 2025. doi: 10.1109/CVPR52734.2025.02778.

Vaibhav Mishra and William Lotter. Comparing computational pathology foundation models using representational similarity analysis. In *Proceedings of the Fifth Machine Learning for Health Symposium*, volume 297 of *Proceedings of Machine Learning Research*, pp. 1007–1022. PMLR, 2026.

Peter Neidlinger, Omar S. M. El Nahhas, Hannah Sophie Muti, Tim Lenz, Michael Hoffmeister, Hermann Brenner, Marko van Treeck, Rupert Langer, Bastian Dislich, Hans Michael Behrens, Christoph Röcken, Sebastian Foersch, Daniel Truhn, Antonio Marra, Oliver Lester Saldanha, and Jakob Nikolas Kather. Benchmarking foundation models as feature extractors for weakly supervised computational pathology. *Nature Biomedical Engineering*, 10:1113–1123, 2026. doi: 10.1038/s41551-025-01516-3.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. 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. *Transactions on Machine Learning Research*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 8748–8763. PMLR, 2021.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6076–6085, 2017.

Mathieu Salzmann, Carl Henrik Ek, Raquel Urtasun, and Trevor Darrell. Factorized orthogonal latent spaces. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pp. 701–708. PMLR, 2010.

Peter H. Schoenemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31:1–10, 1966. doi: 10.1007/BF02289451.

Daniel Shao, Richard J. Chen, Andrew H. Song, Joel Runevic, Ming Y. Lu, Tong Ding, and Faisal Mahmood. Do multiple instance learning models transfer? In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 54219–54238. PMLR, 2025.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2136–2147, 2021.

The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61–70, 2012. doi: 10.1038/nature11412.

Zhilong Weng, Alexander Seper, Alexey Pryalukhin, Fabian Mairinger, Claudia Wickenhauser, Marcus Bauer, Lennert Glamann, Hendrik Bläker, Thomas Lingscheidt, Wolfgang Hulla, Danny Jonigk, Simon Schallenberg, Andrey Bychkov, Junya Fukuoka, Martin Braun, Birgid Schömig-Markiefka, Sebastian Klein, Andreas Thiel, Katarzyna Bozek, George J. Netto, Alexander Quaas, Reinhard Büttner, and Yuri Tolkach. GrandQC: A comprehensive solution to quality control problem in digital pathology. *Nature Communications*, 15:10685, 2024. doi: 10.1038/s41467-024-54769-y.

Yuxiang Xiao, Yang Hu, Bin Li, Tianyang Zhang, Zexi Li, Huazhu Fu, Jens Rittscher, and Kaixiang Yang. Understanding synergistic interactions among pathology foundation models via adaptive fusion. *arXiv preprint arXiv:2608.01370*, 2026.

Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

Andrew Zhang, Guillaume Jaume, Anurag Vaidya, Tong Ding, and Faisal Mahmood. Accelerating data processing and benchmarking of AI models for pathology. *arXiv preprint arXiv:2502.06750*, 2025.

Eric Zimmermann, Eugene Vorontsov, Julian Viret, Adam Casson, Michal Zelechowski, George Shaikovski, Neil Tenenholtz, James Hall, David Klimstra, Razik Yousfi, Thomas Fuchs, Nicolò Fusi, Siqi Liu, and Kristen Severson. Virchow2: Scaling self-supervised mixed magnification models in pathology. *arXiv preprint arXiv:2408.00738*, 2024.

A Additional Method Details

A.1 Closed-form semi-orthogonal alignment

Let $C = A_c^\top H_c \in \mathbb{R}^{d_a \times d_h}$. Expanding the objective in Eq. 2 gives

$$\|A_c R - H_c\|_F^2 = \|A_c\|_F^2 + \|H_c\|_F^2 - 2\,\mathrm{tr}(R^\top C). \tag{15}$$

Thus, the constrained problem maximizes $\mathrm{tr}(R^\top C)$. Consider the thin SVD

$$C = U\Sigma V^\top, \qquad U \in \mathbb{R}^{d_a \times d_a}, \quad \Sigma \in \mathbb{R}^{d_a \times d_a}, \quad V \in \mathbb{R}^{d_h \times d_a}. \tag{16}$$

This SVD retains d_a singular directions, including those with zero singular values when C is rank-deficient. For any feasible R, the matrix $U^\top R V$ is a contraction, so the trace inequality gives

$$\mathrm{tr}(R^\top C) = \mathrm{tr}\big(\Sigma V^\top R^\top U\big) \le \mathrm{tr}(\Sigma). \tag{17}$$

The upper bound is attained by

$$R = UV^\top, \qquad RR^\top = I_{d_a}, \tag{18}$$

which is therefore a closed-form minimizer. When C has full row rank, the polar factor $UV^\top$ is unique, including when positive singular values are repeated. Non-uniqueness arises when C is rank-deficient, in which case the action of R on the zero-singular-value subspace is not identified by the Procrustes objective.

The constraint also gives the geometric preservation used in the method. For any $u, v \in \mathbb{R}^{1 \times d_a}$,

$$\langle uR, vR\rangle = uRR^\top v^\top = \langle u, v\rangle, \qquad \|uR\|_2 = \|u\|_2, \qquad \|(u-v)R\|_2 = \|u - v\|_2. \tag{19}$$

Therefore, R changes the orientation of auxiliary coordinates in the anchor space while preserving their inner products, norms, and pairwise distances. The scale factor β in Eq. 7 subsequently adjusts the magnitude of the embedded representation.

A.2 Reachable decomposition and exact preservation

From $RR^\top = I_{d_a}$,

$$P_R^2 = R^\top (RR^\top) R = R^\top R = P_R, \qquad P_R^\top = P_R. \tag{20}$$

Moreover, $\mathrm{range}(P_R) = \mathrm{range}(R^\top)$, so P_R is the orthogonal projector onto $\mathcal{S}_R$. For the decomposition in Eq. 6,

$$H_\parallel = Z_H R, \qquad H_\perp = H_c - Z_H R. \tag{21}$$

The orthogonality relation follows directly from

$$H_\perp R^\top = H_c(I - P_R)R^\top = H_c\big(R^\top - R^\top R R^\top\big) = 0. \tag{22}$$

Consequently, every aligned auxiliary vector belongs to $\mathcal{S}_R$, while $H_\perp$ is unaffected by replacing Z_H with any other reachable coordinate matrix.

A.3 Minimum-change reconstruction

Let Z be any prescribed reachable-coordinate matrix and consider

$$\underset{G \in \mathbb{R}^{N \times d_h}}{\mathrm{argmin}} \ \|G - H_c\|_F^2 \quad \text{subject to} \quad GR^\top = Z. \tag{23}$$

The constraint fixes $GP_R = (GR^\top)R = ZR$, whereas $G(I - P_R)$ is free. Since P_R and $I - P_R$ are orthogonal projectors,

$$\|G - H_c\|_F^2 = \|GP_R - H_c P_R\|_F^2 + \|G(I - P_R) - H_c(I - P_R)\|_F^2. \tag{24}$$

The first term is fixed by the constraint and the second is uniquely minimized by $G(I - P_R) = H_c(I - P_R) = H_\perp$. Hence the unique solution is

$$G = H_\perp + ZR. \tag{25}$$

Setting $Z = Z_\alpha$ and restoring the anchor mean gives the PPFuse reconstruction in Eq. 11. The exact preservation property in Eq. 13 is an immediate consequence of this minimum-change solution.

The information-preserving reparameterization and its distinction from complement-scale preservation are given in Eq. 14 in Section 3.3.

Table 6: Complete dataset evaluation protocols. Sample counts reflect the experimental data used in this work.

Dataset	Domain	Task	Classes / labels	Samples	Protocol
BRACS	Pathology WSI	Fine-grained lesion classification	7	546 WSIs	Official split
CamelyonPlus	Pathology WSI	Lymph-node metastasis classification	4	1,350 WSIs	Patient-level 5-fold
TCGA-BRCA	Pathology WSI	PAM50 subtype classification	4	480 WSIs	Patient-level 5-fold
TCGA-NSCLC	Pathology WSI	LUAD vs. LUSC few-shot classification	2	936 WSIs	Patient-level 5-fold
Stanford Cars	Natural image	Fine-grained image classification	196	16,185 images	Official split
MM-IMDb	Image–text	Multi-label genre classification	23	25,959 pairs	Official split

A.4 PPFuse procedures

Algorithm 1 separates train-only fitting from application to a new paired collection. The fitted statistics define a fixed transform for subsequent training, validation, and test observations. Fitting assumes $d_a \le d_h$ and $\|A_c\|_F > 0$; ThinSVD uses the dimensions in Eq. 16.

Algorithm 1 PPFuse fitting and fusion.

procedure FitPPFuse($\mathcal{T}_{\text{fit}}, \mathcal{T}_{\text{tr}}$)
  Stack $\mathcal{T}_{\text{fit}}$ as H, A
  Compute μ_H, μ_A and $H_c \leftarrow H - \mathbf{1}_N \mu_H^\top$, $A_c \leftarrow A - \mathbf{1}_N \mu_A^\top$
  $U\Sigma V^\top \leftarrow \text{ThinSVD}(A_c^\top H_c)$; $R \leftarrow UV^\top$
  Stack $\mathcal{T}_{\text{tr}}$ as $H^{\text{tr}}, A^{\text{tr}}$ and center them using μ_H, μ_A
  $Z_H^{\text{tr}} \leftarrow H_c^{\text{tr}} R^\top$; $\beta \leftarrow \|Z_H^{\text{tr}}\|_F / \|A_c^{\text{tr}}\|_F$
  return $\Theta = (\mu_H, \mu_A, R, \beta)$
end procedure

procedure Fuse($H^{(q)}, A^{(q)}, \Theta, \alpha$)
  $(\mu_H, \mu_A, R, \beta) \leftarrow \Theta$
  $H_c^{(q)} \leftarrow H^{(q)} - \mathbf{1}_{N_q} \mu_H^\top$; $A_c^{(q)} \leftarrow A^{(q)} - \mathbf{1}_{N_q} \mu_A^\top$
  $Z_H^{(q)} \leftarrow H_c^{(q)} R^\top$; $H_\perp^{(q)} \leftarrow H_c^{(q)} - Z_H^{(q)} R$
  $Z_\alpha^{(q)} \leftarrow \alpha Z_H^{(q)} + (1-\alpha)\beta A_c^{(q)}$
  return $F_\alpha^{(q)} = \mathbf{1}_{N_q} \mu_H^\top + H_\perp^{(q)} + Z_\alpha^{(q)} R$
end procedure

B Dataset and Feature-Extraction Details

B.1 Dataset protocols

BRACS contains 547 slides in the official release, whereas 546 slides are used in our experiments. Specifically, `BRACS_1404.svs` in the official training split had exceptionally faint coloration;

the GrandQC tissue–background segmenter (Weng et al., 2024), as integrated in TRIDENT (Zhang et al., 2025), did not identify a tissue foreground for this slide. It was excluded before patch extraction; all remaining slides retain their official labels and split assignments.

BRACS uses the official train/validation/test split. CamelyonPlus, TCGA-BRCA, and TCGA-NSCLC use patient-level five-fold splits. CamelyonPlus has four labels: negative, isolated tumor cells, micrometastasis, and macrometastasis. The Normal-like PAM50 subtype is excluded from TCGA-BRCA because only eight samples are available, yielding the four-class task used in the paper. For TCGA-NSCLC, each fold contains nested training subsets of 4, 8, and 16 patients per class; validation and test partitions remain complete. Exact patient- and slide-level split manifests for all patient-level five-fold experiments, together with the TCGA-NSCLC few-shot patient-selection manifests, are provided in the supplementary material.

B.2 Pathology feature extraction

All pathology features were produced with the Mahmood Lab TRIDENT pipeline (Zhang et al., 2025). WSIs were tessellated at $20\times$ magnification into non-overlapping 256×256 patches. The same patch coordinates were retained across encoders, yielding exact same-patch correspondence. The main experiments use the publicly released UNI2-h checkpoint (`MahmoodLab/UNI2-h`) for the 1536-dimensional anchor representation and CONCHv1.5 (`MahmoodLab/conchv1_5`) for the 768-dimensional auxiliary representation. The representation-configuration study additionally uses Virchow2 (`paige-ai/Virchow2`), whose 2560-dimensional output is treated as two 1280-dimensional summaries,

$$2560 = 1280_{\mathrm{CLS}} \oplus 1280_{\mathrm{Mean}}. \tag{26}$$

For the constituent analyses, the 1536-dimensional UNI2 representation is the anchor and either CLS or Mean is the auxiliary, with an independent map and scale factor for each pairing. Full-descriptor fusion instead uses 2560-dimensional Virchow2 as anchor and UNI2 as auxiliary. The plotted balanced-accuracy deltas use the corresponding raw Virchow2 descriptor (CLS, Mean, or full) as the reference.

C MIL Training and Reproducibility

C.1 MIL implementation and optimization

The ABMIL, DSMIL, and TransMIL experiments use the public MIL-Lab implementation from the Mahmood Lab (Shao et al., 2025). All MIL models use the base configurations supplied by that implementation. Unless stated otherwise, a bag is processed with batch size one and models are trained with AdamW (learning rate 10^{-4}, weight decay 10^{-5}) for at most 50 epochs, with early stopping patience 10 and minimum validation improvement 0. The training runs are deterministic. The five BRACS runs use seeds 42–46. For full-data patient-level five-fold experiments, each fold uses seed 42 and variability is reported across folds. Unless otherwise stated, OP fitting uses patch-sampling seed 42 for the at-most-512 paired patches per training WSI; this seed is distinct from the MIL training and few-shot patient-selection seeds.

For TCGA-NSCLC few-shot experiments, the maximum number of epochs is 100 while the optimizer, learning rate, weight decay, batch size, patience, and deterministic setting are unchanged. The patient-sampling and aggregation protocol is detailed in Appendix E.

Computational resources. All downstream models were trained on a single NVIDIA GeForce RTX 4090 GPU (24 GB) per run. Pathology alignment maps were fitted on CPU.

C.2 Validation and checkpointing

MIL checkpoints maximize validation balanced accuracy for the multiclass WSI tasks and validation ROC-AUC for TCGA-NSCLC. Non-pathology checkpointing is specified in Appendix H. PPFuse uses the training-only fitting procedure in Section 3, with the patch sampling described above. We fix $\alpha = 0.5$ except in the sensitivity analysis.

Standard deviations are computed with $\mathrm{ddof} = 0$ over the corresponding runs or folds.

D Bootstrap Protocol and Main Comparisons

We report paired bootstrap intervals for the primary comparison,

$$\Delta\mathrm{BA} = \mathrm{BA}_{\mathrm{PPFuse}} - \mathrm{BA}_{\mathrm{baseline}},$$

where positive values favor PPFuse. All intervals use 10,000 percentile replicates and apply the same resampled units to PPFuse and its comparison baseline within each replicate. BRACS-7 uses patient-cluster resampling over the 31 unique patients in the official test set, with patient membership taken from the provided `BRACS.xlsx` metadata. Each replicate samples 31 patient IDs with replacement and includes all WSIs of every sampled patient with multiplicity. The same patient draw is shared by all methods and five training seeds; BA is computed separately per seed before paired differences are averaged across seeds. Sampling is unstratified because a patient may have WSIs from different classes; draws omitting any class are repeated. The resulting intervals quantify test-patient sampling uncertainty conditional on the five trained models. CamelyonPlus uses patient-cluster resampling within each test fold, which contains 116–119 patient clusters. Each CAMELYON16 file named `random_slide_<id>` forms a singleton patient cluster; CAMELYON17 files named `patient_<id>_node_⋆` are grouped by patient ID. Patients are sampled with replacement, retaining all slides within each sampled patient. Since a patient can have slides with different labels, sampling is unstratified; draws that omit any class are repeated. All methods share the same resampled patients. Metrics are computed separately within each fold and then averaged equally across the five folds. PAM50 uses class-stratified patient resampling within each test fold, also with equal averaging across five folds. Brackets denote unadjusted 95% bootstrap intervals. BRACS-7 comparisons shared with Table 11 use identical bootstrap replicates.

PPFuse has positive BA differences against RMS Concat in all nine settings. The intervals exclude zero for CamelyonPlus DSMIL and all three PAM50 backbones. Its differences against PCA RMS Concat are also positive in all nine settings, with eight intervals entirely above zero. Three intervals against CONCHv1.5 are entirely above zero, while every interval against UNI2 contains zero. The sole interval entirely below zero is against MLP Fusion on CamelyonPlus with TransMIL.

Table 8 reports PPFuse's absolute BA and Macro-F1 with intervals from the same patient-cluster bootstrap.

E Few-Shot Results

For each patient-level fold, five sampling seeds independently generate few-shot training subsets at 4, 8, and 16 patients per class. Within each sampling seed and fold, the subsets are nested, such that the 4-shot subset is contained in the 8-shot subset, which is in turn contained in the 16-shot subset. The MIL training seed is fixed to 42 for all runs so that the reported variation isolates sensitivity to the selected few-shot patients rather than model initialization. All WSIs from each selected patient enter the training subset; validation and test partitions retain all their WSIs. For each sampling seed, metrics are first averaged across the five patient-level folds. Table 9 reports the mean $\pm$ standard deviation across the resulting five sampling-seed means. For each sampling seed, fold, and shot count, PPFuse is fitted only on the selected training patients. The means and semi-orthogonal map use at most 512 paired patches per training WSI with patch sampling seed 42, and β uses all paired patches from the selected training patients. RMS Concat independently estimates its two scalar RMS factors from the selected few-shot training patients for each sampling seed, fold, and shot count, and then freezes them for validation and test data. Other training settings follow Appendix C.

F Fusion-Design and Mechanistic Controls

F.1 Scale Normalization for Concatenation

RMS Concat divides UNI2 and CONCHv1.5 by separate train-only scalar RMS values, $s_X = \sqrt{\sum_{i=1}^{N_{\mathrm{train}}} \sum_{j=1}^{d} X_{ij}^2/(N_{\mathrm{train}} d)}$, where $X \in \mathbb{R}^{N_{\mathrm{train}} \times d}$ is the training feature matrix for that view and $s_X > 0$, before concatenation. The scalars are then fixed for validation and test data. Thus, Raw Concat and RMS Concat have the same 2,304-dimensional feature interface, isolating the effect of global per-view scale on concatenation.

Table 7: Paired bootstrap balanced-accuracy deltas across datasets. Each cell reports PPFuse minus the indicated baseline above its unadjusted 95% percentile interval from 10,000 paired replicates. Values are on the BA scale of Table 2.

(a) Individual representations and Raw Concat

MIL	vs. UNI2	vs. CONCHv1.5	vs. Raw Concat
BRACS-7			
ABMIL	+0.0277 [−0.0414, 0.0960]	+0.0669 [−0.0009, 0.1345]	+0.0285 [−0.0148, 0.0701]
DSMIL	+0.0401 [−0.0097, 0.0730]	+0.1228 [0.0512, 0.1806]	+0.0168 [−0.0414, 0.0564]
TransMIL	+0.0242 [−0.0360, 0.0732]	+0.0406 [0.0002, 0.0815]	+0.0215 [−0.0283, 0.0743]
CamelyonPlus-4			
ABMIL	+0.01024 [−0.01975, 0.04264]	+0.00229 [−0.03238, 0.03587]	+0.02630 [−0.00132, 0.05739]
DSMIL	+0.02773 [−0.00881, 0.06681]	+0.01443 [−0.01719, 0.05404]	−0.01824 [−0.05546, 0.01939]
TransMIL	−0.00177 [−0.03620, 0.03538]	+0.01755 [−0.02783, 0.06400]	−0.00217 [−0.04071, 0.03911]
PAM50			
ABMIL	+0.0154 [−0.0233, 0.0536]	+0.0276 [−0.0133, 0.0679]	+0.0341 [−0.0062, 0.0750]
DSMIL	+0.0339 [−0.0086, 0.0756]	+0.0804 [0.0333, 0.1283]	+0.0475 [0.0041, 0.0891]
TransMIL	+0.0428 [−0.0053, 0.0897]	+0.0248 [−0.0222, 0.0708]	+0.0463 [−0.0006, 0.0919]

(b) Scale-normalized, PCA, and learned fusion baselines

MIL	vs. RMS Concat	vs. PCA RMS Concat	vs. MLP Fusion
BRACS-7			
ABMIL	+0.0431 [−0.0084, 0.1208]	+0.1063 [0.0161, 0.1872]	+0.0224 [−0.0086, 0.0528]
DSMIL	+0.0607 [−0.0100, 0.1276]	+0.2015 [0.0876, 0.2835]	+0.0196 [−0.0427, 0.0820]
TransMIL	+0.0233 [−0.0075, 0.0676]	+0.0980 [0.0343, 0.1667]	+0.0518 [0.0003, 0.0939]
CamelyonPlus-4			
ABMIL	+0.007152 [−0.031155, 0.045236]	+0.0206 [−0.0075, 0.0506]	+0.00886 [−0.02103, 0.04008]
DSMIL	+0.038500 [0.007230, 0.075595]	+0.0898 [0.0591, 0.1273]	+0.00236 [−0.03277, 0.04002]
TransMIL	+0.001863 [−0.038733, 0.044299]	+0.0891 [0.0518, 0.1298]	−0.04718 [−0.08893, −0.00562]
PAM50			
ABMIL	+0.0595 [0.0212, 0.0971]	+0.0482 [0.0036, 0.0922]	+0.0048 [−0.0354, 0.0456]
DSMIL	+0.0465 [0.0004, 0.0930]	+0.1037 [0.0600, 0.1494]	+0.0204 [−0.0230, 0.0621]
TransMIL	+0.0641 [0.0186, 0.1079]	+0.0814 [0.0354, 0.1267]	+0.0352 [−0.0123, 0.0812]

Table 8: PPFuse performance on CamelyonPlus with patient-cluster bootstrap intervals. Point estimates and unadjusted 95% percentile intervals are in percent; each replicate averages the five fold metrics equally.

MIL	BA (%)	95% CI	Macro-F1 (%)	95% CI
ABMIL	65.387	[62.074, 69.263]	65.428	[61.379, 68.800]
DSMIL	65.821	[62.900, 69.293]	64.865	[61.399, 67.718]
TransMIL	66.119	[62.837, 69.858]	65.841	[61.915, 69.003]

Table 9: Few-shot TCGA-NSCLC classification measured by ROC-AUC. Each shot denotes one selected training patient per class, with all of that patient's WSIs included. Entries are mean ± standard deviation across five sampling-seed means, each computed over five patient-level folds. Highest means for each MIL backbone and shot count are shown in bold.

Representation	4-shot	8-shot	16-shot
ABMIL			
UNI2	0.7779 ± 0.0440	0.8672 ± 0.0316	0.9218 ± 0.0304
CONCHv1.5	0.8372 ± 0.0199	0.9054 ± 0.0121	0.9402 ± 0.0087
Raw Concat	0.8246 ± 0.0501	0.8972 ± 0.0143	$\mathbf{0.9494 \pm 0.0110}$
RMS Concat	0.7700 ± 0.0552	0.8595 ± 0.0214	0.9341 ± 0.0231
MLP Fusion	0.8239 ± 0.0388	0.8993 ± 0.0197	0.9480 ± 0.0138
PPFuse	$\mathbf{0.8406 \pm 0.0347}$	$\mathbf{0.9056 \pm 0.0195}$	0.9475 ± 0.0160
DSMIL			
UNI2	0.6844 ± 0.0430	0.7465 ± 0.0316	0.8447 ± 0.0271
CONCHv1.5	0.6999 ± 0.0354	0.7606 ± 0.0208	0.8550 ± 0.0281
Raw Concat	0.6826 ± 0.0535	0.8008 ± 0.0613	0.8831 ± 0.0426
RMS Concat	0.5743 ± 0.0173	0.5965 ± 0.0222	0.7113 ± 0.0537
MLP Fusion	0.7778 ± 0.0272	0.8451 ± 0.0128	0.9236 ± 0.0165
PPFuse	$\mathbf{0.7790 \pm 0.0600}$	$\mathbf{0.8612 \pm 0.0229}$	$\mathbf{0.9257 \pm 0.0217}$
TransMIL			
UNI2	0.6967 ± 0.0390	0.7601 ± 0.0195	0.8506 ± 0.0141
CONCHv1.5	0.7130 ± 0.0407	0.8125 ± 0.0184	0.9001 ± 0.0135
Raw Concat	0.7453 ± 0.0405	0.8349 ± 0.0215	$\mathbf{0.9147 \pm 0.0141}$
RMS Concat	0.7058 ± 0.0387	0.7806 ± 0.0268	0.8783 ± 0.0251
MLP Fusion	0.7014 ± 0.0332	0.7905 ± 0.0282	0.9120 ± 0.0154
PPFuse	$\mathbf{0.7713 \pm 0.0382}$	$\mathbf{0.8533 \pm 0.0101}$	0.9133 ± 0.0139

F.2 Ridge CCA Concatenation

The pathology Ridge CCA Concat baseline uses the same paired training-patch sampling manifests as OP fitting, with at most 512 patches per training WSI. We fit one transform on the official BRACS training split and one per training fold for CamelyonPlus and PAM50, without labels. Let $X \in \mathbb{R}^{N\times 1536}$ and $Y \in \mathbb{R}^{N\times 768}$ denote centered UNI2 and CONCHv1.5 features. We estimate C_{XX}, C_{YY}, and C_{XY}, and apply ridge regularization

$$\lambda_X = 10^{-3}\frac{\mathrm{tr}(C_{XX})}{1536}, \qquad \lambda_Y = 10^{-3}\frac{\mathrm{tr}(C_{YY})}{768}. \tag{27}$$

Using the corresponding regularized whitening transforms B_X and B_Y, we compute

$$B_X^\top C_{XY} B_Y = U\Sigma V^\top \tag{28}$$

and form $W_X = B_X U_{:,1:768}$ and $W_Y = B_Y V_{:,1:768}$. Each patch is represented by the 1536-dimensional concatenation $[\,x_c W_X,\ y_c W_Y\,]$. All fitted statistics and projection matrices are held

Table 10: Feature constructions for the BRACS fusion-design controls. Reachable-only uses $\lambda = 0$; full-space OP interpolation uses $\lambda = \alpha = 0.5$. $L_\phi : \mathbb{R}^{768} \to \mathbb{R}^{1536}$ is an affine map, and M_θ in Eq. 29 has dimensions $768 \to 768 \to 1536$, with bias and no dropout.

Design	**Output dim.**	**Feature construction**
Raw Concat	2304	$X = \mathrm{concat}(U, C)$
Raw-Linear	1536	$X = \frac{1}{2}U + \frac{1}{2}L_\phi(C)$
MLP Fusion	1536	$X = \frac{1}{2}U + \frac{1}{2}M_\theta(C)$
OP Concat	3072	$X = \mathrm{concat}(U, C_{\mathrm{OP}})$
Reachable-only	1536	$F = \mathbf{1}\mu_U^\top + Z_\alpha R$
Full-space OP interp.	1536	$F = \mathbf{1}\mu_U^\top + 0.5H_\perp + Z_\alpha R$
PPFuse	1536	$F = \mathbf{1}\mu_U^\top + H_\perp + Z_\alpha R$

fixed for validation and test data. No additional feature scaling or canonical-correlation weighting is applied. This ridge-regularized pathology control is distinct from the thin-SVD/pseudoinverse CCA controls used for the non-pathology experiments in Appendix H.

F.3 RMS Concat + PCA

As a dimension-matched compression baseline, we first apply the same training-set RMS normalization used by RMS Concat to UNI2 and CONCHv1.5, concatenate the resulting features into a 2304-dimensional representation, and fit PCA using only training-set patches. PCA uses global training-set centering without whitening and retains the top 1536 components, matching the dimensionality of the UNI2 anchor. Across the 11 training partitions (one official BRACS-7 split and five patient-level folds each for CamelyonPlus and PAM50), the retained 1536 components explain 99.768–99.778% of the training-set variance (mean 99.775%).

F.4 Feature Constructions

For the BRACS controls in Table 3, let $U \in \mathbb{R}^{N\times 1536}$ and $C \in \mathbb{R}^{N\times 768}$ denote the raw UNI2 and CONCHv1.5 bags, respectively. For the alignment-based controls, the means, R, and β are estimated from paired training patches exactly as for PPFuse. In particular, $C_{\mathrm{OP}} = \mathbf{1}\mu_U^\top + \beta C_c R$ is the 1536-dimensional aligned auxiliary representation. The trainable maps L_ϕ and M_θ are learned jointly with the downstream MIL model; the remaining constructions use fixed transformations. Using the row-wise feature convention, the MLP is

$$M_\theta(C) = \mathrm{ReLU}(CW_1^\top + \mathbf{1}b_1^\top)W_2^\top + \mathbf{1}b_2^\top, \tag{29}$$

where $W_1 \in \mathbb{R}^{768\times 768}$, $W_2 \in \mathbb{R}^{1536\times 768}$, $b_1 \in \mathbb{R}^{768}$, and $b_2 \in \mathbb{R}^{1536}$.

F.5 BRACS Fusion Results

Table 11 reports the BRACS-7 fusion controls. The final column gives PPFuse-minus-baseline BA differences and intervals from the bootstrap protocol in Appendix D.

F.6 Complement Interventions on PAM50

We repeat the complement interventions on PAM50 with UNI2 as anchor and CONCHv1.5 as auxiliary, using the constructions in Table 10. Full-space OP interpolation scales the anchor complement by $\lambda = \alpha = 0.5$, whereas Reachable-only removes it ($\lambda = 0$); the reachable fusion is unchanged. Table 13 reports these results alongside the complete scale sweep.

Table 12 reports PPFuse-minus-control differences using the class-stratified patient bootstrap in Appendix D for both BA and Macro-F1. Differences and interval bounds are in percentage points.

PPFuse has higher mean BA and Macro-F1 than full-space OP interpolation with all three backbones, although all six intervals include zero. Relative to Reachable-only, both metrics have posi-

Table 11: BRACS-7 fusion-design controls. Metrics are mean ± standard deviation over seeds 42–46. The final column gives unadjusted 95% paired bootstrap intervals for PPFuse minus the listed baseline in balanced accuracy, using 10,000 patient-cluster replicates.

MIL	Design	BA	Macro-F1	ΔBA (PPFuse − row)
ABMIL	Raw Concat	0.4777 ± 0.0286	0.4577 ± 0.0622	$+0.0285\,[-0.0148,\ 0.0701]$
	Raw-Linear	0.4627 ± 0.0343	0.4187 ± 0.0403	$+0.0435\,[-0.0015,\ 0.0947]$
	MLP Fusion	0.4838 ± 0.0105	0.4513 ± 0.0167	$+0.0224\,[-0.0086,\ 0.0528]$
	OP Concat	0.4724 ± 0.0469	0.4516 ± 0.0514	$+0.0338\,[-0.0006,\ 0.0882]$
	Reachable-only	0.4406 ± 0.0638	0.4285 ± 0.0683	$+0.0656\,[0.0206,\ 0.1350]$
	Full-space OP interp.	0.4780 ± 0.0443	0.4680 ± 0.0513	$+0.0282\,[-0.0112,\ 0.0805]$
	PPFuse	0.5062 ± 0.0512	0.4963 ± 0.0518	—
DSMIL	Raw Concat	0.4691 ± 0.0355	0.4415 ± 0.0419	$+0.0168\,[-0.0414,\ 0.0564]$
	Raw-Linear	0.3970 ± 0.0315	0.3573 ± 0.0373	$+0.0890\,[0.0304,\ 0.1490]$
	MLP Fusion	0.4664 ± 0.0317	0.4411 ± 0.0210	$+0.0196\,[-0.0427,\ 0.0820]$
	OP Concat	0.4579 ± 0.0483	0.4401 ± 0.0455	$+0.0281\,[-0.0197,\ 0.0687]$
	Reachable-only	0.4408 ± 0.0279	0.4154 ± 0.0282	$+0.0451\,[-0.0010,\ 0.1093]$
	Full-space OP interp.	0.4599 ± 0.0153	0.4362 ± 0.0238	$+0.0261\,[-0.0237,\ 0.0827]$
	PPFuse	0.4859 ± 0.0334	0.4665 ± 0.0307	—
TransMIL	Raw Concat	0.4178 ± 0.0430	0.3842 ± 0.0420	$+0.0215\,[-0.0283,\ 0.0743]$
	Raw-Linear	0.4124 ± 0.0303	0.3789 ± 0.0328	$+0.0268\,[-0.0244,\ 0.0812]$
	MLP Fusion	0.3874 ± 0.0603	0.3449 ± 0.0613	$+0.0518\,[0.0003,\ 0.0939]$
	OP Concat	0.4206 ± 0.0442	0.3973 ± 0.0423	$+0.0187\,[-0.0310,\ 0.0592]$
	Reachable-only	0.4214 ± 0.0297	0.3919 ± 0.0390	$+0.0178\,[-0.0310,\ 0.0562]$
	Full-space OP interp.	0.4377 ± 0.0284	0.4248 ± 0.0242	$+0.0016\,[-0.0392,\ 0.0268]$
	PPFuse	0.4393 ± 0.0438	0.3998 ± 0.0625	—

Table 12: PAM50 complement-intervention differences ($\mathrm{PPFuse} - \mathrm{control}$) and unadjusted 95% percentile intervals from 10,000 paired bootstrap replicates with class-stratified patient resampling within each test fold and equal fold averaging. Deltas and interval bounds are in percentage points.

Control	MIL	ΔBA	95% CI	ΔMacro-F1	95% CI
Full-space OP interp.	ABMIL	$+0.804$	$[-1.781, 3.283]$	$+1.285$	$[-1.534, 4.090]$
	DSMIL	$+2.998$	$[-0.590, 6.630]$	$+2.439$	$[-1.377, 6.234]$
	TransMIL	$+4.404$	$[-0.836, 9.633]$	$+4.996$	$[-0.281, 10.418]$
Reachable-only	ABMIL	-0.183	$[-3.085, 2.678]$	$+0.469$	$[-2.543, 3.536]$
	DSMIL	$+2.449$	$[-0.951, 5.878]$	$+1.943$	$[-1.531, 5.516]$
	TransMIL	$+6.486$	$[2.042, 11.189]$	$+6.389$	$[1.842, 11.402]$

tive differences with DSMIL and TransMIL; the TransMIL intervals exclude zero for BA ($+6.486$ points, $[2.042, 11.189]$) and Macro-F1 ($+6.389$ points, $[1.842, 11.402]$). With ABMIL, the BA difference is slightly negative and the Macro-F1 difference is positive, with both intervals including zero. The PAM50 results extend the complement analysis beyond BRACS and show that the benefit of retaining the complement varies with the MIL backbone.

F.7 Complement-Scale Sweep

We evaluate Eq. 9 on BRACS-7 and PAM50 using UNI2 as anchor, CONCHv1.5 as auxiliary, $\alpha = 0.5$, and $\lambda \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2\}$. Only the coefficient of $H_\perp$ is varied; the output dimensionality and all training-only statistics (μ_H, μ_A, R, β) remain fixed across λ. All MIL training settings and data partitions follow Appendix C. BRACS uses seeds 42–46 on the official

Table 13: Complete complement-scale sweep at $\alpha = 0.5$. BA and Macro-F1 are mean $\pm$ standard deviation in percent over five training seeds for BRACS-7 and five patient-level folds for PAM50. All outputs are 1536-dimensional.

	ABMIL		**DSMIL**		**TransMIL**	
λ	**BA**	**Macro-F1**	**BA**	**Macro-F1**	**BA**	**Macro-F1**
BRACS-7						
0	44.06 ± 6.38	42.85 ± 6.83	44.08 ± 2.79	41.54 ± 2.82	42.14 ± 2.97	39.19 ± 3.90
0.25	44.10 ± 2.82	41.51 ± 2.14	44.52 ± 3.73	41.32 ± 5.68	43.02 ± 4.20	39.17 ± 4.98
0.5	47.80 ± 4.43	46.80 ± 5.13	45.99 ± 1.53	43.62 ± 2.38	43.77 ± 2.84	42.48 ± 2.42
0.75	46.98 ± 7.13	45.16 ± 7.01	43.63 ± 7.74	42.62 ± 8.45	41.53 ± 2.90	39.37 ± 4.56
1 (PPFuse)	50.62 ± 5.12	49.63 ± 5.18	48.59 ± 3.34	46.65 ± 3.07	43.93 ± 4.38	39.98 ± 6.25
1.25	47.73 ± 3.40	45.42 ± 3.15	43.51 ± 4.27	41.16 ± 3.57	43.97 ± 5.25	41.73 ± 5.81
1.5	50.09 ± 4.56	48.31 ± 4.13	47.27 ± 3.82	44.77 ± 4.88	43.00 ± 2.40	39.35 ± 2.10
2	45.74 ± 5.75	44.72 ± 5.51	42.77 ± 1.96	40.18 ± 1.26	44.93 ± 1.90	41.79 ± 3.72
PAM50						
0	57.65 ± 3.06	56.32 ± 3.44	54.68 ± 2.50	54.28 ± 2.50	49.36 ± 2.02	48.55 ± 1.58
0.25	56.51 ± 3.57	55.51 ± 3.81	55.14 ± 3.26	54.28 ± 2.50	53.87 ± 4.57	52.84 ± 5.49
0.5	56.67 ± 4.05	55.51 ± 4.23	54.13 ± 2.38	53.79 ± 3.18	51.44 ± 5.76	49.94 ± 6.13
0.75	57.21 ± 3.82	56.26 ± 4.01	53.68 ± 3.09	52.66 ± 3.00	52.73 ± 4.35	52.22 ± 4.85
1 (PPFuse)	57.47 ± 4.38	56.79 ± 3.59	57.13 ± 4.01	56.23 ± 2.26	55.85 ± 4.34	54.94 ± 5.66
1.25	58.69 ± 2.23	58.74 ± 1.97	55.39 ± 1.37	54.32 ± 2.07	51.25 ± 3.76	50.97 ± 3.97
1.5	58.40 ± 3.46	57.92 ± 3.03	53.39 ± 1.62	52.94 ± 1.79	53.07 ± 2.65	52.93 ± 2.16
2	56.43 ± 1.62	55.77 ± 1.21	52.86 ± 4.11	51.54 ± 5.80	52.11 ± 3.84	51.85 ± 4.03

split, while PAM50 uses five patient-level folds with seed 42; results are reported as mean $\pm$ standard deviation with $\mathrm{ddof} = 0$.

The $\lambda = 0, 0.5, 1$ settings correspond to Reachable-only, full-space OP interpolation, and PPFuse, respectively, and reuse the existing results. The remaining scales are newly evaluated under the same protocol. PPFuse itself always uses $\lambda = 1$; other values are diagnostic interventions. Table 13 reports the full BA and Macro-F1 results.

F.8 Reachable and Orthogonal Component Diagnostics

We evaluate the two anchor components separately to examine the predictive information retained by the complement. These BRACS-7 diagnostics use ABMIL, the training-only OP fit, and the same five training seeds as the main experiments. Let $H_c \in \mathbb{R}^{N \times 1536}$ denote centered UNI2 patch features and $R \in \mathbb{R}^{768 \times 1536}$ the fitted map, with $RR^\top = I_{768}$. Choose a row-orthonormal matrix $S \in \mathbb{R}^{768 \times 1536}$ spanning the orthogonal complement of the rows of R, so that $SR^\top = 0$. The two diagnostic inputs are

$$D_{\mathrm{reach}} = H_c R^\top, \qquad \text{(UNI2 reachable)},$$
$$D_\perp = \gamma_p H_c S^\top, \qquad \text{(UNI2 complement)}.$$

Both inputs are 768-dimensional patch bags evaluated with the same ABMIL interface. The matrix S expresses the anchor complement in orthonormal coordinates; γ_p is obtained by global train-set scale matching and held fixed for evaluation. Table 14 reports the predictive performance of each component.

F.9 Orientation-Control Evaluation

This section specifies the geometry diagnostics in Table 5. In the UNI2–CONCHv1.5 setting, the OP map and train-set means (μ_H, μ_A) are fitted using only paired training patches. Metrics are evaluated on independently sampled paired validation patches; each validation WSI s contributes $m_s \leq 256$ patches, with the same patch indices retained across encoders.

Table 14: Component-level diagnostic inputs on BRACS-7 with ABMIL. All inputs are 768-dimensional patch bags; entries are means ± standard deviations over five training seeds.

Diagnostic input	BA	Macro-F1	Accuracy	Macro OVR AUC
UNI2 reachable	0.4542 ± 0.0416	0.4237 ± 0.0267	0.5080 ± 0.0312	0.8390 ± 0.0135
UNI2 complement	0.2520 ± 0.0503	0.2231 ± 0.0526	0.3195 ± 0.0531	0.6409 ± 0.0636

Random-Q construction. For each control, we draw $G \in \mathbb{R}^{1536\times768}$ with independent standard-normal entries, compute a reduced QR factorization $G = \widetilde{Q}T$ with signs adjusted so that $\mathrm{diag}(T)$ is nonnegative, and set $Q = \widetilde{Q}^\top \in \mathbb{R}^{768\times1536}$. This construction is independent of the paired correspondence and gives $QQ^\top = I_{768}$. We substitute Q for R throughout PPFuse. The map induces

$$P_Q = Q^\top Q, \qquad H_\perp^Q = H_c(I - P_Q), \qquad Z_H^Q = H_c Q^\top.$$

Using the same training-only global means (μ_H, μ_A) as the OP control, we recompute the scale factor according to Eq. 7:

$$\beta_Q = \frac{\|H_c Q^\top\|_F}{\|A_c\|_F}.$$

The resulting representation is

$$F_{\alpha,Q} = H + (1-\alpha)\big(\beta_Q A_c - H_c Q^\top\big)Q.$$

Thus, Random-Q retains the semi-orthogonal geometry and fusion rule while removing the correspondence-fitted orientation; the induced projector, complement, reachable coordinates, and scale are all recomputed for each sampled Q.

Let $a_{si} \in \mathbb{R}^{1\times768}$ be a raw CONCHv1.5 feature and $h_{sj} \in \mathbb{R}^{1\times1536}$ a raw UNI2 feature. We use train-global centering and define

$$z_{si} = (a_{si} - \mu_A^\top)R, \qquad \tilde{h}_{sj} = h_{sj} - \mu_H^\top.$$

For query z_{si}, the candidate set contains all m_s centered UNI2 patches from the same WSI. The unique positive is $\tilde{h}_{si}$; the other $\tilde{h}_{sj}$, $j \neq i$, are negatives. Ranking these within-WSI candidates by decreasing cosine similarity gives

$$\mathrm{rank}_{si} = 1 + \#\{j : \cos(z_{si}, \tilde{h}_{sj}) > \cos(z_{si}, \tilde{h}_{si})\}, \qquad \mathrm{R@}k = \frac{1}{\sum_s m_s}\sum_{s,i} \mathbf{1}[\mathrm{rank}_{si} \le k].$$

Cosine similarities are defined for nonzero feature vectors. Under uniformly random rankings, the patch-micro chance R@1 is $S_{\mathrm{val}}/\sum_s m_s$, where S_{val} is the number of evaluated slides. This is $1/256 \approx 0.391\%$ when each slide contributes 256 candidates.

The same-patch cosine is $c_{si}^+ = \cos(z_{si}, \tilde{h}_{si})$. For $m_s > 1$, its negative counterpart averages all off-diagonal within-WSI candidates,

$$c_{si}^- = \frac{1}{m_s - 1}\sum_{j\neq i} \cos(z_{si}, \tilde{h}_{sj}), \qquad \mathrm{margin}_{si} = c_{si}^+ - c_{si}^-.$$

Same-patch cosine, negative cosine, margin, and retrieval are pooled patch-micro statistics over validation slides. All features are centered using the global training means. The residual similarly uses the unscaled centered OP output (i.e., before β):

$$\rho = \frac{\|A_c R - H_c\|_F}{\|H_c\|_F} = \sqrt{\frac{\sum_{s,i}\|z_{si} - \tilde{h}_{si}\|_2^2}{\sum_{s,i}\|\tilde{h}_{si}\|_2^2}}.$$

The residual is defined when $\|H_c\|_F > 0$ for these validation features. The OP column uses one correspondence-fitted map. The Random-Q column reports mean ± standard deviation over five fixed random semi-orthogonal orientations, evaluated with the same validation protocol.

Table 15 reports the downstream BA of each fixed Random-Q orientation in Figure 2(c) under the identity and linear adapters. Identity supplies the 1536-dimensional fused features directly to AB-MIL, whereas Linear applies a learned $1536 \to 512$ projection before aggregation.

Table 15: BRACS-7 ABMIL balanced accuracy for PPFuse and the five fixed Random-Q orientations under identity and linear downstream adapters. Each orientation is trained with seeds 42–46; entries for individual orientations report mean ± standard deviation over these five training runs. Random-Q 1–5 correspond to independently sampled semi-orthogonal maps generated with seeds 20260811–20260815. The final row reports the mean ± standard deviation across the five Random-Q mean BAs.

Orientation	**Identity**	**Linear**
PPFuse	0.4352 ± 0.0437	0.4642 ± 0.0481
Random-Q 1	0.3962 ± 0.0758	0.4500 ± 0.0303
Random-Q 2	0.3671 ± 0.0999	0.4561 ± 0.0206
Random-Q 3	0.3449 ± 0.1043	0.4417 ± 0.0272
Random-Q 4	0.3584 ± 0.0949	0.4202 ± 0.0286
Random-Q 5	0.3430 ± 0.0800	0.3950 ± 0.0219
Random-Q mean	0.3619 ± 0.0193	0.4326 ± 0.0224

G Complete Virchow2 Representation Results

Tables 16 and 17 give the absolute test results underlying the representation-granularity comparison in Section 4.5. Raw UNI2 is included as the shared reference. Raw Virchow2 concatenates its 1280-dimensional CLS and 1280-dimensional mean-patch summaries; Raw CLS and Raw Mean use each constituent alone. PPFuse Virchow2 uses the full 2560-dimensional Virchow2 anchor with UNI2 as auxiliary. PPFuse CLS and PPFuse Mean use UNI2 as anchor with the corresponding 1280-dimensional constituent as auxiliary. All PPFuse configurations use $\alpha = 0.5$. Thus, the full-descriptor and constituent comparisons also differ in anchor choice and output dimension.

We report means and population standard deviations ($\mathrm{ddof} = 0$) over seeds 42–46 on the official BRACS-7 split and over five patient-level CamelyonPlus folds (seed 42). Each configuration uses ABMIL, DSMIL, and TransMIL with the training and checkpoint-selection protocol in Appendix C. The raw UNI2 rows reproduce the baseline runs in Table 2.

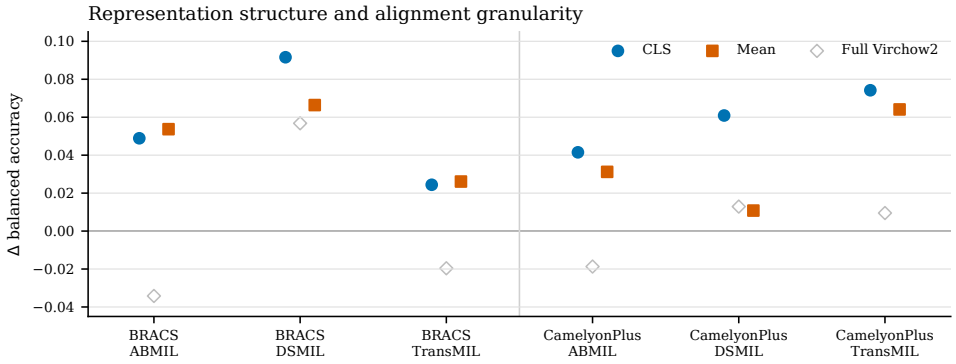


Figure 3: PPFuse BA minus that of the corresponding raw Virchow2 descriptor (CLS, Mean, or full). Constituent configurations have UNI2 as anchor; pale diamonds denote full-descriptor configurations with Virchow2 as anchor. All 12 constituent configurations improve over their corresponding raw constituents; nine also achieve higher mean BA than the stronger of raw UNI2 and the corresponding raw constituent. The full-descriptor results are mixed on BRACS and CamelyonPlus.

Table 16: Complete BRACS-7 results for UNI2, Virchow2, and PPFuse. Entries are test means ± standard deviations over five training seeds. Output dimensions are 1536 (Raw UNI2, PPFuse CLS/Mean), 2560 (Raw/PPFuse Virchow2), and 1280 (Raw CLS/Mean).

MIL	Representation	BA	Macro-F1
ABMIL	Raw UNI2	0.4786 ± 0.0407	0.4611 ± 0.0438
	Raw Virchow2	0.4554 ± 0.0361	0.4226 ± 0.0545
	Raw CLS	0.4335 ± 0.0357	0.3976 ± 0.0444
	Raw Mean	0.4060 ± 0.0187	0.3725 ± 0.0198
	PPFuse Virchow2	0.4212 ± 0.0347	0.3927 ± 0.0192
	PPFuse CLS	0.4824 ± 0.0271	0.4406 ± 0.0312
	PPFuse Mean	0.4597 ± 0.0391	0.4254 ± 0.0418
DSMIL	Raw UNI2	0.4459 ± 0.0351	0.4321 ± 0.0355
	Raw Virchow2	0.3571 ± 0.0475	0.3359 ± 0.0573
	Raw CLS	0.3660 ± 0.0298	0.3518 ± 0.0361
	Raw Mean	0.3720 ± 0.0144	0.3459 ± 0.0218
	PPFuse Virchow2	0.4139 ± 0.0371	0.3957 ± 0.0366
	PPFuse CLS	0.4576 ± 0.0577	0.4158 ± 0.0712
	PPFuse Mean	0.4384 ± 0.0378	0.4013 ± 0.0294
TransMIL	Raw UNI2	0.4151 ± 0.0389	0.4084 ± 0.0448
	Raw Virchow2	0.3899 ± 0.0471	0.3589 ± 0.0475
	Raw CLS	0.3708 ± 0.0190	0.3532 ± 0.0198
	Raw Mean	0.4150 ± 0.0285	0.3928 ± 0.0375
	PPFuse Virchow2	0.3703 ± 0.0554	0.3396 ± 0.0689
	PPFuse CLS	0.3952 ± 0.0247	0.3697 ± 0.0355
	PPFuse Mean	0.4411 ± 0.0424	0.4114 ± 0.0414

H Additional Evaluation Beyond Pathology

We use frozen DINOv2 ViT-B/14 (DINOv2-B) and OpenAI CLIP ViT-B/16 (Oquab et al., 2024; Radford et al., 2021). DINOv2-B uses the released `dinov2_vitb14_pretrain.pth` checkpoint and produces a 768-dimensional CLS representation. CLIP uses the official `ViT-B-16.pt` checkpoint and produces 512-dimensional image or text representations. On Stanford Cars, both encoders process the same source image using their respective preprocessing pipelines. On MM-IMDb, DINOv2-B encodes the poster, while CLIP encodes the first plot synopsis. Synopses exceeding CLIP's context length are split into chunks containing at most 75 content tokens, and the resulting text embeddings are averaged with equal weight.

H.1 Classifier and Evaluation Protocol

The downstream classifier is $\mathrm{Linear}(d, w) \to \mathrm{ReLU} \to \mathrm{Linear}(w, C)$, where d is the input dimension, C is the number of classes or labels, and $w = 512$ for all methods. We train with AdamW, an initial learning rate of 10^{-3}, weight decay 10^{-4}, batch size 256, and cosine annealing over 100 epochs. Both datasets use classifier training seeds 42–46. For each dataset, PPFuse uses $\alpha = 0.5$ and is fitted once from all paired training representations; the resulting fusion statistics are shared across the five classifier seeds. Tables 18 and 19 report means and standard deviations across these seeds.

H.2 Fusion Controls

Raw Concat concatenates the two encoder outputs. RMS Concat first divides each view by its global training-set scalar RMS, as defined in Appendix F.1, and concatenates the normalized features.

The CCA controls center the paired training representations and use thin-SVD/pseudoinverse whitening. CCA-k concat concatenates the first k canonical coordinates from each view, producing

Table 17: Complete CamelyonPlus results for UNI2, Virchow2, and PPFuse. Entries are test means ± standard deviations over five patient-level folds. Output dimensions are 1536 (Raw UNI2, PPFuse CLS/Mean), 2560 (Raw/PPFuse Virchow2), and 1280 (Raw CLS/Mean).

MIL	Representation	BA	Macro-F1
ABMIL	Raw UNI2	0.6436 ± 0.0221	0.6447 ± 0.0259
	Raw Virchow2	0.6537 ± 0.0347	0.6411 ± 0.0239
	Raw CLS	0.6293 ± 0.0273	0.6297 ± 0.0298
	Raw Mean	0.6335 ± 0.0100	0.6333 ± 0.0173
	PPFuse Virchow2	0.6350 ± 0.0306	0.6315 ± 0.0283
	PPFuse CLS	0.6708 ± 0.0162	0.6745 ± 0.0234
	PPFuse Mean	0.6647 ± 0.0236	0.6690 ± 0.0103
DSMIL	Raw UNI2	0.6305 ± 0.0121	0.6324 ± 0.0076
	Raw Virchow2	0.6386 ± 0.0383	0.6365 ± 0.0470
	Raw CLS	0.5964 ± 0.0218	0.5920 ± 0.0214
	Raw Mean	0.6665 ± 0.0229	0.6696 ± 0.0179
	PPFuse Virchow2	0.6515 ± 0.0349	0.6473 ± 0.0337
	PPFuse CLS	0.6573 ± 0.0324	0.6569 ± 0.0326
	PPFuse Mean	0.6773 ± 0.0458	0.6767 ± 0.0448
TransMIL	Raw UNI2	0.6630 ± 0.0341	0.6527 ± 0.0369
	Raw Virchow2	0.6600 ± 0.0531	0.6407 ± 0.0428
	Raw CLS	0.6311 ± 0.0284	0.6157 ± 0.0292
	Raw Mean	0.6287 ± 0.0168	0.6306 ± 0.0181
	PPFuse Virchow2	0.6695 ± 0.0194	0.6589 ± 0.0367
	PPFuse CLS	0.7053 ± 0.0344	0.6963 ± 0.0389
	PPFuse Mean	0.6928 ± 0.0168	0.7017 ± 0.0243

$2k$ dimensions. All fusion statistics are fitted on training representations and then held fixed for validation and test observations.

H.3 Stanford Cars

Stanford Cars uses its official train/test split, with a fixed stratified 20% subset of the official training set reserved for validation using seed 42, leaving 6,515 images for classifier training. Validation accuracy selects the checkpoint.

PPFuse exceeds Raw Concat and both CCA concatenation controls in all reported metrics (Table 18). It ranks second in BA, 0.457 percentage points below RMS Concat.

Table 18: Stanford Cars with the 512-unit MLP classifier and AdamW. Entries are mean ± standard deviation. Bold denotes the highest mean in each metric.

Representation	Dim.	BA	Macro-F1	Accuracy
DINOv2-B	768	0.85098 ± 0.00216	0.85052 ± 0.00221	0.85273 ± 0.00215
CLIP image	512	0.85140 ± 0.00146	0.85079 ± 0.00159	0.85173 ± 0.00142
Raw Concat	1280	0.88411 ± 0.00090	0.88346 ± 0.00110	0.88586 ± 0.00097
RMS Concat	1280	$\mathbf{0.89561 \pm 0.00107}$	$\mathbf{0.89522 \pm 0.00104}$	$\mathbf{0.89735 \pm 0.00119}$
CCA-384 concat	768	0.88151 ± 0.00175	0.88224 ± 0.00163	0.88337 ± 0.00163
CCA-512 concat	1024	0.88625 ± 0.00061	0.88666 ± 0.00058	0.88795 ± 0.00057
PPFuse	768	0.89104 ± 0.00113	0.89088 ± 0.00108	0.89272 ± 0.00107

H.4 MM-IMDB

MM-IMDb uses the official split of 15,552/2,608/7,799 training/validation/test pairs with 23 genre labels. Its classifier is trained with binary cross-entropy with logits (`BCEWithLogitsLoss`). A fixed probability threshold of 0.5 converts sigmoid outputs to label predictions, and validation Macro-F1 selects the checkpoint. Macro-AP and Macro-AUC use continuous prediction scores.

PPFuse exceeds Raw Concat and both CCA concatenation controls in all reported metrics (Table 19). It ranks second in Macro-F1, 0.946 percentage points below RMS Concat.

Table 19: MM-IMDb with the 512-unit MLP classifier and AdamW on the official split. Entries are mean $\pm$ standard deviation. Bold denotes the highest mean in each metric across the listed methods.

Representation	**Dim.**	**Macro-F1**	**Macro-AP**	**Macro-AUC**
DINOv2-B	768	0.36517 ± 0.00408	0.36741 ± 0.01070	0.78310 ± 0.00766
CLIP text	512	0.54130 ± 0.00163	0.60218 ± 0.00723	0.90638 ± 0.00335
Raw Concat	1280	0.53988 ± 0.00321	0.57993 ± 0.00643	0.89432 ± 0.00372
RMS Concat	1280	$\mathbf{0.56745 \pm 0.00283}$	$\mathbf{0.62232 \pm 0.00601}$	$\mathbf{0.91217 \pm 0.00347}$
CCA-384 concat	768	0.53321 ± 0.00183	0.58721 ± 0.00647	0.89315 ± 0.00318
CCA-512 concat	1024	0.53614 ± 0.00213	0.59595 ± 0.00298	0.89581 ± 0.00163
PPFuse	768	0.55799 ± 0.00333	0.60741 ± 0.00861	0.90515 ± 0.00446