GRASP: GRAPH-REFINED ANCHORED SOFT PREDICTIONS FOR TRAINING-FREE VISION-LANGUAGE MODEL ADAPTATION ACROSS STREAM STRUCTURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Training-free test-time adaptation of vision-language models must operate on test streams whose local structure can vary substantially in deployment. Samples observed together may share useful temporal or semantic context, yet heterogeneous batches can make the same relational evidence unreliable. This creates a central challenge for relational adaptation: exploiting local structure without assuming that it is always informative. We demonstrate this failure mode with a simple batch-local graph refinement, which performs strongly on class-correlated streams but degrades sharply under IID ordering.

We introduce **GRASP** (**G**raph-**R**efined **A**nchored **S**oft **P**redictions), which retains this local relational structure while controlling its influence. GRASP anchors graph refinement to a cache-refined posterior through a fidelity-regularized objective and attenuates edges whose endpoint posteriors disagree. The graph is restricted to the current batch and discarded after inference; only the bounded cache persists across the test stream. GRASP requires no target labels, gradient updates, prompt optimization, or knowledge of the stream regime.

Across 11 image-classification benchmarks and five stream regimes, GRASP achieves the highest cross-regime mean among evaluated methods at 71.64%. Relative to TDA, which provides its default prediction anchor, GRASP trails by 0.82 percentage points under IID ordering but improves by 8.40 points under class-sequential ordering. These results show that local relational evidence can provide substantial test-time gains while limiting degradation under weak local structure.

1 INTRODUCTION

Vision-language models (VLMs) such as CLIP (Radford et al., 2021) provide strong zero-shot recognition, but their performance can deteriorate under distribution shift (Shu et al., 2022). Test-time adaptation (TTA) uses unlabeled target data available at inference to improve predictions on the shifted distribution. Training-free TTA is particularly attractive for vision-language models because it leaves the pretrained encoder unchanged and avoids gradient-based optimization.

Existing training-free methods exploit different forms of test-time evidence. Cache-based approaches retain information from observed test samples (Karmanov et al., 2024), while statistical methods estimate properties of the target distribution online (Zanella et al., 2025). Relational methods instead exploit structure among test samples, often by connecting nearby samples in feature space and propagating information across these neighborhoods. ECALP (Li et al., 2025), for example, performs label propagation over a dynamically expanding graph containing textual class prototypes and observed test samples.

In deployment, however, the amount of local structure in the incoming stream can vary: temporally or semantically related samples may occur together, while other batches may be heterogeneous. A useful relational method should therefore exploit such structure when present without depending on it for robust prediction. The challenge is that feature proximity does not necessarily imply that two samples should reinforce one another. In locally coherent batches, neighboring samples can provide useful relational evidence; in heterogeneous batches, similar-looking samples may instead support different predictions, making relational reliability depend on more than feature similarity.

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.

To study this trade-off directly, we construct **GRASP-Base**, a deliberately minimal batch-local relational baseline that propagates cache-refined predictions over a cosine k-nearest-neighbor (kNN) graph. GRASP-Base is highly effective when local relational structure is strong, reaching 77.08% average accuracy under class-sequential ordering, but falls to 52.60% under IID ordering. On individual datasets, IID propagation can fall more than 22 percentage points below zero-shot CLIP. Together, these results show that local relational evidence can provide substantial gains, but naive propagation makes performance strongly dependent on stream composition.

This leads to the question we study in this work: *how can local test-time relationships be exploited when their reliability varies with an unknown stream structure?* We investigate this question under five test-stream regimes: IID ordering, three Dirichlet-generated settings from StatA (Zanella et al., 2025) with $\gamma \in \{0.1, 0.01, 0.001\}$, and the class-sequential *Separate* setting. Together, these protocols span increasing levels of local class structure available to relational inference.

We introduce **GRASP** (**G**raph-**R**efined **A**nchored **S**oft **P**redictions), which retains the batch-local relational structure of GRASP-Base while controlling its influence. GRASP starts from a cache-refined posterior and treats it as the reference for graph refinement. A fidelity term penalizes unnecessary deviation from this prediction, while Jensen-Shannon divergence between cache-refined posteriors attenuates feature-space edges with posterior disagreement. Relational information can therefore modify the cache prediction without letting feature proximity alone dominate it.

The resulting design is deliberately local. Unlike ECALP, GRASP does not maintain an expanding graph over previously observed test samples or use textual prototypes as graph nodes. The graph is constructed only for the current batch and discarded after inference; only the bounded test-time cache persists. GRASP therefore uses a more restricted relational state while retaining the ability to exploit short-range structure when it is informative.

A single GRASP configuration is used across all datasets and stream regimes. Across 11 benchmarks, GRASP achieves the highest cross-regime mean of 71.64%. Relative to its default TDA anchor, it trails by 0.82 points under IID ordering but improves by 8.40 under class-sequential ordering. The unchanged relational operator shows the same structure-dependent trend with two alternative upstream adapters. Overall, GRASP retains the gains of local propagation when relational structure is informative while avoiding its severe IID degradation. GRASP builds on a standard fidelity-smoothness objective. Its distinguishing design combines a cache-refined prediction anchor, posterior-consistent affinities, and a batch-local graph to address variation in test-stream structure.

Contributions

1. We characterize the dependence of local relational adaptation on test-stream structure. GRASP-Base, a minimal batch-local graph refinement, yields large gains on class-correlated streams but degrades sharply under IID ordering.
2. We introduce **GRASP** (Graph-Refined Anchored Soft Predictions), which controls relational influence through cache-anchored diffusion and posterior-consistent edge attenuation.
3. We analyze the refinement operator, derive a bound on graph-induced displacement from the cache prediction, and relate GRASP to conventional label propagation. Edge diagnostics and a matched no-JS comparison further characterize posterior-consistent attenuation.
4. We evaluate GRASP on 11 image-classification benchmarks across five stream regimes using a single fixed configuration, and show that the unchanged relational operator transfers to two alternative upstream adapters without retuning.

2 Background

Zero-shot vision-language inference CLIP (Radford et al., 2021) learns aligned image and text representations through contrastive pre-training. At inference, a frozen image encoder maps an input $\mathbf{x}$ to an ℓ_2-normalized feature $\mathbf{z} \in \mathbb{S}^{d-1}$, while class c is represented by a text embedding t_c. The zero-shot logit for class c is

$$s_c^{\mathrm{ZS}}(\mathbf{z}) = \tau\, \mathbf{z}^\top t_c, \tag{1}$$

where τ is the logit scale.

Cache-based adaptation TDA (Karmanov et al., 2024) performs training-free adaptation using online positive and negative caches constructed from test samples. Cache retrieval augments the zero-shot logits as

$$\mathbf{s}^{\text{cache}}(\mathbf{z}) = \mathbf{s}^{\text{ZS}}(\mathbf{z}) + \beta^{+}\mathbf{s}^{\mathcal{C}^{+}}(\mathbf{z}) + \beta^{-}\mathbf{s}^{\mathcal{C}^{-}}(\mathbf{z}), \tag{2}$$

yielding the cache-refined posterior

$$\mathbf{p}_{\text{cache}}(\mathbf{z}) = \text{softmax}\big(\mathbf{s}^{\text{cache}}(\mathbf{z})\big) \in \mathbb{R}^{C}. \tag{3}$$

The cache persists across the test stream. In GRASP, these cache-refined posteriors provide the predictions available before relational refinement.

Statistical adaptation StatA (Zanella et al., 2025) formulates training-free VLM adaptation through regularized probabilistic clustering of target features. It estimates class-conditional statistics online and anchors them using information from the pretrained text classifier. StatA explicitly considers test streams with varying class composition and correlation, making it an important statistical baseline for our cross-regime evaluation.

Relational adaptation ECALP (Li et al., 2025) performs incremental graph-based inference over textual class prototypes and observed test samples. Its graph is expanded as new samples arrive, allowing relational information from earlier test samples to persist during inference. GRASP uses a more local form of relational reasoning: textual prototypes are not graph nodes, and the test-sample graph is restricted to the current batch and discarded after prediction. Only the bounded test-time cache persists across batches.

GRASP-Base: batch-local label propagation We use **GRASP-Base** as the minimal relational instantiation of our framework, following the classical label-propagation principle of repeatedly mixing an initial prediction with predictions propagated over a similarity graph (Zhou et al., 2003). Starting from the cache-refined posterior $\mathbf{P}_{\text{cache}}$, it constructs a cosine kNN graph over the current batch and applies row-normalized propagation,

$$\mathbf{F}^{(t+1)} = \alpha\mathbf{P}_{\text{cache}} + (1-\alpha)\mathbf{D}^{-1}\mathbf{W}\mathbf{F}^{(t)}, \qquad \alpha = \frac{1}{1+\lambda}, \qquad \mathbf{F}^{(0)} = \mathbf{P}_{\text{cache}}. \tag{4}$$

The affinity matrix $\mathbf{W}$ is symmetrized before row normalization. GRASP-Base uses $T = 3$ propagation steps and discards the graph after each batch. Full GRASP instead uses posterior-consistent affinities and an anchored diffusion objective based on the unnormalized graph Laplacian.

3 GRASP: Graph-Refined Anchored Soft Predictions

GRASP retains the batch-local relational structure of GRASP-Base, but controls how strongly this evidence can modify the prediction. For each test batch, GRASP (i) obtains a cache-refined posterior, (ii) constructs a posterior-consistent neighborhood graph that is discarded after inference, and (iii) performs anchored graph refinement. As illustrated in Figure 1, the persistent cache provides the prediction anchor, while the batch graph contributes relational evidence.

3.1 Online test-time setting

At test step t, we receive an unlabeled batch $\mathcal{B}_t = \{\mathbf{x}_i\}_{i=1}^{B}$ together with the persistent cache state $\mathcal{C}_t = (\mathcal{C}_t^{+}, \mathcal{C}_t^{-})$ accumulated from previous test samples. A frozen image encoder maps each image to an ℓ_2-normalized feature $\mathbf{z}_i \in \mathbb{S}^{d-1}$. No target labels or model-parameter updates are used, and the stream regime is unknown.

3.2 Cache-refined anchor

For each sample $\mathbf{x}_i$ in the current batch, the persistent cache defined in Section 2 yields cache-refined logits $\mathbf{s}_i^{\text{cache}} \in \mathbb{R}^{C}$ and posterior

$$P_{ic}^{\text{cache}} = \text{softmax}\big(\mathbf{s}_i^{\text{cache}}\big)_c, \qquad i = 1, \dots, B, \quad c = 1, \dots, C. \tag{5}$$

Thus $\mathbf{p}_i^{\text{cache}} = (P_{i1}^{\text{cache}}, \dots, P_{iC}^{\text{cache}}) \in \mathbb{R}^{C}$ is the cache-refined posterior for sample i. Stacking these posteriors gives $\mathbf{P}_{\text{cache}} \in \mathbb{R}^{B\times C}$, which serves as the reference for relational refinement.

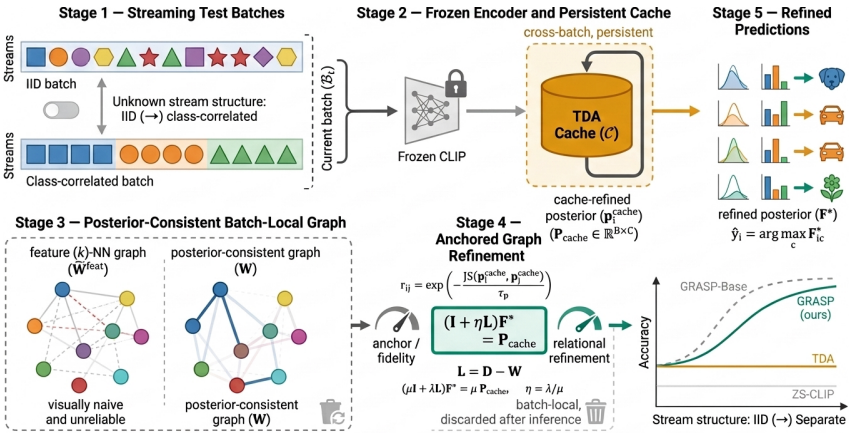


Figure 1: **GRASP.** A persistent cache provides the prediction anchor, while a batch-local cosine kNN graph is attenuated according to disagreement between cache posteriors. GRASP combines them through the anchored solve $(\mathbf{I} + \eta\mathbf{L})\mathbf{F}^* = \mathbf{P}_{\text{cache}}$. The graph is discarded after each batch, while the bounded cache persists, constraining relational influence when local structure is unreliable.

3.3 Posterior-consistent neighborhood graph

GRASP constructs a k-nearest-neighbor graph over the frozen image features. Let $\mathcal{N}_k(i)$ denote the indices of the k nearest neighbors of sample i under cosine similarity. The directed feature affinity is

$$\widetilde{W}_{ij}^{\text{feat}} = \begin{cases} \mathbf{z}_i^\top \mathbf{z}_j, & j \in \mathcal{N}_k(i), \\ 0, & \text{otherwise}, \end{cases} \qquad i \neq j, \tag{6}$$

where ℓ_2-normalization makes the inner product equal to cosine similarity.

Feature proximity alone, however, does not imply that two samples should reinforce one another. We therefore assess whether a feature-space relation is also consistent with the corresponding cache posteriors. The Jensen-Shannon divergence between $\mathbf{p}_i^{\text{cache}}$ and $\mathbf{p}_j^{\text{cache}}$ is

$$\text{JS}(\mathbf{p}_i^{\text{cache}}, \mathbf{p}_j^{\text{cache}}) = \frac{1}{2}\,\text{KL}(\mathbf{p}_i^{\text{cache}} \| M_{ij}) + \frac{1}{2}\,\text{KL}(\mathbf{p}_j^{\text{cache}} \| M_{ij}), \qquad M_{ij} = \frac{\mathbf{p}_i^{\text{cache}} + \mathbf{p}_j^{\text{cache}}}{2}. \tag{7}$$

We convert this disagreement into a posterior-consistency weight

$$r_{ij} = \exp\left(-\frac{\text{JS}(\mathbf{p}_i^{\text{cache}}, \mathbf{p}_j^{\text{cache}})}{\tau_p}\right), \qquad 0 < r_{ij} \leq 1, \tag{8}$$

where τ_p controls the strength of posterior-based attenuation. Collecting these weights in $\mathbf{R} = [r_{ij}]$, the directed affinities become

$$\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^{\text{feat}} \odot \mathbf{R}. \tag{9}$$

The final graph is symmetrized and made non-negative:

$$W_{ij} = \left[\frac{\widetilde{W}_{ij} + \widetilde{W}_{ji}}{2}\right]_+, \qquad W_{ii} = 0, \tag{10}$$

where $[x]_+ = \max(x, 0)$. Thus posterior agreement preserves a candidate feature-space relation, while disagreement attenuates its influence.

Let $\mathbf{D}$ be the degree matrix, $D_{ii} = \sum_j W_{ij}$. GRASP uses the unnormalized graph Laplacian

$$\mathbf{L} = \mathbf{D} - \mathbf{W}. \tag{11}$$

The consistency weights depend only on model predictions; no target labels are used.

3.4 CACHE-ANCHORED GRAPH REFINEMENT

GRASP refines the cache posterior by solving

$$\mathbf{F}^* = \arg\min_{\mathbf{F}} \frac{1}{2}\|\mathbf{F} - \mathbf{P}_{\text{cache}}\|_F^2 + \frac{\eta}{2}\operatorname{tr}(\mathbf{F}^\top \mathbf{L}\mathbf{F}), \tag{12}$$

where $\eta > 0$ controls the strength of relational refinement. The first term preserves the cache prediction, while the second encourages smooth predictions along posterior-consistent graph edges. The stationary condition gives

$$\boxed{(\mathbf{I} + \eta\mathbf{L})\mathbf{F}^* = \mathbf{P}_{\text{cache}}}, \tag{13}$$

and therefore

$$\mathbf{F}^* = (\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{P}_{\text{cache}}. \tag{14}$$

Equivalently, using fidelity and graph weights μ and λ gives

$$\frac{\mu}{2}\|\mathbf{F} - \mathbf{P}_{\text{cache}}\|_F^2 + \frac{\lambda}{2}\operatorname{tr}(\mathbf{F}^\top \mathbf{L}\mathbf{F}), \tag{15}$$

where only the ratio $\eta = \lambda/\mu$ affects the solution. We use $\mu = \lambda = 10$ ($\eta = 1$), $k = 5$, and $\tau_p = 1.0$ throughout. We additionally apply a row-wise simplex projection (Duchi et al., 2008) to $\mathbf{F}^*$ for numerical stability; under the conditions in Section 4, the exact solution already lies in the probability simplex. The final label is

$$\hat{y}_i = \arg\max_c F_{ic}^*. \tag{16}$$

where F_{ic}^* is the refined probability assigned to class c.

3.5 ONLINE INFERENCE

GRASP computes zero-shot predictions, sequentially updates the persistent caches, obtains $\mathbf{P}_{\text{cache}}$, constructs the batch-local graph, refines predictions, and discards the graph. See Algorithm 1 in Appendix A.1.

4 THEORETICAL ANALYSIS

GRASP refines the cache posterior as

$$\mathbf{F}^* = (\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{P}_{\text{cache}}, \tag{17}$$

where $\mathbf{L}$ is the unnormalized Laplacian of the posterior-consistent batch graph. Since $\mathbf{L} \succeq 0$, the solution is unique for $\eta \geq 0$ and, for GRASP's symmetric non-negative graph, preserves the probability simplex. Proofs are provided in Appendix A.2.

4.1 BOUNDED RELATIONAL INFLUENCE

From equation 17,

$$\mathbf{F}^* - \mathbf{P}_{\text{cache}} = -\eta(\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{L}\mathbf{P}_{\text{cache}}. \tag{18}$$

Let $\ell_{\max} = \lambda_{\max}(\mathbf{L})$.

Proposition 1 (Bounded relational influence)**.** *GRASP satisfies*

$$\boxed{\|\mathbf{F}^* - \mathbf{P}_{\text{cache}}\|_F \leq \frac{\eta\ell_{\max}}{1 + \eta\ell_{\max}}\|\mathbf{P}_{\text{cache}}\|_F.} \tag{19}$$

Appendix A.2.1 proves the bound, which quantifies the maximum graph-induced deviation from the cache prediction and makes the stabilizing role of the cache reference explicit. Posterior-consistent weighting further tightens this control. Let $\mathbf{L}_{\text{feat}}$ denote the Laplacian of the corresponding feature-only graph. Since $0 \leq r_{ij} \leq 1$, the implemented symmetric graph satisfies

$$\mathbf{L} \preceq \mathbf{L}_{\text{feat}}, \qquad \lambda_{\max}(\mathbf{L}) \leq \lambda_{\max}(\mathbf{L}_{\text{feat}}). \tag{20}$$

Thus posterior attenuation can only tighten the worst-case bound in equation 19 relative to the same feature graph without attenuation. This bound concerns prediction displacement and does not guarantee improved classification accuracy. The derivation is given in Appendix A.2.2.

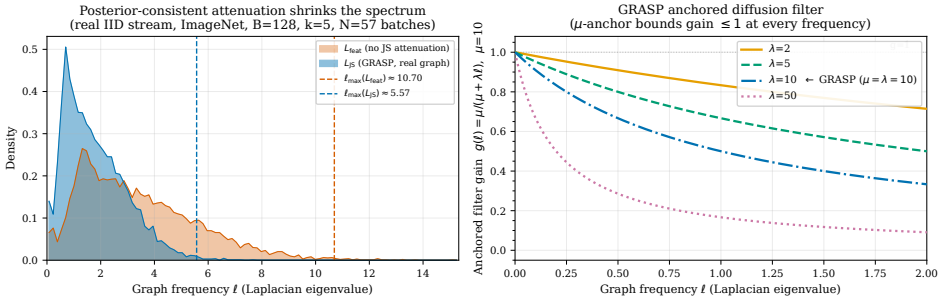


Figure 2: **Spectral effect of posterior-consistent refinement. Left**: Eigenvalue distributions of the unnormalized Laplacians $\mathbf{L}_{\text{feat}}$ and $\mathbf{L}_{\text{JS}}$ over 57 ImageNet IID batches (3 random seeds, B=128, k=5, ViT-B/16). Posterior-consistent attenuation reduces the mean largest eigenvalue from 10.70 to 5.57, consistent with $\mathbf{L}_{\text{JS}} \preceq \mathbf{L}_{\text{feat}}$ and the tighter bound in Proposition 1. **Right**: Diffusion response $g(\ell) = 1/(1+\eta\ell)$ for different refinement strengths. At the default η=1, low-frequency components are preserved more strongly while higher-frequency variation is progressively attenuated.

4.2 Spectral view of graph refinement

Let

$$\mathbf{L} = \mathbf{U}\,\mathrm{diag}(\ell_0, \ldots, \ell_{B-1})\mathbf{U}^\top, \qquad 0 = \ell_0 \le \cdots \le \ell_{B-1}. \tag{21}$$

Writing $\widehat{\mathbf{P}} = \mathbf{U}^\top \mathbf{P}_{\text{cache}}$ and $\widehat{\mathbf{F}} = \mathbf{U}^\top \mathbf{F}^*$, the m-th graph-frequency component satisfies

$$\widehat{\mathbf{F}}_m = g(\ell_m)\widehat{\mathbf{P}}_m, \qquad \boxed{g(\ell) = \frac{1}{1+\eta\ell}}, \tag{22}$$

where $\widehat{\mathbf{P}}_m$ and $\widehat{\mathbf{F}}_m$ are the components associated with Laplacian eigenvalue ℓ_m. Equivalently, $g(\ell) = \mu/(\mu + \lambda\ell)$.

GRASP therefore acts as a graph low-pass filter: low-frequency components are largely preserved, while higher-frequency variation is attenuated. Since $0 < g(\ell) \le 1$, refinement does not amplify variation, while null-space components are preserved exactly.

The same spectral view also characterizes how much each component can move: $h(\ell) = 1 - g(\ell) = \eta\ell/(1+\eta\ell)$. Since $h(\ell)$ is maximized at $\ell_{\max}$, it recovers the bound in Proposition 1. The spectral response and displacement bound therefore describe complementary aspects of the same refinement operator. Figure 2 shows the empirical Laplacian spectrum and corresponding diffusion response.

4.3 Relation to GRASP-Base

GRASP-Base uses the finite-step propagation in equation 4. At convergence, its fixed point is

$$\mathbf{F}^*_{\text{Base}} = [\mathbf{I} + \lambda(\mathbf{I} - \mathbf{S})]^{-1}\mathbf{P}_{\text{cache}}, \qquad \mathbf{S} = \mathbf{D}^{-1}\mathbf{W}.$$

GRASP-Base uses feature-only affinities and T=3 propagation steps, whereas GRASP uses posterior-consistent affinities, the unnormalized Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$, and directly solves the anchored system. The full derivation is given in Appendix A.2.3.

5 Experiments

5.1 Experimental Setup

Model We use CLIP ViT-B/16 (Radford et al., 2021), implemented with OpenCLIP (Ilharco et al., 2021), with a frozen encoder, ℓ_2-normalized image features, logit scale $\tau = 100$, and class text embeddings constructed from an ensemble of 80 prompt templates.

Table 1: Cross-regime training-free adaptation on 11 datasets (ViT-B/16). All methods use identically ordered test streams. Worst is the minimum and Mean is the average across stream regimes.

	Stream regime					Cross-regime	
Method	IID	γ=.1	γ=.01	γ=.001	Separate	Worst↑	Mean↑
ZS-CLIP	65.31±0.00	65.31±0.00	65.31±0.00	65.31±0.00	65.31±0.00	65.31	65.31
TDA	67.73±0.17	67.74±0.26	67.12±0.33	67.09±0.30	66.85±0.40	66.85	67.30
DMN	68.29±0.08	67.55±0.29	66.94±0.32	66.58±0.15	66.56±0.49	66.56	67.18
MTA	66.67±0.04	66.66±0.05	66.67±0.05	66.69±0.03	66.69±0.03	66.66	66.68
ECALP	68.42±0.12	68.11±0.25	66.62±0.32	65.30±0.39	64.71±0.38	64.71	66.63
StatA	64.55±0.11	66.27±0.21	68.45±0.21	69.46±0.38	69.12±0.14	64.55	67.57
TT-RAA	67.80±0.31	67.02±1.20	65.98±0.95	65.82±0.41	65.87±0.50	65.82	66.50
DOTA	69.60±0.09	69.33±0.19	68.66±0.24	68.28±0.20	68.33±0.28	68.28	68.84
ADAPT	**70.44±0.14**	**69.83±0.32**	68.92±0.49	68.43±0.34	68.63±0.60	**68.43**	69.25
GRASP-Base	52.60±0.25	56.55±0.38	66.90±0.47	74.29±0.38	**77.08±0.50**	52.60	65.49
GRASP	66.91±0.21	69.10±0.31	**72.43±0.44**	**74.52±0.35**	75.25±0.47	66.91	**71.64**

Datasets We evaluate on 11 image-classification benchmarks: Caltech-101 (Fei-Fei et al., 2004), DTD (Cimpoi et al., 2014), Oxford Flowers-102 (Nilsback & Zisserman, 2008), Oxford Pets (Parkhi et al., 2012), FGVC Aircraft (Maji et al., 2013), UCF-101 (Soomro et al., 2012), EuroSAT (Helber et al., 2019), Stanford Cars (Krause et al., 2013), SUN-397 (Xiao et al., 2010), Food-101 (Bossard et al., 2014), and ImageNet (Deng et al., 2009).

Evaluation protocol We consider five stream regimes: uniformly shuffled **IID**; Dirichlet-sampled streams (Zanella et al., 2025) with $\gamma \in \{0.1, 0.01, 0.001\}$, where smaller γ induces stronger class imbalance and local correlation; and **Separate**, where classes arrive sequentially. All methods use identically ordered test streams. For the main comparison, results are averaged over 10 random seeds per dataset and stream regime.

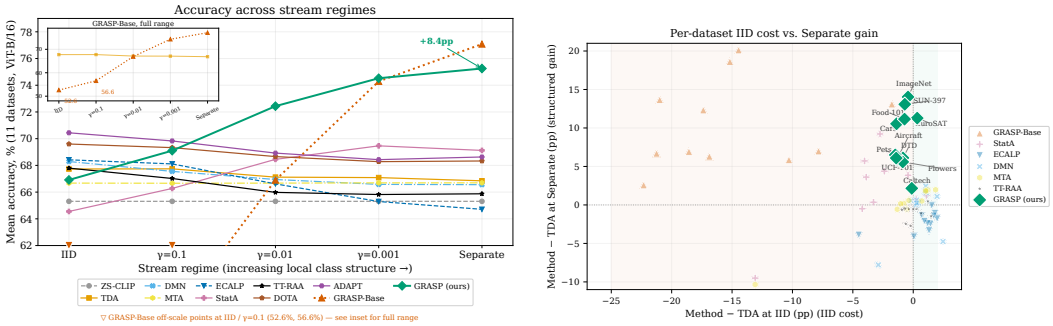
Baselines and implementation details We compare against ZS-CLIP, TDA (Karmanov et al., 2024), StatA (Zanella et al., 2025), ECALP (Li et al., 2025), DMN (Zhang et al., 2024), MTA (Zanella & Ben Ayed, 2024), TT-RAA (Fan et al., 2025), DOTA (Han et al., 2025), ADAPT (Zhang et al., 2025), and GRASP-Base. ZLaP (Stojnić et al., 2024) assumes a fixed unlabeled reference set and is excluded from the online-stream comparison; gradient-based methods such as TENT (Wang et al., 2021) and TPT (Shu et al., 2022) are excluded from the primary training-free comparison. GRASP-Base uses T=3, k=5, λ=10, and row-normalized propagation $\mathbf{S} = \mathbf{D}^{-1}\mathbf{W}$. For TDA, we use positive/negative cache capacities 3/2, (α_+, β_+)=(2, 5), (α_-, β_-)=(0.117, 1), and negative-cache entropy range $[0.2, 0.5]$. Under our batched online protocol with B=128, each batch is inserted before cache-augmented prediction, allowing retrieval from current and earlier batches; TDA, GRASP-Base, and GRASP share this convention. GRASP uses μ=λ=10 (η=1), k=5, and τ_p=1 throughout. Appendix A.7 reports sensitivity around this setting.

5.2 Main Results: Behavior Across Stream Structure

We first compare all methods across the five stream regimes. Figure 3(a) shows mean accuracy across the 11 datasets for each stream regime. Table 1 also reports worst-regime and cross-regime mean.

Overall pattern Methods exhibit markedly different responses to stream structure. ADAPT is strongest under weak structure, reaching 70.44% under IID and 69.83% at γ=0.1, but declines as structure increases. DOTA shows a milder decline, while StatA improves under more structured streams. GRASP-Base depends most strongly on stream composition, rising from 52.60% under IID to 77.08% under Separate. Full GRASP moderates this trade-off, rising from 66.91% under IID to 75.25% under Separate and achieving the highest cross-regime mean of 71.64%.

Structured-stream gains Relative to TDA, GRASP moves from -0.82 points under IID to $+1.36$, $+5.31$, $+7.43$, and $+8.40$ points as stream structure increases. Against the stronger weak-structure ADAPT baseline, the corresponding gap changes from -3.53 and -0.73 points under IID and



(a) Cross-regime accuracy **(b) IID cost vs. Separate gain**

Figure 3: **Behavior across stream structure. (a)** Mean accuracy across five stream regimes on 11 datasets (ViT-B/16), showing the different responses of training-free methods to increasing local stream structure. **(b)** Per-dataset change relative to TDA under IID and Separate ordering, showing the resulting trade-offs across methods.

Table 2: Operator ablation from GRASP-Base to GRASP. W^{feat}: symmetrized cosine kNN affinity (k=5); $D = \text{diag}(W\mathbf{1})$, $S = D^{-1}W$; W^{JS} adds posterior-consistency weighting. Base uses T=3; anchored variants are solved directly. Mean accuracy (%) over four datasets and 10 random seeds.

Variant	Operator	IID	Separate
GRASP-Base	$S = D^{-1}W^{\text{feat}}$	55.29	74.02
→ converged	$S = D^{-1}W^{\text{feat}}$	59.49	75.09
→ anchored, random-walk	$L = I - D^{-1}W^{\text{feat}}$	68.38	70.16
→ anchored, unnormalized	$L = D - W^{\text{feat}}$	66.77	73.99
→ posterior-consistent	$L = D - W^{\text{JS}}$	67.68	73.40

γ=0.1 to +3.51, +6.09, and +6.62 under the three more structured regimes. Thus, methods that are strongest under weak structure need not benefit most as local relational structure increases.

Per-dataset trade-off Figure 3(b) compares changes in accuracy relative to TDA for selected methods under IID and Separate ordering. GRASP remains close to TDA under IID while retaining substantial gains on Separate streams. GRASP-Base obtains larger gains under strong structure but at a substantially larger IID cost, whereas StatA and ECALP exhibit different regime-dependent trade-offs. Full per-dataset results are provided in Tables 3-7 in Appendices A.3-A.5.

5.3 From GRASP-Base to GRASP

Table 2 compares successive variants on DTD, Caltech-101, Flowers-102 and EuroSAT under IID and Separate ordering. Running GRASP-Base propagation to convergence improves IID accuracy by 4.20 percentage points and Separate accuracy by 1.07 points. The random-walk Laplacian configuration yields the largest IID improvement in this sequence. Compared with converged propagation, it improves IID accuracy by 8.89 percentage points but reduces accuracy under Separate ordering by 4.93 percentage points. Switching to the unnormalized Laplacian improves accuracy under Separate ordering by 3.83 percentage points but reduces IID accuracy by 1.61 percentage points. Posterior-consistent weighting then adds 0.91 IID points while reducing Separate accuracy by 0.59 points. Per-dataset results are reported in Tables 14 and 15 in Appendix A.13.

5.4 Controlling Relational Influence

Effect of posterior-consistency attenuation We isolate posterior attenuation using a matched no-JS variant that retains the same anchored refinement but sets $r_{ij} = 1$ for all candidate edges. Removing attenuation changes the cross-regime mean only slightly, from 71.64% to 71.53%, but reduces worst-regime accuracy from 66.91% to 66.06%. Full GRASP is stronger under IID, γ=0.1, and γ=0.01,

whereas no-JS is stronger under γ=0.001 and Separate. Thus, posterior attenuation controls the operating point rather than uniformly improving every regime. See Table 8 in Appendix A.6.

Natural temporal structure To complement the synthetic regimes, we compare a naturally ordered UCF-101 stream, preserving within-clip temporal order, with a global shuffle. Across four seeds, GRASP improves from 66.99% to 72.98% (+5.99 points), with a positive gain for every seed; GRASP-Base shows a larger 20.56-point gap, indicating greater order dependence (Appendix A.10).

Generalization across backbones Using the same configuration, GRASP shows similar cross-regime behavior on CLIP ViT-B/32 and ViT-L/14 (Radford et al., 2021), and SigLIP-L (Zhai et al., 2023). Separate gains over TDA range from 6.61 to 9.54 points, and GRASP achieves the highest cross-regime mean in each case. Figure 5 in Appendix A.12 reports the results for all four backbones.

Relational refinement across upstream adapters The default GRASP uses TDA as its anchor. We also apply the same relational operator to DOTA and ADAPT while preserving their native online updates. Its contribution grows with stream structure: from -0.47 percentage points under IID to $+7.37$ points under Separate for DOTA, and from -0.07 to $+4.45$ points for ADAPT. Thus, stronger upstream adapters incur little IID cost while gaining as local structure increases. See Appendix A.14.

6 Related Work

Training-free and relational VLM adaptation Training-free methods adapt frozen VLMs using test-time memory or target statistics. TDA (Karmanov et al., 2024) and DMN (Zhang et al., 2024) maintain test-time memories, while StatA (Zanella et al., 2025) and ADAPT (Zhang et al., 2025) estimate target statistics online. COSMIC (Huang et al., 2025) combines frozen CLIP and DINOv2 representations through multi-space cache and clique inference. DOTA (Han et al., 2025) estimates test-stream distributions online, while ACE (Nguyen et al., 2025) uses adaptive class-specific thresholds for reliable cache entries. Relational methods such as ZLaP (Stojnić et al., 2024), TransCLIP (Zanella et al., 2024), and ECALP (Li et al., 2025) exploit relations among unlabeled target samples. GRASP instead uses only current-batch graph, retains a bounded cache across batches, and controls relational influence through cache anchoring and posterior-consistent edge weighting.

Gradient-based test-time adaptation A separate line adapts model or prompt parameters at inference. TENT (Wang et al., 2021) minimizes entropy through parameter updates, while TPT (Shu et al., 2022) and DiffTPT (Feng et al., 2023) optimize text prompts, and PromptAlign (Abdul Samadh et al., 2023) adds distribution alignment. These methods complement our training-free setting.

Graph regularization and label propagation GRASP builds on classical graph smoothness and label propagation (Zhou et al., 2003; Zhu et al., 2003), also used in transductive few-shot inference (Liu et al., 2019; Lazarou et al., 2021). LATA (Bozorgtabar et al., 2026) applies Laplacian refinement to conformal uncertainty in training-free medical VLM inference over a calibration–test pool. GRASP instead uses a transient batch-local graph, cache anchoring, and posterior-agreement edge attenuation for online adaptation.

7 Conclusion

We studied how relational evidence in training-free VLM adaptation depends on test-stream structure. GRASP-Base shows that local propagation can excel on structured streams but degrade under heterogeneous batches. GRASP controls this trade-off through cache anchoring and posterior-consistent attenuation, achieving a 71.64% cross-regime mean across 11 datasets with one fixed configuration. These results support relational evidence as a controlled refinement of cache predictions rather than an unconstrained replacement.

Limitations GRASP requires batched inference and depends on frozen-feature quality. Posterior agreement measures compatibility rather than correctness, so consistent errors may reinforce one another. GRASP-Base can remain stronger under highly coherent local structure, and evaluation is limited to image classification.

AI USE STATEMENT

In this work, we used generative AI tools to provide reviewer-style research feedback and to assist with tasks such as mathematical formulation, code development, plotting, block-diagram creation, and manuscript editing. All AI-assisted content was reviewed and verified for correctness by the authors.

REFERENCES

Jameel Abdul Samadh, Mohammad Hanan Gani, Noor Hussein, Muhammad Uzair Khattak, Muhammad Muzammal Naseer, Fahad Shahbaz Khan, and Salman H. Khan. Align your prompts: Test-time prompting with distribution alignment for zero-shot generalization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 80396–80413, 2023.

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101: Mining discriminative components with random forests. In *European Conference on Computer Vision*, volume 8694 of *Lecture Notes in Computer Science*, pp. 446–461, 2014.

Behzad Bozorgtabar, Dwarikanath Mahapatra, Sudipta Roy, Muzammal Naseer, Imran Razzak, and Zongyuan Ge. LATA: Laplacian-assisted transductive adaptation for conformal uncertainty in medical VLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36311–36320, 2026.

Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3606–3613, 2014.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 272–279, 2008.

Xinqi Fan, Xueli Chen, Luoxiao Yang, Chuin Hong Yap, Rizwan Qureshi, Qi Dou, Moi Hoon Yap, and Mubarak Shah. Test-time retrieval-augmented adaptation for vision-language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8810–8819, 2025.

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp. 178–178, 2004.

Chun-Mei Feng, Kai Yu, Yong Liu, Salman Khan, and Wangmeng Zuo. Diverse data augmentation with diffusions for effective test-time prompt tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2704–2714, 2023.

Zongbo Han, Jialong Yang, Guangyu Wang, Junfan Li, Qianli Xu, Mike Zheng Shou, and Changqing Zhang. DOTA: Distributional test-time adaptation of vision-language models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 158436–158462, 2025.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

Fanding Huang, Jingyan Jiang, Qinting Jiang, Hebei Li, Faisal Nadeem Khan, and Zhi Wang. COSMIC: Clique-oriented semantic multi-space integration for robust CLIP test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9772–9781, 2025.

Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. OpenCLIP, 2021.

Adilbek Karmanov, Dayan Guan, Shijian Lu, Abdulmotaleb El Saddik, and Eric Xing. Efficient test-time adaptation of vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14162–14171, 2024.

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 Workshops*, pp. 554–561, 2013.

Michalis Lazarou, Tania Stathaki, and Yannis Avrithis. Iterative label cleaning for transductive and semi-supervised few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8751–8760, 2021.

Yushu Li, Yongyi Su, Adam Goodge, Kui Jia, and Xun Xu. Efficient and context-aware label propagation for zero-/few-shot training-free adaptation of vision-language model. In *International Conference on Learning Representations*, 2025.

Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, and Yi Yang. Learning to propagate labels: Transductive propagation network for few-shot learning. In *International Conference on Learning Representations*, 2019.

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.

Khanh-Binh Nguyen, Phuoc-Nguyen Bui, Hyunseung Choo, and Duc Thanh Nguyen. Adaptive cache enhancement for test-time adaptation of vision-language models. *arXiv preprint arXiv:2508.07570*, 2025.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 722–729, 2008.

Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3498–3505, 2012.

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 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, and Chaowei Xiao. Test-time prompt tuning for zero-shot generalization in vision-language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 14274–14289, 2022.

Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*, 2012.

Vladan Stojnić, Yannis Kalantidis, and Giorgos Tolias. Label propagation for zero-shot classification with vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23209–23218, 2024.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. TENT: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.

Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3485–3492, 2010.

Maxime Zanella and Ismail Ben Ayed. On the test-time zero-shot generalization of vision-language models: Do we really need prompt learning? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23783–23793, 2024.

Maxime Zanella, Benoît Gérin, and Ismail Ben Ayed. Boosting vision-language models with transduction. In *Advances in Neural Information Processing Systems*, volume 37, pp. 62223–62256, 2024.

Maxime Zanella, Clément Fuchs, Christophe De Vleeschouwer, and Ismail Ben Ayed. Realistic test-time adaptation of vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25103–25112, 2025.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11975–11986, 2023.

Yabin Zhang, Wenjie Zhu, Hui Tang, Zhiyuan Ma, Kaiyang Zhou, and Lei Zhang. Dual memory networks: A versatile adaptation approach for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28718–28728, 2024.

Youjia Zhang, Youngeun Kim, Young-Geun Choi, Hongyeob Kim, Huiling Liu, and Sungeun Hong. Backpropagation-free test-time adaptation via probabilistic Gaussian alignment. In *Advances in Neural Information Processing Systems*, volume 38, pp. 84921–84951, 2025.

Dengyong Zhou, Olivier Bousquet, Thomas N. Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*, volume 16, pp. 321–328, 2003.

Xiaojin Zhu, Zoubin Ghahramani, and John D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 912–919, 2003.

A Appendix

A.1 GRASP Inference Algorithm

Algorithm 1 GRASP inference for one test batch

Require: Test batch $\mathcal{B}_t$, cache $\mathcal{C}_t$, frozen VLM, η, k, τ_p
Encode $\mathcal{B}_t$ and compute zero-shot logits
for $i = 1, \ldots, B$ **do**
    Update the positive/negative cache using sample i
end for
Compute cache-augmented logits using the updated cache
$\mathbf{P}_{\text{cache}} \leftarrow \text{softmax}(\text{cache-augmented logits})$
Construct directed cosine kNN affinities $\widetilde{\mathbf{W}}^{\text{feat}}$
Compute posterior-consistency weights $\mathbf{R}$ using equation 8
$\widetilde{\mathbf{W}} \leftarrow \widetilde{\mathbf{W}}^{\text{feat}} \odot \mathbf{R}$
Symmetrize $\widetilde{\mathbf{W}}$ to obtain $\mathbf{W}$ using equation 10
Form $\mathbf{L} = \mathbf{D} - \mathbf{W}$
Solve $(\mathbf{I} + \eta\mathbf{L})\mathbf{F}^* = \mathbf{P}_{\text{cache}}$
Project rows of $\mathbf{F}^*$ onto the probability simplex ▷ numerical safeguard
Predict $\hat{y}_i = \arg\max_c F^*_{ic}$ for $i = 1, \ldots, B$
Discard the current batch graph
return $\{\hat{y}_i\}_{i=1}^B$ and updated cache

A.2 Additional Theoretical Details

We provide supporting derivations for the properties used in Section 4. Recall that

$$\mathbf{F}^* = (\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{P}_{\text{cache}}, \tag{23}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the unnormalized Laplacian of the symmetric, non-negative, zero-diagonal GRASP graph.

Existence and simplex preservation Since $\mathbf{L} \succeq 0$, $\mathbf{I} + \eta\mathbf{L}$ is positive definite for $\eta \geq 0$, and hence equation 23 has a unique solution.

The solution also preserves the probability simplex in exact arithmetic. Let $\mathbf{A} = \mathbf{I} + \eta\mathbf{L}$. Its diagonal entries are positive, its off-diagonal entries are non-positive, and

$$A_{ii} = 1 + \eta D_{ii} > \eta D_{ii} = \sum_{j \neq i} |A_{ij}|.$$

Thus $\mathbf{A}$ is a nonsingular M-matrix and $\mathbf{A}^{-1} \geq 0$ entrywise. Since $\mathbf{P}_{\text{cache}} \geq 0$, this gives $\mathbf{F}^* \geq 0$. Moreover, $\mathbf{L}\mathbf{1}_B = \mathbf{0}$ implies $\mathbf{A}^{-1}\mathbf{1}_B = \mathbf{1}_B$. Since each row of $\mathbf{P}_{\text{cache}}$ sums to one,

$$\mathbf{F}^*\mathbf{1}_C = \mathbf{A}^{-1}\mathbf{P}_{\text{cache}}\mathbf{1}_C = \mathbf{1}_B. \tag{24}$$

Hence every row of $\mathbf{F}^*$ is non-negative and sums to one. The evaluated implementation nevertheless applies a row-wise simplex projection after the solve for numerical stability.

A.2.1 Proof of the displacement bound

From equation 23,

$$\mathbf{F}^* - \mathbf{P}_{\text{cache}} = -\eta(\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{L}\mathbf{P}_{\text{cache}}. \tag{25}$$

Using the eigendecomposition $\mathbf{L} = \mathbf{U}\,\text{diag}(\ell_m)\mathbf{U}^\top$,

$$\eta(\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{L} = \mathbf{U}\,\text{diag}\left(\frac{\eta\ell_m}{1 + \eta\ell_m}\right)\mathbf{U}^\top. \tag{26}$$

Since $\eta\ell/(1 + \eta\ell)$ is increasing for $\ell \geq 0$,

$$\left\|\eta(\mathbf{I} + \eta\mathbf{L})^{-1}\mathbf{L}\right\|_2 = \frac{\eta\ell_{\max}}{1 + \eta\ell_{\max}}. \tag{27}$$

Therefore,

$$\|\mathbf{F}^* - \mathbf{P}_{\text{cache}}\|_F \leq \frac{\eta \ell_{\max}}{1 + \eta \ell_{\max}} \|\mathbf{P}_{\text{cache}}\|_F, \tag{28}$$

which proves Proposition 1.

A.2.2 Effect of posterior-consistent attenuation

Let $\mathbf{W}_{\text{feat}}$ denote the graph obtained using the same kNN construction and symmetrization as GRASP, but without posterior attenuation. Since Jensen-Shannon divergence is symmetric, $r_{ij} = r_{ji}$, and the implemented graph satisfies

$$W_{ij} = r_{ij} W_{ij}^{\text{feat}}, \qquad 0 \leq r_{ij} \leq 1. \tag{29}$$

Hence $\Delta\mathbf{W} = \mathbf{W}_{\text{feat}} - \mathbf{W}$ is symmetric and entrywise non-negative. Let $\mathbf{L}_\Delta$ denote its Laplacian. Then $\mathbf{L}_\Delta \succeq 0$ and

$$\mathbf{L}_{\text{feat}} = \mathbf{L} + \mathbf{L}_\Delta, \tag{30}$$

which gives

$$\boxed{\mathbf{L} \preceq \mathbf{L}_{\text{feat}}} \qquad \text{and} \qquad \lambda_{\max}(\mathbf{L}) \leq \lambda_{\max}(\mathbf{L}_{\text{feat}}). \tag{31}$$

Combining this with Proposition 1 shows that posterior attenuation cannot increase the corresponding worst-case displacement bound. This does not imply smaller realized displacement for every individual sample.

A.2.3 Relation to GRASP-Base

GRASP-Base uses the update

$$\mathbf{F}^{(t+1)} = \alpha \mathbf{P}_{\text{cache}} + (1 - \alpha)\mathbf{S}\mathbf{F}^{(t)}, \qquad \alpha = \frac{1}{1 + \lambda}, \qquad \mathbf{S} = \mathbf{D}^{-1}\mathbf{W}. \tag{32}$$

Since $\mathbf{S}$ is row-stochastic and $0 < \alpha \leq 1$, $\rho((1 - \alpha)\mathbf{S}) < 1$, so the iteration converges to a unique fixed point. At convergence,

$$[\mathbf{I} - (1 - \alpha)\mathbf{S}]\mathbf{F}^*_{\text{Base}} = \alpha \mathbf{P}_{\text{cache}}. \tag{33}$$

Substituting $\alpha = 1/(1 + \lambda)$ gives

$$\boxed{\mathbf{F}^*_{\text{Base}} = [\mathbf{I} + \lambda(\mathbf{I} - \mathbf{S})]^{-1}\mathbf{P}_{\text{cache}}.} \tag{34}$$

Thus GRASP-Base and GRASP both admit anchored graph-resolvent forms but use different relational operators. GRASP-Base uses feature-only affinities, row-normalized propagation, and $T = 3$ iterations in the evaluated method, whereas GRASP uses posterior-consistent affinities, the unnormalized Laplacian $\mathbf{D} - \mathbf{W}$, and directly solves the anchored diffusion system.

A.3 Per-dataset Results: IID Regime

Table 3: Per-dataset accuracy (%) under IID ordering. ViT-B/16, B=128, 10 random seeds.

Dataset	ZS-CLIP	TDA	DMN	MTA	TT-RAA	ECALP	StatA	DOTA	ADAPT	GRASP-Base	GRASP
Aircraft	24.81	25.34	25.83	27.16	25.04	26.82	21.41	26.10	28.49	17.55	23.99
Caltech-101	93.27	93.44	93.70	94.13	92.88	93.45	93.66	94.19	94.48	71.17	93.32
Cars	65.51	66.74	67.86	67.92	66.13	68.12	63.95	67.61	68.56	49.41	65.36
DTD	44.09	45.84	45.70	45.04	47.12	46.49	46.89	47.81	54.37	35.60	44.99
EuroSAT	48.36	60.67	63.18	47.55	62.21	56.11	47.55	62.68	66.75	58.95	61.00
Flowers-102	70.73	72.29	74.30	71.73	72.62	74.04	68.22	75.06	75.31	55.46	71.42
Food-101	85.88	86.10	83.25	87.09	85.95	87.36	85.63	86.92	83.86	65.16	85.38
ImageNet	66.72	68.26	69.28	69.27	67.58	69.40	65.83	70.67	70.77	53.85	67.83
Pets	89.04	89.83	90.16	89.44	89.89	90.71	88.01	92.04	91.97	68.63	88.34
SUN-397	62.56	65.88	66.96	64.81	65.00	67.81	62.55	69.64	70.33	50.75	65.18
UCF-101	67.41	70.63	70.97	69.28	71.34	72.28	66.37	72.85	69.91	52.11	69.22
Avg	65.31	67.73	68.29	66.67	67.80	68.42	64.55	69.60	70.44	52.60	66.91

Table 6: Per-dataset accuracy (%) under the $\gamma = 0.01$ (moderately structured) stream.

Dataset	ZS-CLIP	TDA	DMN	MTA	TT-RAA	ECALP	StatA	DOTA	ADAPT	GRASP-Base	GRASP
Aircraft	24.81	25.27	25.82	27.23	24.62	26.24	26.68	26.17	28.29	27.12	28.68
Caltech-101	93.27	93.53	93.61	94.15	92.26	90.84	94.25	93.96	94.69	85.45	94.77
Cars	65.51	66.47	67.38	67.78	65.89	66.24	71.59	67.11	68.29	64.69	72.72
DTD	44.09	45.54	45.83	45.06	46.61	44.02	46.91	47.39	52.73	46.67	49.67
EuroSAT	48.36	55.51	54.16	47.63	57.18	54.36	51.51	56.39	55.64	68.03	66.27
Flowers	70.73	72.64	73.56	71.62	72.08	72.39	75.06	74.84	75.75	68.54	76.03
Food-101	85.88	85.76	80.27	87.09	75.71	84.51	89.20	86.42	81.78	93.29	94.65
ImageNet	66.72	68.31	69.31	69.23	67.50	68.94	68.37	70.70	70.72	56.26	72.29
Pets	89.04	89.07	89.26	89.48	88.91	88.44	94.68	91.03	90.08	91.22	94.44
SUN-397	62.56	65.79	66.98	64.82	64.81	66.66	65.42	69.49	70.16	64.38	72.55
UCF-101	67.41	70.44	70.19	69.32	70.20	70.14	69.30	71.79	69.96	70.24	74.70
Avg	65.31	67.12	66.94	66.67	65.98	66.62	68.45	68.66	68.92	66.90	72.43

A.4 Per-dataset Results: Separate (Class-Sequential) Regime

Table 4: Per-dataset accuracy (%) under the Separate (fully class-sequential) stream.

Dataset	ZS-CLIP	TDA	DMN	MTA	TT-RAA	ECALP	StatA	DOTA	ADAPT	GRASP-Base	GRASP
Aircraft	24.81	25.33	25.69	27.37	24.78	25.30	28.89	26.23	28.06	32.27	31.28
Caltech-101	93.27	93.53	93.57	94.11	91.05	89.45	94.17	93.99	95.15	96.06	95.69
Cars	65.51	65.90	67.56	67.91	65.36	63.49	75.07	66.73	67.94	78.17	76.44
DTD	44.09	44.79	45.42	45.05	45.24	43.48	45.97	47.10	52.73	50.59	50.90
EuroSAT	48.36	57.95	53.26	47.68	56.47	54.10	48.39	56.56	56.17	70.99	69.23
Flowers-102	70.73	72.18	73.35	71.69	71.64	71.16	77.85	74.43	75.84	78.43	77.77
Food-101	85.88	85.14	77.41	87.09	82.40	81.87	88.94	86.07	79.37	98.75	96.29
ImageNet	66.72	67.37	69.25	69.24	66.89	65.03	71.70	70.35	69.93	87.44	81.40
Pets	89.04	88.69	89.38	89.35	88.12	86.63	95.18	90.51	90.61	95.33	95.19
SUN-397	62.56	64.60	67.11	64.83	63.95	62.87	64.90	69.03	69.50	83.16	77.68
UCF-101	67.41	69.82	70.14	69.32	68.73	68.45	69.28	70.66	69.69	76.69	75.91
Avg	65.31	66.85	66.56	66.69	65.87	64.71	69.12	68.33	68.63	77.08	75.25

A.5 Per-dataset Results: $\gamma = 0.1$, $\gamma = 0.01$, $\gamma = 0.001$

Table 5: Per-dataset accuracy (%) under the $\gamma = 0.1$ (mildly structured) stream.

Dataset	ZS-CLIP	TDA	DMN	MTA	TT-RAA	ECALP	StatA	DOTA	ADAPT	GRASP-Base	GRASP
Aircraft	24.81	25.18	25.60	27.15	24.91	26.72	22.45	26.29	28.48	19.44	24.64
Caltech-101	93.27	93.33	93.66	94.12	92.78	92.76	93.84	94.25	94.30	71.62	93.42
Cars	65.51	66.60	67.34	67.86	65.93	67.57	65.37	67.35	68.39	49.39	66.48
DTD	44.09	45.80	45.69	44.97	47.71	45.56	47.52	47.67	53.28	39.33	46.86
EuroSAT	48.36	60.24	57.60	47.61	60.97	56.60	52.72	59.45	61.19	69.70	68.78
Flowers	70.73	72.42	73.84	71.74	72.33	73.72	70.09	75.38	75.56	55.97	72.72
Food-101	85.88	86.08	82.66	87.10	78.63	86.65	87.98	86.82	83.52	76.63	90.03
ImageNet	66.72	68.37	69.31	69.24	67.58	69.56	65.83	70.77	70.78	54.04	67.90
Pets	89.04	89.79	89.73	89.40	89.80	89.84	92.18	91.48	91.44	78.87	91.95
SUN-397	62.56	66.03	67.06	64.78	65.02	68.02	62.89	69.77	70.46	49.99	66.01
UCF-101	67.41	71.32	70.57	69.32	71.57	72.22	68.14	73.43	70.75	57.12	71.33
Avg	65.31	67.74	67.55	66.66	67.02	68.11	66.27	69.33	69.83	56.55	69.10

A.6 Mechanistic analysis

We further examine how posterior consistency affects relational refinement, using ground-truth labels only for post-hoc diagnosis and never during inference. For candidate kNN edges, agreement between cache posteriors helps predict whether the connected samples belong to the same class. On five datasets (EuroSAT, DTD, Caltech-101, UCF-101, Food-101) and three random seeds, same-class edges have substantially lower Jensen-Shannon (JS) divergence and higher consistency weights r_{ij} than different-class edges (Table 8). Under IID, mean JS is 0.083 versus 0.567, with $r_{ij} = 0.927$ versus 0.579; under Separate, the corresponding values are 0.086 versus 0.425 and 0.926 versus 0.668. Using r_{ij} to distinguish same-class from different-class edges gives AUROC 0.958 under IID and 0.911 under Separate. Thus, cache-posterior agreement provides an informative proxy for

Table 7: Per-dataset accuracy (%) under the $\gamma = 0.001$ (strongly structured) stream.

Dataset	ZS-CLIP	TDA	DMN	MTA	TT-RAA	ECALP	StatA	DOTA	ADAPT	GRASP-Base	GRASP
Aircraft	24.81	24.93	25.70	27.32	24.68	25.96	28.28	26.11	27.99	30.21	29.86
Caltech-101	93.27	93.58	93.67	94.13	91.52	89.54	94.30	93.94	94.91	92.34	95.44
Cars	65.51	66.02	67.47	67.85	65.40	64.21	74.53	66.98	68.19	74.83	75.51
DTD	44.09	45.32	45.46	45.13	46.17	43.18	46.32	47.07	52.81	49.10	49.98
EuroSAT	48.36	58.31	52.98	47.51	56.18	53.70	52.68	55.42	54.80	71.13	69.29
Flowers	70.73	72.33	73.37	71.68	71.77	71.45	76.60	74.36	74.88	75.18	77.24
Food-101	85.88	85.52	78.03	87.09	79.55	82.57	89.24	86.15	79.79	97.27	95.67
ImageNet	66.72	67.95	69.27	69.24	67.26	67.01	71.56	70.47	70.33	78.79	79.44
Pets	89.04	88.88	89.46	89.39	88.44	87.74	95.04	90.75	90.01	93.98	94.75
SUN-397	62.56	65.13	67.02	64.84	64.37	64.48	65.75	69.12	69.73	79.32	76.87
UCF-101	67.41	69.97	69.93	69.41	68.69	68.41	69.80	70.76	69.25	75.03	75.72
Avg	65.31	67.09	66.58	66.69	65.82	65.30	69.46	68.28	68.43	74.29	74.52

Table 8: Mechanistic analysis of posterior-consistent attenuation. **A:** Edge compatibility on five datasets and three random seeds (ViT-B/16, B=128). Ground-truth labels are used only for post-hoc diagnosis. **B:** Matched full-GRASP versus no-JS comparison on all 11 datasets and 10 random seeds. The variants are identical except that no-JS sets $r_{ij} = 1$ for all candidate feature edges. Worst is the minimum and Mean is the average across the five stream regimes.

A: Edge compatibility

Regime	JS (same)	JS (diff.)	r (same)	r (diff.)	AUROC
IID	0.083	0.567	0.927	0.579	0.958
Separate	0.086	0.425	0.926	0.668	0.911

B: Effect of posterior-consistency attenuation

Variant	IID	$\gamma = .1$	$\gamma = .01$	$\gamma = .001$	Separate	Worst	Mean
No-JS	66.06	68.07	71.97	**75.06**	**76.46**	66.06	71.53
GRASP	**66.91**	**69.10**	**72.43**	74.52	75.25	**66.91**	**71.64**

the compatibility of candidate feature-space relations. Table 8 also reports the complete 11-dataset comparison with the matched no-JS variant discussed in Section 5.4.

A.7 Extended Hyperparameter Sensitivity

We report post-hoc sensitivity to GRASP's three effective hyperparameters, refinement strength $\eta = \lambda/\mu$, neighborhood size k, and posterior-consistency temperature τ_p, independently around the fixed configuration $(\eta, k, \tau_p) = (1, 5, 1)$. Results are averaged over five datasets (EuroSAT, UCF-101, ImageNet, Food-101, and Caltech-101), four random seeds, and both IID and Separate streams. Figure 4 and Table 9 show the trade-off between relational strength and IID performance: increasing η or k generally benefits Separate streams at some IID cost, while τ_p has a milder effect. Within the central ranges $\eta \in [0.5, 2]$, $k \in [3, 7]$, and $\tau_p \in [0.5, 2]$, the maximum absolute deviations from the default accuracy are 1.32 percentage points under IID and 3.24 points under Separate ordering.

A.8 Graph-Structure Controls

To verify that the gains of GRASP-Base arise from meaningful neighborhood structure rather than message passing alone, we replace the cosine kNN graph with random and shuffled alternatives under the *Separate* stream. Table 10 shows substantial accuracy drops across all six datasets when the graph structure is randomized or shuffled. These results suggest that GRASP-Base benefits from meaningful local relationships.

A.9 Batch-Size Sensitivity

Because GRASP constructs its relational graph within each test batch, batch size determines the samples available for relational refinement. We vary B from 8 to 1024 while keeping all other

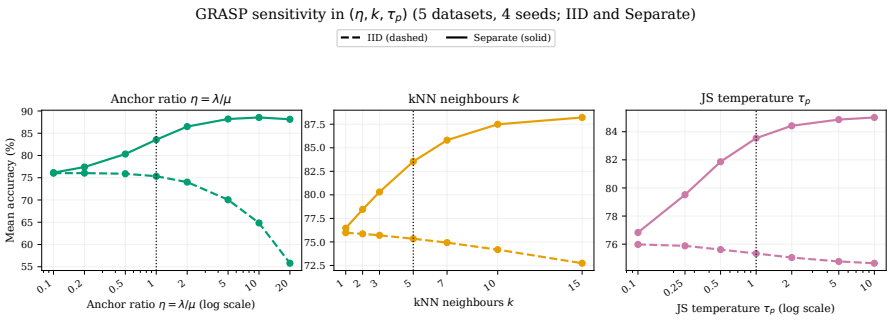


Figure 4: GRASP hyperparameter sensitivity on five datasets and four random seeds. Solid and dashed curves denote Separate and IID streams, respectively; each hyperparameter is varied independently around $(\eta, k, \tau_p) = (1, 5, 1)$.

Table 9: GRASP hyperparameter sensitivity. Mean accuracy (%) under IID and Separate streams. Each parameter is varied independently around $(\eta, k, \tau_p) = (1, 5, 1)$; default values are shown in bold.

$\eta = \lambda/\mu$			k			τ_p		
Value	IID	Sep.	Value	IID	Sep.	Value	IID	Sep.
0.1	76.03	76.15	1	75.98	76.46	0.1	75.98	76.83
0.2	76.04	77.40	2	75.86	78.45	0.25	75.88	79.52
0.5	75.89	80.32	3	75.70	80.30	0.5	75.61	81.87
1	**75.34**	**83.54**	**5**	**75.34**	**83.54**	**1**	**75.34**	**83.54**
2	74.02	86.49	7	74.93	85.81	2	75.05	84.42
5	70.05	88.19	10	74.18	87.49	5	74.77	84.86
10	64.84	88.54	15	72.74	88.22	10	74.64	85.01
20	55.79	88.14						

hyperparameters fixed. Results are averaged over five datasets (EuroSAT, UCF-101, ImageNet, Food-101, and Caltech-101) and four random seeds, under IID and Separate streams.

Table 11 shows a smooth trade-off with batch size. Increasing B from 8 to 1024 raises IID accuracy from 73.78% to 76.26%, while Separate accuracy decreases from 85.52% to 82.02%. Around the default $B = 128$, performance is relatively stable: changing to $B = 64$ or $B = 256$ changes IID accuracy by only -0.36 and $+0.18$ points, and Separate accuracy by $+0.63$ and -0.53 points, respectively. Thus, the default is not narrowly tuned, although batch size naturally influences the balance between IID and structured-stream performance.

A.10 Natural Temporal Structure

We evaluate UCF-101 using the same frame pool under two orderings. Natural randomizes video-clip order while preserving temporal order within each clip; Shuffled globally permutes the same frames. Table 12 reports the full comparison over four random seeds. The Natural-Shuffled difference is positive for GRASP in all four seeds. GRASP-Base exhibits a much larger ordering gap, while ECALP changes by only -0.40 points and its paired 95% confidence interval spans zero.

Graph	Caltech	Cars	DTD	Pets	SUN-397	UCF-101
Cosine kNN (GRASP-Base)	**96.06**	**78.17**	**50.59**	**95.33**	**83.16**	**76.69**
Random graph	65.17	46.60	33.34	71.84	50.36	47.08
Shuffled graph	59.45	45.83	33.15	72.28	49.83	47.22

Table 10: Accuracy (%) for GRASP-Base with cosine, random, and shuffled graphs under the *Separate* stream.

Table 11: Batch-size sensitivity of GRASP on five datasets and four random seeds. All other hyperparameters are fixed.

Batch size B	8	16	32	64	128	256	512	1024
IID	73.78	74.23	74.55	74.98	75.34	75.52	75.95	76.26
Separate	85.52	85.36	84.69	84.17	83.54	83.01	82.50	82.02

Table 12: Natural temporal-stream evaluation on UCF-101. Natural randomizes clip order while preserving within-clip temporal order; Shuffled globally permutes the same frames. Values are mean $\pm$ standard deviation over four seeds; Δ is the paired Natural-Shuffled difference.

Method	Natural (%)	Shuffled (%)	Δ (pp)
TDA	69.07 ± 0.28	68.62 ± 0.13	$+0.45 \pm 0.16$
GRASP-Base	72.88 ± 0.38	52.32 ± 0.15	$+20.56 \pm 0.49$
StatA	67.86 ± 0.23	63.58 ± 0.03	$+4.28 \pm 0.25$
ECALP	71.03 ± 0.31	71.43 ± 0.04	-0.40 ± 0.26
GRASP	72.98 ± 0.35	66.99 ± 0.02	$+5.99 \pm 0.36$

A.11 Runtime cost

We measure adaptation latency after the shared frozen VLM forward pass using precomputed features and logits, with float32 inference on a single Tesla V100 32GB GPU. At $B = 128$, GRASP requires 0.207 ± 0.037 ms per sample, compared with 0.145 ± 0.024 ms for TDA and 0.167 ± 0.025 ms for GRASP-Base (Table 13). Thus, the batch-local graph construction and anchored refinement add 0.062 ms per sample over TDA. The relational graph is discarded after each batch, so the only state retained across the test stream is the fixed-capacity positive and negative cache.

A.12 Generalization across backbones

We additionally evaluate ViT-B/32, ViT-L/14, and SigLIP-L using the same fixed GRASP configuration. For SigLIP, we use the ViT-SO400M-14-SigLIP checkpoint with the webli pretrained weights from OpenCLIP and resize inputs to 224×224 pixels.

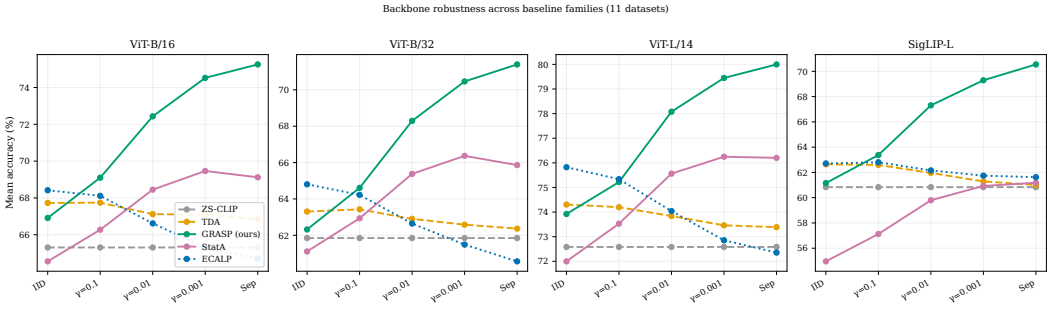


Figure 5: **Cross-regime generalization across VLM backbones.** GRASP uses the same fixed configuration for all backbones and achieves the highest cross-regime mean among the five evaluated methods in each case. Separate gains over TDA range from 6.61 to 9.54 percentage points.

Table 13: Post-encoding adaptation latency at B=128, averaged over five datasets, IID and Separate streams, and four random seeds.

Method	ms/sample ↓	Relative to TDA
TDA	0.145 ± 0.024	$1.00\times$
GRASP-Base	0.167 ± 0.025	$1.15\times$
GRASP	0.207 ± 0.037	$1.43\times$

A.13 Operator Ablation: Per-Dataset Results and Closed-Form Equivalence

This section complements Section 5.3 with the per-dataset results and the fixed-point derivation underlying the *converged propagation* step.

Closed-form equivalence GRASP-Base applies the recurrence

$$\mathbf{F}^{(t+1)} = \alpha \mathbf{P}_{\text{cache}} + (1-\alpha)\mathbf{S}\mathbf{F}^{(t)}, \qquad \alpha = \frac{1}{1+\lambda}, \tag{35}$$

where $\mathbf{S} = \mathbf{D}^{-1}\mathbf{W}$ is the row-normalized propagation matrix. GRASP-Base uses only $T = 3$ iterations. If the same update is instead run to convergence, its fixed point satisfies

$$\mathbf{F}^{*}_{\text{Base}} = \alpha \left[\mathbf{I} - (1-\alpha)\mathbf{S}\right]^{-1} \mathbf{P}_{\text{cache}} = \left[\mathbf{I} + \lambda(\mathbf{I} - \mathbf{S})\right]^{-1} \mathbf{P}_{\text{cache}}. \tag{36}$$

Thus, the second row of Table 2 changes only the number of propagation steps, from $T = 3$ to the converged solution. The subsequent rows then modify the refinement operator itself, allowing the contribution of each change from GRASP-Base to GRASP to be evaluated separately.

Per-dataset results Tables 14 and 15 report the five ablation stages separately for each dataset under IID and Separate streams. The anchored random-walk Laplacian formulation produces the largest overall recovery under IID, although the magnitude varies across datasets. Switching to the unnormalized Laplacian then recovers structured-stream performance on several datasets; for example, EuroSAT improves from 62.56% to 70.44% under Separate. The final posterior-consistent weighting makes a smaller adjustment, consistent with the aggregate results in Table 2.

Table 14: Per-dataset accuracy (%) for the operator ablation ladder, IID stream (4 datasets, 10 seeds).

Variant	DTD	Caltech-101	Flowers-102	EuroSAT	Avg
GRASP-Base (T=3)	35.60	71.17	55.46	58.95	55.29
→ converged propagation	38.20	79.26	64.08	56.43	59.49
→ anchored solve (random-walk Laplacian)	45.84	93.60	72.20	61.86	68.38
→ anchored solve (unnormalized Laplacian)	44.14	91.92	70.46	60.57	66.77
→ posterior-consistent edges (GRASP)	44.99	93.32	71.42	61.00	67.68

Table 15: Per-dataset accuracy (%) for the operator ablation ladder, Separate (structured) stream (4 datasets, 10 seeds).

Variant	DTD	Caltech-101	Flowers-102	EuroSAT	Avg
GRASP-Base (T=3)	50.59	96.06	78.43	70.99	74.02
→ converged propagation	50.82	95.71	78.50	75.32	75.09
→ anchored solve (random-walk Laplacian)	48.14	94.64	75.31	62.56	70.16
→ anchored solve (unnormalized Laplacian)	51.34	96.00	78.19	70.44	73.99
→ posterior-consistent edges (GRASP)	50.90	95.69	77.77	69.23	73.40

A.14 Complementarity Across Test-Time Adapters

We test whether the GRASP relational operator remains effective when applied to alternative test-time adapters rather than the default TDA cache. For each adapter A, its native adaptation procedure and state updates are preserved, while its posterior $\mathbf{P}_{\text{A}}$ replaces $\mathbf{P}_{\text{cache}}$ as both the refinement anchor and the source of posterior-consistency weights:

$$r_{ij} = \exp\left(-\frac{\text{JS}(\mathbf{p}^{\text{A}}_i, \mathbf{p}^{\text{A}}_j)}{\tau_p}\right).$$

The resulting graph uses the same cosine kNN construction and unnormalized Laplacian as GRASP, and refinement solves

$$(\mu\mathbf{I} + \lambda\mathbf{L})\mathbf{F}^{*} = \mu\mathbf{P}_{\text{A}}.$$

All relational hyperparameters remain unchanged: k=5, τ_p=1, and μ=λ=10 (η=1). Refined predictions are used only as outputs and are not fed back into the underlying adapter.

DOTA DOTA retains its original sequential adaptation pipeline. For each image, it forms its fused prediction from the CLIP logits and online distributional estimate, then updates its running class statistics before processing the next sample. We collect the resulting DOTA features and fused posteriors in a rolling buffer of $B{=}128$ samples and apply GRASP when the buffer is full. Thus, DOTA's sample-wise updates remain unchanged and are decoupled from the batch-local relational refinement.

ADAPT ADAPT likewise retains its native sample-wise online adaptation. For each image, it updates its per-class knowledge bank, re-estimates its closed-form Gaussian discriminant analysis (GDA) parameters when the bank changes, and forms its native fused prediction from the resulting adaptation state. We collect the corresponding CLIP features and fused ADAPT posteriors in a rolling buffer of $B{=}128$ samples and apply GRASP when the buffer is full. The refined predictions do not modify ADAPT's knowledge bank or subsequent state updates.

Table 16: Complementarity of GRASP across test-time adapters. DOTA and DOTA+GRASP, and ADAPT and ADAPT+GRASP, use matched sets of three random seeds. GRASP (TDA)† is reproduced from the 10-seed main evaluation as the default-method reference. Values are mean accuracy (%) across 11 datasets. Δ (+GRASP) denotes the matched change from adding GRASP to the corresponding upstream adapter. Differences are computed from unrounded matched results.

Method	IID	$\gamma{=}0.1$	$\gamma{=}0.01$	$\gamma{=}0.001$	Separate	Worst	Mean
GRASP (TDA)†	66.91	69.10	72.43	74.52	75.25	66.91	71.64
DOTA	69.52	69.19	68.72	68.33	68.11	68.11	68.77
DOTA + GRASP	69.05	70.04	72.69	74.41	75.48	69.05	72.34
Δ (+GRASP)	-0.47	$+0.85$	$+3.97$	$+6.08$	$+7.37$	$+0.94$	$+3.57$
ADAPT	70.58	70.16	69.36	68.55	68.85	68.55	69.50
ADAPT + GRASP	70.51	70.82	71.87	72.41	73.31	70.51	71.78
Δ (+GRASP)	-0.07	$+0.66$	$+2.51$	$+3.86$	$+4.45$	$+1.96$	$+2.28$

Table 16 shows that the gains from relational refinement increase monotonically with stream structure for both adapters. Across both adapters, the contribution of relational refinement increases monotonically with stream structure. Relative to DOTA, the change from GRASP ranges from -0.47 points under IID to $+7.37$ under Separate, increasing the cross-regime mean from 68.77% to 72.34% and worst-regime accuracy from 68.11% to 69.05%. For ADAPT, the corresponding change increases from -0.07 points under IID to $+4.45$ under Separate, while the cross-regime mean rises from 69.50% to 71.78% and worst-regime accuracy from 68.55% to 70.51%. Thus, the same GRASP operator causes only a small decrease in IID accuracy to these stronger upstream adapters while contributing progressively more as local stream structure increases. This consistent pattern across two distinct adapters supports the portability of GRASP's relational refinement beyond the default TDA cache.