CHIPS: Efficient CLIP Adaptation via Curvature-aware Hybrid Influence-based Data Selection

Supplementary Material

Appendix Contents

A. Related Work

CLIP Adaptation. Current approaches for effective CLIP adaptation to specialized vertical domains (e.g., Medical) can be broadly categorized into two paradigms: model-centric and data-centric methods [66]. Model-centric approaches primarily focus on developing novel training strategies, including probabilistic fine-tuning [25] and many-to-many contrastive learning [24], as well as parameter-efficient fine-tuning (PEFT) techniques [18, 46, 58, 65]. In contrast, data-centric approaches emphasize the collection of large-scale domain-specific datasets for continual pre-training. Within the general-purpose medical domain, PMC-CLIP [34] leveraged approximately 2M samples, BioMedCLIP [64] curated 15M samples, BIOMEDICA [36] assembled 24M samples, and MedTrinity aggregated 25M samples. For more specialized medical subdomains, QUILT [23] compiled 1M samples for Histopathology, EYECLIP [48] gathered approximately 3M samples for Ophthalmology, and PanDerm [59] curated 2M samples for Dermatology. In the biology and biodiversity domains, even larger-scale datasets ranging from 5M to 214M samples have been collected [4, 15, 17, 50, 54, 60]. In this work, rather than pursuing resource-intensive dataset scaling, we investigate the intrinsic characteristics of existing domain-specific datasets and employ strategic data selection to achieve more efficient CLIP adaptation.

Data Attribution. Data attribution methods quantify the influence of individual training samples on model training efficacy. [28] initially introduced Influence Functions (IF), proposing the LiSSA approximation to efficiently compute the inverse Hessian-vector product, wherein the computation entails calculating the Hessian inverse, iterating through validation and training samples, and deriving influence scores through gradient-based calculations. [43] proposed TracIn, which offers a more intuitive first-order approximation that tracks model training dynamics across multiple checkpoints. [42] introduced EL2N, a self-influence metric that computes the L2 norm of prediction error vectors without requiring explicit validation samples. [47] proposed the Arnoldi iteration method, which leverages dominant eigenvalue decomposition to address the computational overhead and non-H-invariance issues inherent in random projection approaches. [27] introduced GEX, reformulating IF to capture bimodal influence distributions by replacing gradient-based Taylor approximations with direct loss multiplications and employing ensemble models to mitigate Hessian singularity bias. [62] proposed FVM, which optimizes for flat validation minima to enhance the stability of influence estimations for mislabeled sample detection. In this work, we propose an IF variant specifically tailored for CLIP influence estimation.

B. Proofs

This section provides consolidated derivations and proofs for the three core components referenced in Sec. 2.1, Sec. 2.2, and Sec. 2.3. Unless stated otherwise, expectations are taken over the mini-batch sampling and data randomness at the current parameter iterate θ. We write $\widehat{\mathbf{g}} = \frac{1}{B}\sum_{i=1}^{B} \nabla_\theta \ell(z_i;\theta)$, $\mathbf{g}_q = \mathbb{E}_{z\sim q}[\nabla_\theta \ell(z;\theta)]$, $\mathbf{u} = \nabla_\theta \mathcal{L}_{\text{eval}}(\theta)$, and use the Euclidean inner product and norm. Spectral and Frobenius norms are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$.

B.1. Alignment Direction

In a neighborhood of θ assume either (i) $\mathcal{L}_{\text{eval}} \in C^2$ with Hessian-Lipschitz constant L_H, or (ii) $\mathcal{L}_{\text{eval}}$ is ρ-smooth.

Second-order upper bound. For $\Delta\theta$, the second-order Taylor expansion with integral remainder gives

$$\mathcal{L}_{\text{eval}}(\theta+\Delta\theta) = \mathcal{L}_{\text{eval}}(\theta) + \mathbf{u}^\top\Delta\theta + \tfrac{1}{2}\,\Delta\theta^\top \mathbf{H}_{\text{eval}}(\Theta)\,\Delta\theta + R_3, \tag{13}$$

where Θ lies on the segment between θ and $\theta+\Delta\theta$, and $|R_3| \le \frac{L_H}{6}\|\Delta\theta\|^3$. With the one-step SGD move $\Delta\theta = -\eta\,\widehat{\mathbf{g}}$ and taking expectation,

$$\begin{aligned}\mathbb{E}\big[\Delta\mathcal{L}_{\text{eval}}\big] \le\; & -\eta\,\mathbf{g}_q^\top\mathbf{u} \\ & + \tfrac{1}{2}\,\eta^2\,\mathbb{E}\big[\widehat{\mathbf{g}}^\top\mathbf{H}_{\text{eval}}(\Theta)\widehat{\mathbf{g}}\big] \\ & + \tfrac{L_H}{6}\,\eta^3\,\mathbb{E}\big[\|\widehat{\mathbf{g}}\|^3\big],\end{aligned}$$

which matches Eq. (1) in the main text and shows that descent is driven by the alignment term $-\mathbf{g}_q^\top\mathbf{u}$.

First-order upper bound. If $\mathcal{L}_{\text{eval}}$ is ρ-smooth, the Descent Lemma yields

$$\mathcal{L}_{\text{eval}}(\theta+\Delta\theta) \le \mathcal{L}_{\text{eval}}(\theta) + \mathbf{u}^\top\Delta\theta + \tfrac{\rho}{2}\|\Delta\theta\|^2. \tag{14}$$

Setting $\Delta\theta = -\eta\,\widehat{\mathbf{g}}$ and taking expectations gives Eq. (2).

Reegarding mini-batch variance, let $\Sigma_q = \mathrm{Cov}_{z\sim q}[\nabla_\theta\ell(z;\theta)]$, then

$$\begin{aligned}\mathbb{E}\big[\|\widehat{\mathbf{g}}\|^2\big] &= \|\mathbf{g}_q\|^2 + \tfrac{1}{B}\,\mathrm{tr}(\Sigma_q), \\ \mathbb{E}\big[\|\widehat{\mathbf{g}} - \mathbf{g}_q\|^2\big] &= \tfrac{1}{B}\,\mathrm{tr}(\Sigma_q).\end{aligned} \tag{15}$$

This makes explicit how the batch size B moderates the quadratic term in Eq. (2). In a local quadratic model the steepest descent direction for $\mathcal{L}_{\text{eval}}$ is $\mathbf{H}_{\text{eval}}^{-1}\mathbf{u}$. Hence a selection distribution q that increases $\mathbf{g}_q^\top\mathbf{H}_{\text{eval}}^{-1}\mathbf{u}$ is desirable, which motivates the proxy alignment in Sec. 2.2.

AdamW-aware form. At iteration t, let $\mathbf{g}_t = \frac{1}{B}\sum_{i=1}^{B}\nabla_\theta\ell(z_i;\theta_t)$. Consider Adam moments and bias corrections as the following forms

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1)\mathbf{g}_t, \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2)\,(\mathbf{g}_t \odot \mathbf{g}_t), \\ \hat{m}_t &= \frac{m_t}{1-\beta_1^t}, \\ \hat{v}_t &= \frac{v_t}{1-\beta_2^t}, \\ \mathbf{P}_t &= \mathrm{diag}\left((\sqrt{\hat{v}_t}+\epsilon)^{-1}\right),\end{aligned}$$

where $\odot$ denotes the element-wise product. With decoupled weight decay $w_d > 0$ and diagonal mask $\mathbf{D}$, the update is

$$\Delta\theta_t \triangleq \theta_{t+1} - \theta_t = -\eta_t\big(\mathbf{P}_t\hat{m}_t + w_d\,\mathbf{D}\,\theta_t\big).$$

Under the C^2 and Hessian-Lipschitz assumption, Eq. (13) gives

$$\begin{aligned}\mathbb{E}\big[\Delta\mathcal{L}_{\text{eval}}\big] \le\; & -\eta_t\,\mathbb{E}\big[\hat{m}_t^\top\mathbf{P}_t\mathbf{u}\big] - \eta_t\,w_d\,(\mathbf{D}\theta_t)^\top\mathbf{u} \\ & + \tfrac{1}{2}\,\eta_t^2\,\mathbb{E}\big[\Delta\theta_t^\top\mathbf{H}_{\text{eval}}(\Theta_t)\Delta\theta_t\big] \\ & + \tfrac{L_H}{6}\,\eta_t^3\,\mathbb{E}\big[\|\Delta\theta_t\|^3\big],\end{aligned}$$

Moreover, if under ρ-smoothness, Eq. (14) yields

$$\begin{aligned}\mathbb{E}\big[\Delta\mathcal{L}_{\text{eval}}\big] \le\; & -\eta_t\,\mathbb{E}\big[(\mathbf{P}_t\hat{m}_t)^\top\mathbf{u}\big] \\ & - \eta_t\,w_d\,(\mathbf{D}\theta_t)^\top\mathbf{u} \\ & + \tfrac{\rho}{2}\,\eta_t^2\,\mathbb{E}\big[\|\Delta\theta_t\|^2\big].\end{aligned}$$

In both cases the alignment term becomes $\mathbb{E}[(\mathbf{P}_t\hat{m}_t)^\top\mathbf{u}]$, which is in line with SDG in the main content of this paper.

B.2. Proxy Alignment in Subspace

In this section, we prove Theorem 1 and state two practical corollaries.

In the end-point subspace let $\mathbf{g} = \mathbf{g}_\vartheta(z)$, $\boldsymbol{\mu} = \mathbb{E}[\mathbf{g}]$, $\mathbf{g}_c = \mathbf{g} - \boldsymbol{\mu}$, and $\Sigma_g = \mathrm{Cov}[\mathbf{g}] \succeq 0$. Locally,

$$\nabla_\theta\ell(z) = \mathbf{J}\,\mathbf{g} + \mathbf{r}(z), \qquad \mathbf{u} = \bar{\mathbf{J}}\,\mathbf{u}_\vartheta + \boldsymbol{\varepsilon},$$

with $\mathbf{J}, \bar{\mathbf{J}}$ linear maps, residual $\mathbf{r}(z)$, and deterministic mismatch $\boldsymbol{\varepsilon}$. Define

$$\mathbf{S} = \tfrac{1}{2}(\mathbf{J}^\top\bar{\mathbf{J}} + \bar{\mathbf{J}}^\top\mathbf{J}), \quad \mathbf{A} = \tfrac{1}{2}(\mathbf{J}^\top\bar{\mathbf{J}} - \bar{\mathbf{J}}^\top\mathbf{J}),$$

and consider

$$X(z) = \mathbf{u}_\vartheta^\top\mathbf{g}_c, \qquad Y(z) = (\nabla_\theta\ell(z))^\top\mathbf{u}.$$

Collect the remaining random terms into

$$\zeta(z) = (\mathbf{J}\mathbf{g}_c)^\top\boldsymbol{\varepsilon} + \mathbf{r}(z)^\top\bar{\mathbf{J}}\mathbf{u}_\vartheta + \mathbf{r}(z)^\top\boldsymbol{\varepsilon} + \mathbf{g}_c^\top\mathbf{A}\,\mathbf{u}_\vartheta,$$

and write $\sigma_\zeta^2 = \mathrm{Var}[\zeta(z)]$.

Proof of Theorem 1. Let $\tilde{\mathbf{g}} = \Sigma_g^{-1/2}\mathbf{g}_c$, $\alpha = \Sigma_g^{1/2}\mathbf{u}_\vartheta$, $B = \Sigma_g^{1/2}\mathbf{S}\Sigma_g^{-1/2}$, and $Z(z) = \mathbf{g}_c^\top\mathbf{S}\mathbf{u}_\vartheta = \alpha^\top B\tilde{\mathbf{g}}$. If ζ is uncorrelated with $\tilde{\mathbf{g}}$ then $\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,Z)$ and $\mathrm{Var}(Y) = \mathrm{Var}(Z) + \sigma_\zeta^2$. Using

$$\begin{aligned}\mathrm{Var}(X) &= \|\alpha\|_2^2, \\ \mathrm{Cov}(X,Z) &= \alpha^\top\mathrm{sym}(B)\,\alpha, \\ \mathrm{Var}(Z) &\le \|B\|_2^2\,\|\alpha\|_2^2.\end{aligned}$$

and the Rayleigh-Ritz principle, gives

$$\rho_{XY} \;\geq\; \frac{\lambda_{\min}\big(\mathrm{sym}(B)\big)}{\sqrt{\|B\|_2^2 + \sigma_\zeta^2/(\mathbf{u}_\vartheta^\top \Sigma_g \mathbf{u}_\vartheta)}}, \tag{16}$$

which is Eq. (5) in the main content after identifying B. Without the uncorrelatedness assumption, Cauchy-Schwarz implies the fallback bound

$$\rho_{XY} \;\geq\; \frac{\lambda_{\min}(\mathrm{sym}(B))\,\|\alpha\|_2 - \sigma_\zeta}{\|B\|_2\,\|\alpha\|_2 + \sigma_\zeta}. \tag{17}$$

B.3. Negative Pair Curvature

In the projection-temperature subspace, we write $\mathbf{g}(z) = \mathbf{g}_\vartheta(z)$ and $\mathbf{u} = \mathbf{u}_\vartheta$. For softmax-type losses, a generalized Gauss-Newton curvature admits the population decomposition

$$\begin{aligned} \mathbf{H}_\vartheta &= \Phi_{\text{pos}} + \Phi_{\text{neg}}, \\ \Phi_{\text{pos}} &= \mathbb{E}_z\big[\mathbf{g}(z)\mathbf{g}(z)^\top\big], \\ \Phi_{\text{neg}} &= \mathbb{E}_{z\neq z'}\big[\mathbf{g}(z)\mathbf{g}(z')^\top\big], \end{aligned} \tag{18}$$

which mirrors the random-negative mechanism of InfoNCE. Note that $\Phi_{\text{pos}} \succeq 0$ while Φ_{neg} need not be PSD, and the sum $\mathbf{H}_\vartheta$ is PSD.

For a mini-batch $\{z_i\}_{i=1}^B$ let $\mathbf{g}_i = \mathbf{g}_\vartheta(z_i)$, then

$$\widehat{\Phi}_{\text{pos}} = \tfrac{1}{B}\sum_{i=1}^{B} \mathbf{g}_i\mathbf{g}_i^\top, \qquad \widehat{\Phi}_{\text{neg}} = \tfrac{1}{B(B-1)}\sum_{i\neq j} \mathbf{g}_i\mathbf{g}_j^\top. \tag{19}$$

As in Eq. (7), define

$$\mathbf{H}_\vartheta^{(\alpha)} = (1-\alpha)\,\Phi_{\text{pos}} + \alpha\,\Phi_{\text{neg}}, \qquad \mathbf{M} = \mathbf{H}_\vartheta^{(\alpha)} + \lambda\,\mathbf{I}, \tag{20}$$

with $\alpha \in [0,1]$ and $\lambda > 0$. Let $A^\star(z) = \mathbf{g}(z)^\top \mathbf{H}_\vartheta^{-1}\mathbf{u}$, $A_\alpha(z) = \mathbf{g}(z)^\top \mathbf{M}^{-1}\mathbf{u}$, and $\Delta_\alpha = \mathbf{H}_\vartheta - \mathbf{H}_\vartheta^{(\alpha)}$. Using the resolvent identity,

$$\begin{aligned} \mathbf{M}^{-1} - \mathbf{H}_\vartheta^{-1} &= \mathbf{H}_\vartheta^{-1}\,(\Delta_\alpha - \lambda\mathbf{I})\,\mathbf{M}^{-1}, \\ \big\|\mathbf{M}^{-1} - \mathbf{H}_\vartheta^{-1}\big\|_2 &\leq \|\mathbf{H}_\vartheta^{-1}\|_2\,\|\Delta_\alpha\|_2\,\|\mathbf{M}^{-1}\|_2. \end{aligned} \tag{21}$$

which leads to

$$\mathbb{E}_z\big[(A_\alpha - A^\star)^2\big] \leq C_2\,\|\Delta_\alpha\|_F^2\,\|\mathbf{H}_\vartheta^{-1}\mathbf{u}\|_2^2,$$

for a constant $C_2 > 0$ depending only on $\|\Phi_{\text{pos}}\|_2$, $\|\mathbf{H}_\vartheta^{-1}\|_2$, and $\|\mathbf{M}^{-1}\|_2$. Whenever cross-example coupling is present, any $\alpha > 0$ that injects off-diagonal mass reduces $\|\Delta_\alpha\|_F$. Let $\Pi_k \in \mathbb{R}^{k\times d_\vartheta}$ be a JL transform. For any fixed $\mathbf{a}, \mathbf{b}$, $|\mathbf{a}^\top\mathbf{b} - (\Pi_k\mathbf{a})^\top(\Pi_k\mathbf{b})| \leq \varepsilon\|\mathbf{a}\|\,\|\mathbf{b}\|$ holds with probability at least $1-\delta$, where $k = \Omega(\varepsilon^{-2}\log(1/\delta))$. Define $\mathbf{M}_k = \Pi_k\,\mathbf{M}\,\Pi_k^\top$ and $\bar{\mathbf{u}}_k = \mathbf{M}_k^{-1}\Pi_k\mathbf{u}$. The sketched score

$$\widehat{A}_\alpha(z) = (\Pi_k\mathbf{g}(z))^\top\bar{\mathbf{u}}_k$$

satisfies $\mathbb{E}_z[(\widehat{A}_\alpha - A_\alpha)^2] \leq C_1\log(1/\delta)/k$ for a constant $C_1 > 0$ depending on $\|\mathbf{M}^{-1}\|_2$, $\|\mathbf{u}\|_2$, and $\mathbb{E}\|\mathbf{g}(z)\|_2^2$. Combining with the mixing bias via the triangle inequality yields Eq. (8) in the main content:

$$\mathbb{E}_z\big[\big(\widehat{A}_\alpha(z) - A^\star(z)\big)^2\big] \leq \underbrace{\frac{C_1\log(1/\delta)}{k}}_{\text{projection variance}} + \underbrace{C_2\,\|\Delta_\alpha\|_F^2\,\|\mathbf{H}_\vartheta^{-1}\mathbf{u}_\vartheta\|_2^2}_{\text{curvature bias}}. \tag{22}$$

C. Data Selection Implementation Details

This section provides all components needed to reproduce CHIPS and baseline methods.

C.1. Problem Setup

Let the training pool be $\mathcal{D}_{\text{pool}} = \{(x_i, y_i)\}_{i=1}^N$ and the evaluation set be $\mathcal{D}_{\text{eval}} = \{(x_m, y_m)\}_{m=1}^M$. Given a data retention ratio $r \in (0,1]$, we select $n = \lfloor r \times |\mathcal{D}_{\text{pool}}| \rfloor$ samples from $\mathcal{D}_{\text{pool}}$ to perform CPT. CHIPS operates in CLIP's end-point subspace $\vartheta = \{\mathbf{W}_v, \mathbf{W}_t, \tau\}$ where $\mathbf{W}_v$ and $\mathbf{W}_t$ are the projection heads for vision and text, and $\tau > 0$ is the logit scale. For computational efficienciy, backbone encoders are used to produce features and are kept fixed for computing selection utility scores (we only need training dynamics of ϑ instead of the complete model).

Let $\mathbf{h} = \mathrm{CLIP}_{\text{img}}(x)$ and $\mathbf{t} = \mathrm{CLIP}_{\text{txt}}(y)$ be backbone features. Define L2-normalized end-point embeddings

$$\hat{\mathbf{x}} = \frac{\mathbf{W}_v\mathbf{h}}{\|\mathbf{W}_v\mathbf{h}\|}, \qquad \hat{\mathbf{y}} = \frac{\mathbf{W}_t\mathbf{t}}{\|\mathbf{W}_t\mathbf{t}\|},$$

and similarities $s_{ij} = \tau\,\hat{\mathbf{x}}_i^\top\hat{\mathbf{y}}_j$. For a batch of size B, write $\mathbf{S} = [s_{ij}]_{i,j=1}^B$ and define the bidirectional softmax probabilities

$$p_{ij}^{i2t} = \frac{\exp(s_{ij})}{\sum_{j'}\exp(s_{ij'})}, \qquad p_{ij}^{t2i} = \frac{\exp(s_{ij})}{\sum_{i'}\exp(s_{i'j})}.$$

The symmetric InfoNCE loss for sample i is

$$\ell_i(\vartheta) = \tfrac{1}{2}\big(\mathrm{CE}(\mathbf{S}_{i,:},\, i) + \mathrm{CE}(\mathbf{S}_{:,i},\, i)\big),$$

where CE denotes the cross-entropy loss.

Let $\mathbf{g}_\vartheta(z) = \nabla_\vartheta\ell(z;\vartheta) \in \mathbb{R}^d$ denote the end-point gradient of a sample, and let

$$\mathbf{u}_\vartheta \;\triangleq\; \mathbb{E}_{z\sim\mathcal{D}_{\text{eval}}}[\mathbf{g}_\vartheta(z)]$$

be the evaluation mean gradient in the same subspace. In practice we maintain $\mathbf{u}_\vartheta$ by an exponential moving average over random evaluation minibatches.

Moreover, to reduce computational cost we use a JL map matrix $\Pi_k \in \mathbb{R}^{k\times d}$ and work with sketched vectors

$\mathbf{g}_k = \Pi_k \mathbf{g}_\vartheta$ and $\mathbf{u}_k = \Pi_k \mathbf{u}_\vartheta$. The inner product distortion satisfies $|\mathbf{a}^\top \mathbf{b} - (\Pi_k \mathbf{a})^\top (\Pi_k \mathbf{b})| \le \varepsilon \|\mathbf{a}\| \, \|\mathbf{b}\|$ with probability at least $1 - \delta$ when $k = \Omega(\varepsilon^{-2} \log(1/\delta))$, which is also used in App. B.3.

C.2. CHIPS

CHIPS ranks each $z \in \mathcal{D}_{\text{pool}}$ by the multiplicative utility

$$\mathcal{I}_{\text{CHIPS}}(z) \;=\; \widehat{A}_\alpha(z) \cdot w_{\text{L}}(z) \cdot w_{\text{R}}(z),$$

then selects the top n samples for CPT. The three factors are implemented as follows and the complete process of CHIPS is provided in Alg. 1.

Curvature-aware proxy alignment. We estimate the curvature in ϑ by mixing self and cross moments from symmetric InfoNCE, consistent with Sec. 2.3. Maintain EMA estimators

$$\Phi_{\text{pos}} = \mathbb{E}[\mathbf{g}_\vartheta(z)\mathbf{g}_\vartheta(z)^\top], \qquad \Phi_{\text{neg}} = \mathbb{E}_{z \neq z'}[\mathbf{g}_\vartheta(z)\mathbf{g}_\vartheta(z')^\top],$$

then build

$$\mathbf{M} \;=\; \mathbf{H}_\vartheta^{(\alpha)} + \lambda \mathbf{I}, \qquad \mathbf{H}_\vartheta^{(\alpha)} \;=\; (1-\alpha)\, \Phi_{\text{pos}} + \alpha\, \Phi_{\text{neg}},$$

with $\alpha \in [0,1]$ and ridge $\lambda > 0$. Without sketching, the alignment score is

$$A_\alpha(z) \;=\; \mathbf{g}_\vartheta(z)^\top \, \mathbf{M}^{-1} \, \mathbf{u}_\vartheta.$$

With a JL sketch, form $\mathbf{M}_k = \Pi_k \mathbf{M} \Pi_k^\top$ once per update and solve $\bar{\mathbf{u}}_k = \mathbf{M}_k^{-1} \mathbf{u}_k$. The sketched score is

$$\widehat{A}_\alpha(z) \;=\; (\Pi_k \mathbf{g}_\vartheta(z))^\top \bar{\mathbf{u}}_k.$$

This matches Eq. (3) and App. B.3.

Learnability. Compute correctness and hardest negative margin using the current batch statistics

$$\begin{aligned} p_{\text{corr}}(z) &= \tfrac{1}{2}\big(p_{ii}^{i2t} + p_{ii}^{t2i}\big), \\ m(z) &= s_{ii} - \max\Big\{ \max_{j \neq i} s_{ij},\, \max_{i' \neq i} s_{i'i} \Big\}. \end{aligned} \tag{23}$$

Define the learnability weight as in Eq. (10)

$$w_{\text{L}}(z) = \big(1 - p_{\text{corr}}(z)\big)\big(1 + \sigma(-m(z))\big),$$

which upweights near-boundary samples and downweights saturated ones.

Target-domain relevance. Let the evaluation centroids be $\boldsymbol{\mu}_x = \mathbb{E}[\hat{\mathbf{x}}]$ and $\boldsymbol{\mu}_y = \mathbb{E}[\hat{\mathbf{y}}]$ over $\mathcal{D}_{\text{eval}}$. Following Eq. (11),

$$w_{\text{R}}(z) = \sigma\Big((1-\beta)\cos(\hat{\mathbf{x}}, \boldsymbol{\mu}_x) + \beta \cos(\hat{\mathbf{y}}, \boldsymbol{\mu}_y)\Big),$$

which softly reweights toward the evaluation domain while keeping the selection drift bounded.

Algorithm 1 CHIPS Algorithm

Require: pool $\mathcal{D}_{\text{pool}}$, evaluation set $\mathcal{D}_{\text{eval}}$, retention n, mix α, ridge λ, JL dim k, relevance balance β

Estimate $\mathbf{u}_\vartheta \leftarrow \mathbb{E}_{z \sim \mathcal{D}_{\text{eval}}}[\mathbf{g}_\vartheta(z)]$ using minibatches; cache $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$
Initialize EMA moments $\Phi_{\text{pos}} \leftarrow \mathbf{0}$, $\Phi_{\text{neg}} \leftarrow \mathbf{0}$
for each minibatch $\{z_i\}_{i=1}^B$ sampled from $\mathcal{D}_{\text{pool}}$ **do**
    compute $\mathbf{g}_i = \mathbf{g}_\vartheta(z_i)$ for all i
    update Φ_{pos} and Φ_{neg} using the batch U-statistics and EMA
end for
$\mathbf{M} \leftarrow (1-\alpha)\Phi_{\text{pos}} + \alpha \Phi_{\text{neg}} + \lambda \mathbf{I}$
if $k > 0$ **then**
    sample JL Π_k, set $\mathbf{M}_k = \Pi_k \mathbf{M} \Pi_k^\top$, solve $\bar{\mathbf{u}}_k = \mathbf{M}_k^{-1}(\Pi_k \mathbf{u}_\vartheta)$
end if
for each $z = (x, y) \in \mathcal{D}_{\text{pool}}$ **do**
    compute $\mathbf{g} = \mathbf{g}_\vartheta(z)$ and $\widehat{A}_\alpha(z) = \begin{cases} \mathbf{g}^\top \mathbf{M}^{-1} \mathbf{u}_\vartheta, & k = 0 \\ (\Pi_k \mathbf{g})^\top \bar{\mathbf{u}}_k, & k > 0 \end{cases}$
    compute $p_{\text{corr}}(z)$ and $m(z)$, then $w_{\text{L}}(z) = \big(1 - p_{\text{corr}}(z)\big)\big(1 + \sigma(-m(z))\big)$
    compute $w_{\text{R}}(z) = \sigma\big((1-\beta)\cos(\hat{\mathbf{x}}, \boldsymbol{\mu}_x) + \beta \cos(\hat{\mathbf{y}}, \boldsymbol{\mu}_y)\big)$
    set $\mathcal{I}_{\text{CHIPS}}(z) = \widehat{A}_\alpha(z)\, w_{\text{L}}(z)\, w_{\text{R}}(z)$
end for
return the top n samples by $\mathcal{I}_{\text{CHIPS}}$

More details. During the computation process,

- $\mathbf{M}^{-1}\mathbf{u}_\vartheta$ is reused across all samples, so the per-sample cost is dominated by $\mathbf{g}_\vartheta(z)$ and a dot product in d or k dimensions.
- τ is kept positive by parameterizing $\tau = \exp(\tilde{\tau})$.
- We rescale $\mathcal{I}_{\text{CHIPS}}$ within each shard of $\mathcal{D}_{\text{pool}}$ if memory constraints require sharded processing.
- We recompute $\mathbf{u}_\vartheta$ periodically to track drift during CPT.

C.3. Heuristics-based Baselines

All heuristics select the top n samples under their scores.

- **Random** samples uniformly at random without replacement.
- **CLIPScore** ranks by CLIPScore [21] computed with the frozen base CLIP model.
- **Concept-Balance** performs probabilistic downsampling to flatten concept frequency. Following [36], we downsample overrepresented concepts such as *Plots and Charts*, *Tables*, and *Scientific Formulae and Equations* at rate 0.25, then sample the remainder uniformly.
- **Concept-Filter** keeps only samples whose metadata contain any of the eight whitelist concepts in [36], (*Clinical Imaging*, *Microscopy*, *Immuno Assays*, *Illustrative Dia-*

grams, *Chemical Structures*, *Maps*, *Tools and Materials*, and *Hand Drawn and Screen Based Visuals*) then samples uniformly from the filtered pool.

C.4. Influence-based Baselines in the End-point Subspace

For fairness and to avoid conflicts with the main text, all influence baselines use the same end-point gradients $\mathbf{g}_\vartheta$ and the same optional JL sketch Π_k as CHIPS. Let $\tilde{\mathbf{g}}_i = \Pi_k \mathbf{g}_\vartheta(z_i)$ if $k > 0$ (use $\tilde{\mathbf{g}}_i = \mathbf{g}_\vartheta(z_i)$ when $k = 0$). Let the evaluation mean gradient be $\bar{\mathbf{g}}_{\text{eval}} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{g}}_m^{\text{eval}}$.

Dot [67]. The first-order directional alignment is

$$\mathcal{I}_{\text{Dot}}(i) = \tilde{\mathbf{g}}_i^\top \bar{\mathbf{g}}_{\text{eval}}.$$

TracIn [43]. Given checkpoints $\{\theta_t\}_{t=0}^{T-1}$ and learning rates $\{\eta_t\}$, we accumulate

$$\mathcal{I}_{\text{TracIn}}(i) = \sum_{t=0}^{T-1} \eta_t \left(\tilde{\mathbf{g}}_i^{(t)}\right)^\top \bar{\mathbf{g}}_{\text{eval}},$$

where $\tilde{\mathbf{g}}_i^{(t)} = \Pi_k \nabla_\vartheta \ell_i(\theta_t)$. We keep $\bar{\mathbf{g}}_{\text{eval}}$ fixed for computational stability.

TRAK [40]. We form a regularized second-moment matrix in the same feature space

$$\mathbf{\Phi} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\mathbf{g}}_i \tilde{\mathbf{g}}_i^\top + \lambda_{\text{TRAK}} \mathbf{I},$$

and compute

$$\mathcal{I}_{\text{TRAK}}(i) = \tilde{\mathbf{g}}_i^\top \mathbf{\Phi}^{-1} \bar{\mathbf{g}}_{\text{eval}}.$$

CHIPS for comparison. CHIPS differs from Dot and TRAK by preconditioning with $\mathbf{M}$ that mixes Φ_{pos} and Φ_{neg} as in Sec. C.2, and by multiplying learnability and relevance weights. In the sketched space,

$$\mathcal{I}_{\text{CHIPS}}(i) = \left(\tilde{\mathbf{g}}_i^\top \mathbf{M}_k^{-1} \mathbf{u}_k\right) w_{\text{L}}(i)\, w_{\text{R}}(i),$$

where $\mathbf{M}_k = \Pi_k\big((1-\alpha)\Phi_{\text{pos}} + \alpha\Phi_{\text{neg}} + \lambda\mathbf{I}\big)\Pi_k^\top$.

D. Score Distribution

The score distributions of CLIPScore, Dot, TracIn, TRAK, and CHIPS are provided in Fig. 5, Fig. 6, Fig. 7, Fig. 8, and Fig. 9.

E. Training

E.1. Software and Distributed Setup

Framework and precision. We train with PyTorch in distributed data-parallel mode. On H200 GPUs we use bfloat16 for tensor computation with an fp32 master copy of parameters for the AdamW states. The logit-scale parameter is maintained as $\tilde{\tau} \in \mathbb{R}$ and realized at runtime by $\tau = \exp(\tilde{\tau})$ to guarantee positivity. Softmax and cross-entropy are evaluated with numerically stable log-sum-exp.

Communication. Gradients are synchronized by NCCL with bucket size tuned to saturate NVLink. We enable gradient accumulation when the per-GPU micro-batch would otherwise exceed memory budget. Let G be the number of GPUs, B_{micro} the per-GPU micro-batch, and A the accumulation steps, then the global batch is

$$B_{\text{global}} = G \times B_{\text{micro}} \times A.$$

We choose $G = 8$, $B_{\text{micro}} = 4096$ and $A = 1$ such that $B_{\text{global}} = 32{,}768$ as in the main content.

Determinism. We fix seeds for Python, NumPy, and CUDA RNGs and set cuDNN to deterministic kernels where available. Dataset sharding and sampler seeds are logged per epoch to make re-runs bitwise reproducible up to nondeterminism in fused kernels.

E.2. Data Preprocessing and Batching

Image pipeline. Training images are decoded to RGB and resized to the default input resolution of each MetaCLIP variant. We apply random resized crop that preserves aspect ratio, followed by horizontal flip with probability 0.5. Pixel intensities are normalized by CLIP mean and variance. Center crop is used at evaluation time.

Text pipeline. Texts are normalized by Unicode and then tokenized by the CLIP tokenizer. Sequences are truncated to the CLIP maximum length (77 tokens) with special tokens preserved.

Sharding and streaming. The training pools are sharded into balanced files on disk to avoid hot spots. Each worker maintains a streaming iterator with prefetching in pinned memory. We keep sharding consistent across methods so that all runs process the same examples at the same step counts.

E.3. Optimization Schedule

Optimizer and parameter groups. We use AdamW with $(\beta_1, \beta_2, \epsilon) = (0.9, 0.98, 10^{-6})$ as stated in the main text. Weight decay is applied to all trainable parameters except LayerNorm and bias parameters.

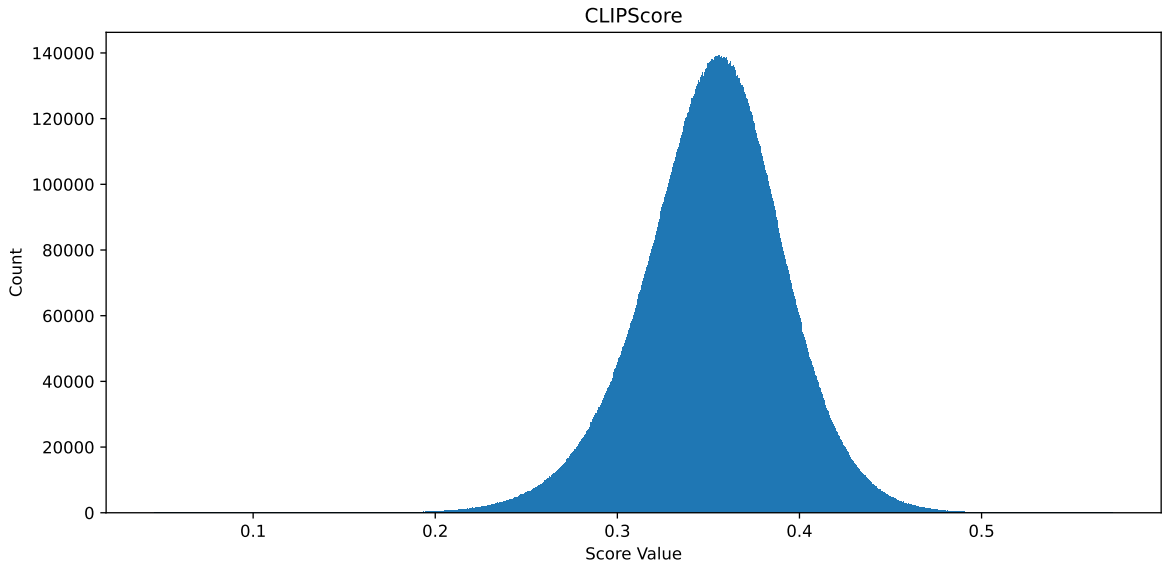

Figure 5. Distribution of CLIPScore on BIOMEDICA.

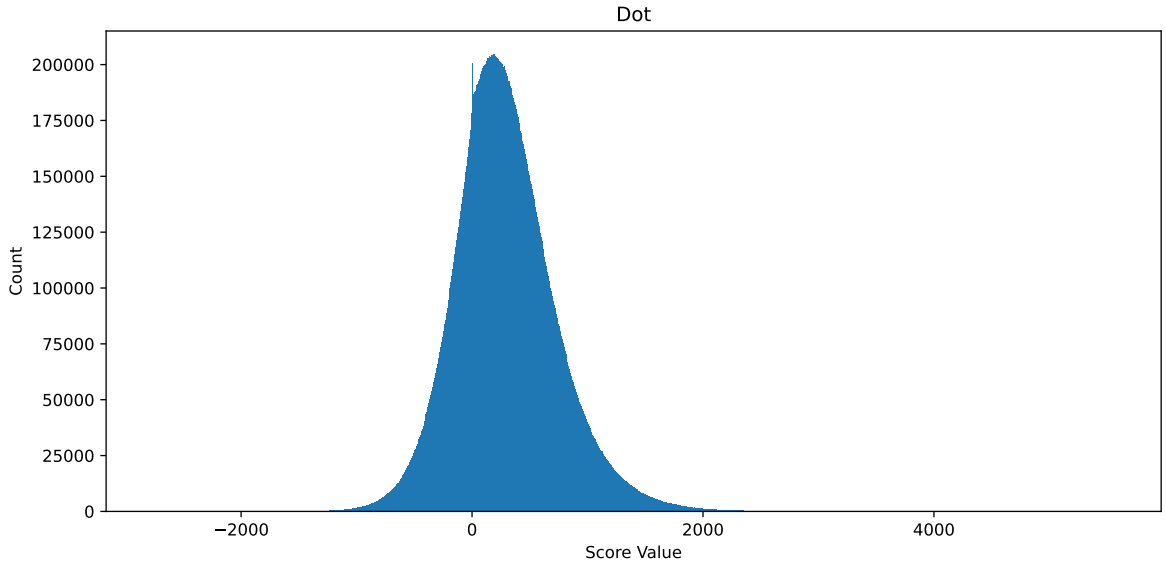

Figure 6. Distribution of Dot on BIOMEDICA.

Learning-rate scheduler. We use a cosine decay scheduler over the total number of optimizer steps. Unless specified otherwise, the peak learning rate equals the initial value 10^{-6}. We optionally use a short linear warmup of T_{warm} steps when training with heavier augmentation or larger B_{global}. The final learning rate floor is set to 0 unless explicit floor is reported.

E.4. Counting Steps and Budget Fairness

Let $n = \lfloor r \times |\mathcal{D}_{\text{train}}| \rfloor$ be the number of retained samples, E the number of epochs, and B_{global} the global batch size. The number of optimizer steps per run is

$$T_{\text{steps}} = \left\lceil \frac{n}{B_{\text{global}}} \right\rceil \times E.$$

All methods, including baselines and CHIPS, are trained for the same T_{steps}. The wall-clock overhead of selection

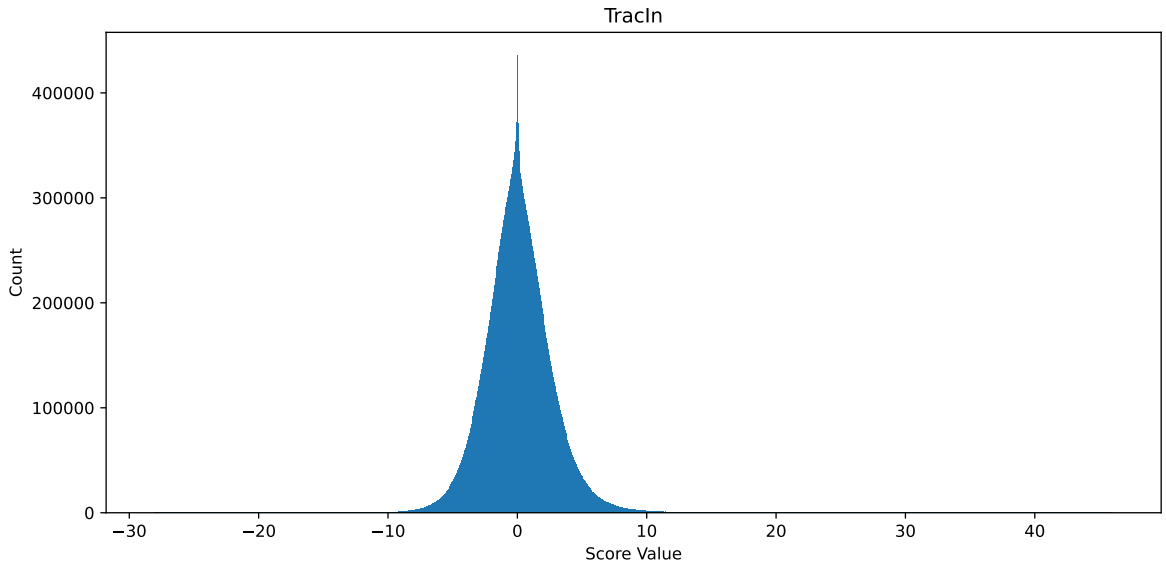

Figure 7. Distribution of TracIn on BIOMEDICA.

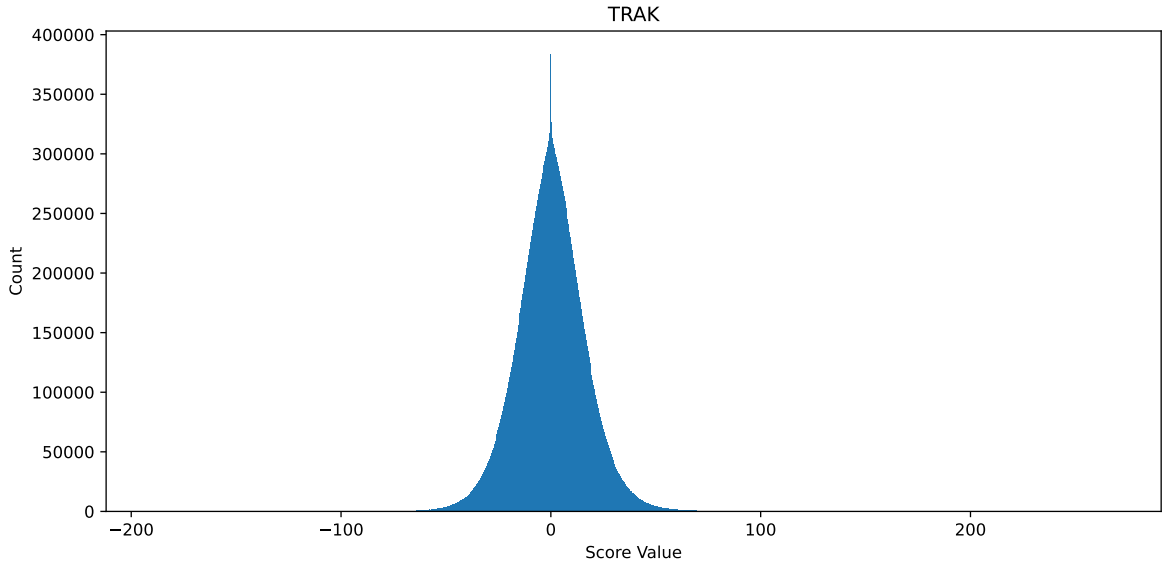

Figure 8. Distribution of TRAK on BIOMEDICA.

is measured separately and reported as a percentage of the total training time.

F. Evaluation

To ensure a comprehensive and reproducible evaluation, we include 48 datasets covering both general-domain and medical-domain visual understanding tasks for this paper. Detailed descriptions are provided below.

F.1. General-Domain Datasets

For General-Domain Datasets, they span a variety of recognition and retrieval tasks, including classification, fine-grained categorization, visual reasoning, robustness evaluation, and multimodal understanding. To be specific, we give an overall description for each as follows:

Object & Scene Classification:

- **ImageNet-1K** [8]: A large-scale object classification benchmark with 1.28 million training images and 1,000

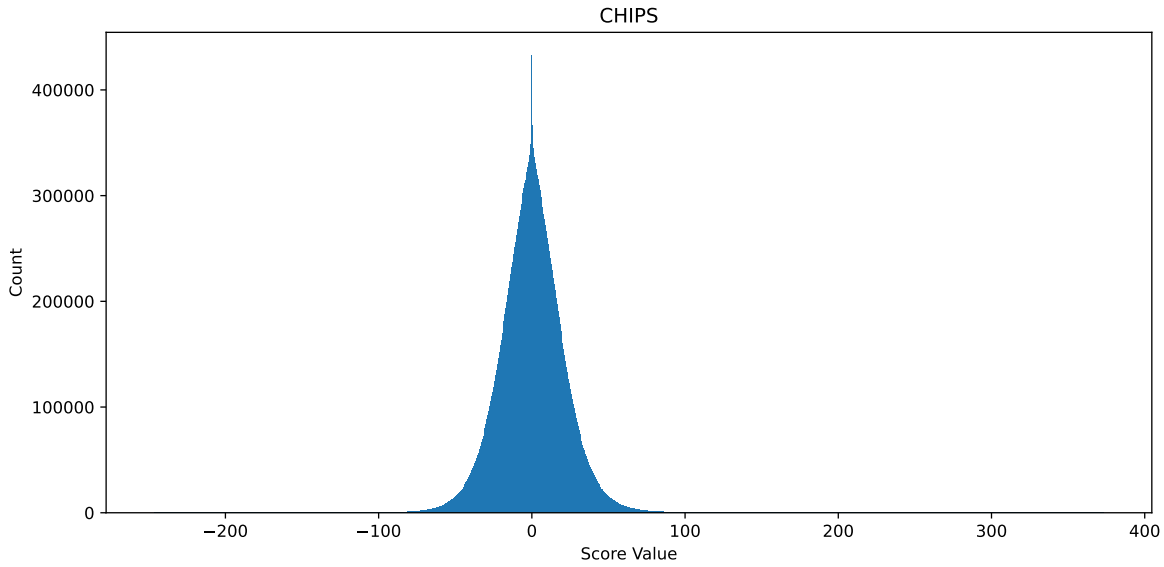

Figure 9. Distribution of CHIPS on BIOMEDICA.

object categories, widely used as a baseline in deep vision.

- **ImageNetV2** [45]: A re-collected test set for ImageNet, designed to assess generalisation under dataset shift for the same 1,000 categories.
- **SUN397** [55]: A scene recognition dataset containing 397 scene categories of diverse indoor/outdoor types, enabling evaluation on scene-level classification.
- **Caltech-101** [13]: A mid-scale object classification benchmark with 101 categories and around 9,000 images, used historically for transfer-learning studies.
- **VOC2007** [12]: The PASCAL Visual Object Classes 2007 dataset, including object classification (and detection) tasks across 20 categories in real-world scenes.

Fine-Grained Recognition:

- **Cars** [29]: A fine-grained vehicle classification dataset containing many car make/model/year classes, evaluating subtle visual differences among similar objects.
- **Aircraft** [37]: A fine-grained aircraft dataset distinguishing among many aircraft models and variants, used for high-granularity recognition research.
- **Food101** [3]: A food image dataset covering 101 food types, with 101,000 images total, for fine-grained categorisation of dishes.
- **Oxford Flowers (Flowers102)** [39]: A flower species classification dataset with 102 categories and 8,000 images, common in fine-grained vision tasks.
- **Pets** [41]: A fine-grained pet breed classification dataset (cats & dogs) with multiple breeds and varied poses/backgrounds.

Texture, Shape, Small-Scale Objects:

- **CIFAR-10/100** [30]: Standard small-scale image classification datasets with 10 (CIFAR-10) and 100 (CIFAR-100) classes, each containing 60,000 colour images of 32×32 size.
- **MNIST** [31]: Classic handwritten digit classification dataset with 70,000 grayscale images of digits 0-9, often used for benchmarking basic vision models.
- **STL-10** [7]: A dataset derived from ImageNet with 10 classes and high-resolution images, used for unsupervised / transfer learning in small-scale settings.
- **smallNORB (Azimuth/Elevation)** [32]: A synthetic 3D object dataset capturing objects under different viewpoints (azimuth/elevation), for studying pose-invariance and representation robustness.
- **SVHN** [38]: Street View House Numbers dataset with real-world digital number images, used for digit recognition in situ.

Robustness & Out-of-Distribution:

- **ImageNet-A** [20]: A subset of ImageNet containing "adversarially filtered" images challenging standard models, designed to test worst-case object recognition robustness.
- **ImageNet-O** [20]: An out-of-distribution test set for ImageNet models, containing images from unknown classes not present in training, to measure OOD detection/generalisation.

Domain-Specific Recognition:

- **KITTI** [14]: A dataset collected for autonomous driving, including object, scene and motion tasks—here used in the classification context of road-scene recognition.
- **EuroSAT** [19]: A remote sensing image classification dataset with 45 scene types, aimed at land-use and

satellite-image recognition.
- **RESISC45** [5]: Another remote sensing scene classification dataset covering 45 classes of aerial images, evaluating models on high-altitude imagery.

Visual Reasoning & Synthetic Tasks:
- **CLEVR (Closest-Object-Distance / Count-All)** [26]: A synthetic benchmark for visual reasoning, where models answer compositional questions about objects (distance/count) in rendered scenes.
- **DTD** [6]: The Describable Texture Dataset, focusing on texture/material recognition with diverse patterns, supporting generalisation in less object-centric tasks.

Image-Text Retrieval / Multimodal:
- **Flickr8k** [22]: An image-caption dataset with 8,000 images, used for evaluating image-to-text and text-to-image retrieval and captioning systems.
- **Flickr30k** [63]: A larger image-caption dataset with 30,000 images, used widely in cross-modal retrieval research.
- **MSCOCO** [33]: A large-scale multimodal dataset with 120K+ images and captions, supporting image detection, segmentation and retrieval tasks.
- **Rendered-SST2** [44]: A dataset created by rendering the sentences from the Stanford Sentiment Treebank v2 into images (positive/negative labels), used to assess optical-character-recognition via image encoders. (Train: 6,920 images; Val: 872; Test: 1,821).

Geographic / Landmark Classification:
- **Country211** [44]: A dataset designed for geolocation classification, filtered from the YFCC100M dataset—211 countries/territories, with 150 train / 50 val / 100 test images per country.

F.2. Medical-Domain Datasets

For Medical-Domain Datasets, we include a broad range of imaging modalities and clinical specialties, covering ophthalmology, radiology, dermatology, hematology, pathology, neuropathology, and non-clinical biology. These datasets collectively assess model performance in clinically relevant scenarios and diagnostic contexts. From a task perspective, they can be broadly grouped into three major categories: *disease classification*, *organ and tissue recognition*, and *pathological image analysis*, with an additional category for *non-clinical biological imaging*. Together, they form a comprehensive benchmark for evaluating generalization and robustness of visual models in biomedical applications. In detail, the brief introduction of included datasets are listed:

Disease Classification:
- **Diabetic Retinopathy** [11]: A fundus-image dataset for grading diabetic retinopathy severity in ophthalmology, used for multi-class disease classification.
- **RetinaMNIST / OCTMNIST** [61]: Retinal fundus and optical coherence tomography (OCT) image datasets for ophthalmic disease classification across multiple categories; part of the MedMNIST benchmark (about 708k 2D images in total for the MedMNIST collection).
- **ChestMNIST** [61]: Based on the NIH-ChestXray14 dataset, ChestMNIST contains 112,120 chest X-ray images, formulated as a multi-label classification task for detecting 14 different diseases.
- **ChestX-ray14** [52]: A large collection of over 100,000 frontal-view chest X-ray images from more than 30,000 patients. The labels for eight common thoracic diseases were automatically generated by text-mining the associated radiological reports.
- **DermaMNIST** [61]: This is a multi-class classification dataset of 10,015 dermatoscopic images for identifying 7 different types of common pigmented skin lesions.
- **BloodMNIST** [61]: A hematology microscope-image dataset containing 17,092 images across 8 classes, used for blood-cell type classification.

Organ and Tissue Recognition:
- **OrganAMNIST / OrganCMNIST / OrganSMNIST** [61]: This dataset is for multi-class classification of 11 body organs, containing 58,850 images derived from axial-view slices of abdominal CT scans and resized to 28x28 pixels.
- **TissueMNIST** [61]: Sourced from the Broad Bioimage Benchmark Collection, TissueMNIST is a large-scale dataset of 236,386 human kidney cortex cells, organized into 8 categories for a multi-class classification task.

Pathological Image Analysis:
- **PCAM** [51]: The dataset is a large-scale collection of 96x96 pixel histopathology image patches extracted from the Camelyon16 challenge, designed for the task of identifying metastatic cancer in lymph node sections.
- **LC25000** [2]: The LC25000 dataset contains 25,000 color histopathological images across five classes, featuring both cancerous and benign tissues from the lung and colon.
- **PathMNIST** [61]: PathMNIST is a multi-class classification dataset derived from colorectal cancer histology slides. It is comprised of 107,180 image patches, categorized into 9 distinct tissue types.
- **Amyloid CAA/Diffuse** [53]: It includes 100495 annotations on 20099 candidate amyloid beta neuropathologies, which is a neuropathology image dataset for subtypes of amyloid pathology in brain tissue, used for subtype classification tasks.

Non-Clinical Biological Imaging:
- **Pollen** [1]: The Pollen13K dataset is a large-scale collection of over 13,000 microscopic pollen grain images from aerobiological samples, used for biological particle classification. It was chosen to test model generalization on complex, non-clinical imagery.

G. FLOPs Computation

This appendix consolidates the FLOPs counting procedure for BIOMEDICA [36] used to report scoring cost.

All totals below measure a single complete *scoring* pass over $\mathcal{D}_{\text{train}}$ and use the batch primitives in Tab. 5: $C_{\text{fwd}}(B)$, $C_{\text{bwd}}(B)$, $C_{\text{fb}}(B) = C_{\text{fwd}}(B) + C_{\text{bwd}}(B)$, and $C_{\text{jvp}}(B)$. The random projection (CountSketch) cost per application is $C_{\text{rp}} \approx 2P$ (see Sec. G). We denote $n_{\text{train}} = \lceil N_{\text{train}}/B_{\text{train}} \rceil$ and $n_{\text{eval}} = \lceil N_{\text{eval}}/B_{\text{eval}} \rceil$. TracIn uses E epochs for accumulation. TRAK and CHIPS use I conjugate-gradient (CG) iterations.

TracIn. TracIn combines a single evaluation-direction construction with per-batch JVPs and trajectory accumulation:

$$C_{\text{TracIn}} = C_{\text{eval-dir}} + C_{\text{JVP-train}} + C_{\text{accum}}, \tag{24}$$

$$C_{\text{eval-dir}} = n_{\text{eval}} \left(C_{\text{fb}}(B_{\text{eval}}) + C_{\text{rp}} \right), \tag{25}$$

$$C_{\text{JVP-train}} = n_{\text{train}}\, C_{\text{jvp}}(B_{\text{train}}), \tag{26}$$

$$C_{\text{accum}} = E\, n_{\text{train}}\, C_{\text{fb}}(B_{\text{train}}). \tag{27}$$

TRAK. TRAK builds a second-order score with CG in the same end-point geometry. We write the shared backbone block once and reuse it:

$$\begin{aligned} \text{Base} = {} & n_{\text{eval}} \left(C_{\text{fb}}(B_{\text{eval}}) + C_{\text{rp}} \right) \; + \; C_{\text{proto-eval}} \\ & + \; n_{\text{train}} \left(C_{\text{fb}}(B_{\text{train}}) + C_{\text{rp}} \right) \\ & + \; I\, n_{\text{train}} \left(C_{\text{jvp}}(B_{\text{train}}) + C_{\text{fb}}(B_{\text{train}}) + C_{\text{rp}} \right) \\ & + \; n_{\text{train}} \left(C_{\text{jvp}}(B_{\text{train}}) + C_{\text{fwd}}(B_{\text{train}}) \right). \end{aligned} \tag{28}$$

The TRAK total is simply

$$C_{\text{TRAK}} = \text{Base}. \tag{29}$$

CHIPS. CHIPS shares Base with TRAK and adds three lightweight terms for (i) negative-pair curvature in the sketched space, (ii) hardest-negative margin bookkeeping, and (iii) two prototype similarities per batch for relevance. To avoid overlong lines we factor these contributions:

$$\Delta_{\text{neg}} \approx c_{\text{neg}}\, k \tag{30}$$

$$\Delta_{\text{margin}} \approx 2\, B_{\text{train}}^2 \tag{31}$$

$$\Delta_{\text{rel}} \approx 4\, B_{\text{train}}\, d \tag{32}$$

which are vector-level ops per CG iteration, hardest-negative search in a $B_{\text{train}} \times B_{\text{train}}$ block, and two prototype similarities per batch.

The CHIPS total is then

$$\begin{aligned} C_{\text{CHIPS}} = {} & \text{Base} \\ & + \; I\, \Delta_{\text{neg}} \\ & + \; n_{\text{train}}\, \Delta_{\text{margin}} \\ & + \; n_{\text{train}}\, \Delta_{\text{rel}}. \end{aligned} \tag{33}$$

C_{rp} is incurred when forming the sketched evaluation direction and wherever sketched vectors are refreshed. Δ_{neg} accounts for a handful of axpy-like operations per CG iteration in the k-dimensional sketched space. Δ_{margin} is a max-reduction over off-diagonal logits already materialized for symmetric InfoNCE. Δ_{rel} computes two dot products per sample with cached prototypes $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$.

G.1. Numerical Totals

We instantiate Eqs. (24)-(33) with the fixed values above. We use $E = 10$ epochs for TracIn and $I = 5$ CG iterations for TRAK and CHIPS.

$$\begin{aligned} \textbf{TracIn}\ (E{=}10): &\quad 5.258869 \times 10^{16}. \\ \textbf{TRAK}\ (I{=}5): &\quad 5.094585 \times 10^{16}. \\ \textbf{CHIPS}\ (I{=}5): &\quad 5.094747 \times 10^{16}. \end{aligned}$$

G.2. Assumptions and Further notes

- Symmetric CLIP computes logits in both directions; $C_{\text{fwd}}(B)$ already includes the two $B \times B$ matmuls in Eq. (5).
- The empirical bound $C_{\text{jvp}}(B) \approx 2\, C_{\text{fwd}}(B)$ holds for our implementation and batch shapes.
- The evaluation mean-gradient uses a single split with $B_{\text{eval}} = 3400$. $C_{\text{proto-eval}}$ counts only projections and L2 norms.
- CountSketch is the default random projection; for other sketches replace C_{rp} with the appropriate cost model (sparse RP $2sP$, SRHT $2m \log_2 m$, dense Gaussian $2kP$).
- FLOPs are operation counts independent of arithmetic precision and exclude file I/O and host preprocessing.
- *Scaling summary.* With fixed B and d, the dominant terms scale as TracIn $= \Theta(E\, n_{\text{train}}\, C_{\text{fb}})$, TRAK $= \Theta(I\, n_{\text{train}}\, C_{\text{fb}})$, CHIPS $= \Theta(I\, n_{\text{train}}\, C_{\text{fb}}) + O(Ik + n_{\text{train}} B^2 + n_{\text{train}} Bd)$. At the reported k and B, the $O(Ik)$ and $O(n_{\text{train}} Bd)$ extras are negligible and Δ_{margin} is amortized by already-computed logits.

H. Full Results

H.1. Main Experiment

The full results of the main experiment are detailed in Tab. 6 and Tab. 7 for the medical domain, and in Tab. 8 through Tab. 11 for the general domain.

Quantity	Meaning	General formula	B_{train}	B_{eval}
C_{lin}	two linear projections to d	$2\,B\,(d_v{+}d_t)\,d$	$4.294967296 \times 10^{10}$	4.456448×10^{9}
C_{norm}	two L2 normalizations	$\approx 6\,B\,d$	1.00663296×10^{8}	1.04448×10^{7}
C_{mm}	logits matmul for one direction	$2\,B^2\,d$	$1.099511627776 \times 10^{12}$	1.183744×10^{10}
C_{fwd}	forward, both directions	$C_{\text{lin}} + C_{\text{norm}} + 2C_{\text{mm}}$	$2.242073591808 \times 10^{12}$	$2.81417728 \times 10^{10}$
C_{bwd}	backward to end-point heads and τ	$\approx 2\,(C_{\text{lin}} + 2C_{\text{mm}})$	$4.483945857024 \times 10^{12}$	5.6262656×10^{10}
C_{fb}	forward + backward	$C_{\text{fwd}} + C_{\text{bwd}}$	$6.726019448832 \times 10^{12}$	$8.44044288 \times 10^{10}$
C_{jvp}	JVP cost upper bound†	$\approx 2\,C_{\text{fwd}}$	$4.484147183616 \times 10^{12}$	$5.62835456 \times 10^{10}$
$C_{\text{proto-eval}}$	eval prototypes only	$C_{\text{lin}}(B_{\text{eval}}) + C_{\text{norm}}(B_{\text{eval}})$	4.4668928×10^{9}	

Table 5. Batch-level FLOPs primitives and instantiated costs. Symmetric CLIP loss computes logits in both directions. $^{\dagger}C_{\text{jvp}}$ uses the empirical bound $C_{\text{jvp}} \approx 2\,C_{\text{fwd}}$ for this architecture and batching.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
PubMedCLIP	63.25	51.46	8.16	25.27	94.12	26.88	8.45	19.01
BioMedCLIP	2.26	47.53	20.19	72.91	1.65	87.92	10.14	3.43
BMCLIP	69.95	73.96	38.00	17.28	1.80	74.72	9.44	8.77
Vanilla	2.58	55.77	20.15	10.56	52.61	63.31	17.95	57.67
Full Dataset	60.89	71.62	44.83	9.81	88.30	77.27	23.33	62.11
r=10%								
Random	67.29	68.43	38.91	10.18	19.85	68.07	20.61	8.44
Concept-Balance	61.18	66.89	39.71	10.87	11.07	82.57	21.81	15.62
Concept-Filter	54.91	68.60	40.00	11.06	6.85	82.21	23.41	9.93
CLIPScore	2.49	60.68	40.67	10.56	72.81	55.49	20.99	27.80
Dot	71.67	57.20	39.89	23.51	72.64	18.34	16.87	5.50
TracIn	65.40	56.75	40.03	26.34	90.73	12.84	16.69	3.69
TRAK	43.23	59.23	39.57	12.88	77.16	14.74	25.93	3.60
CHIPS (ours)	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
r=20%								
Random	65.86	57.13	40.55	11.13	10.30	62.52	21.92	15.43
Concept-Balance	55.50	69.32	40.13	10.37	3.49	82.83	24.00	15.01
Concept-Filter	54.92	69.78	41.04	15.21	3.25	68.44	25.46	6.93
CLIPScore	4.34	59.92	47.44	11.25	48.38	16.31	18.82	11.17
Dot	72.11	60.07	31.49	33.63	62.01	45.43	13.30	11.07
TracIn	72.77	57.24	36.61	19.17	75.28	16.98	19.70	3.00
TRAK	71.76	58.73	41.71	28.16	42.51	15.15	16.75	3.31
CHIPS (ours)	72.50	68.67	48.07	30.36	53.58	23.50	22.49	13.97
r=30%								
Random	70.95	70.79	41.39	10.94	29.82	64.34	18.42	14.17
Concept-Balance	55.50	62.83	41.01	10.43	5.07	66.69	24.61	10.15
Concept-Filter	54.92	52.14	41.15	21.06	5.73	76.31	25.96	3.97
CLIPScore	16.89	53.70	44.16	11.94	29.20	23.78	13.94	5.59
Dot	70.35	57.56	32.45	20.74	44.69	25.38	15.40	6.39
TracIn	72.13	58.48	29.81	22.19	50.72	31.45	14.24	7.28
TRAK	69.39	59.76	42.11	25.14	35.00	16.40	16.92	1.64
CHIPS (ours)	73.42	73.66	49.88	36.34	67.90	25.15	25.02	14.39
r=50%								
Random	63.57	72.99	39.65	10.56	12.41	69.59	18.82	17.01

Table 6. Full medical-domain results (Part 1/2) of the main experiment.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
PubMedCLIP	12.83	9.63	26.50	23.07	22.63	23.81	21.73	18.00	8.45
BioMedCLIP	9.78	1.20	25.00	20.28	22.55	21.92	8.48	16.00	4.66
BMCLIP	9.57	61.65	25.70	27.15	20.37	21.35	43.52	19.75	5.02
Vanilla	9.35	10.82	26.40	10.42	9.91	11.46	22.21	18.75	5.24
Full Dataset	8.64	11.72	21.50	17.51	14.89	14.20	40.91	17.50	12.40
r=10%									
Random	6.57	11.27	23.00	13.80	10.35	10.49	25.75	17.75	4.68
Concept-Balance	7.63	11.67	21.60	15.07	11.14	11.09	27.49	16.75	5.03
Concept-Filter	7.44	11.72	22.30	14.79	10.43	11.46	31.98	17.00	5.14
CLIPScore	6.86	11.47	27.20	8.51	8.99	11.60	25.79	16.75	5.25
Dot	6.17	12.17	20.60	9.48	10.36	11.19	25.42	33.75	4.16
TracIn	6.53	12.02	18.00	13.98	12.85	15.28	22.98	40.50	4.96
TRAK	25.84	12.17	20.90	10.74	10.43	11.22	28.68	26.75	4.46
CHIPS (ours)	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
r=20%									
Random	17.84	11.62	24.00	14.75	10.13	10.93	29.86	17.50	6.60
Concept-Balance	9.46	11.97	25.30	17.65	13.19	12.77	33.75	17.50	5.42
Concept-Filter	8.41	11.97	22.40	17.28	13.01	12.85	34.54	16.75	6.42
CLIPScore	4.94	11.27	27.40	10.10	11.25	11.69	22.34	20.25	4.45
Dot	7.15	11.77	22.00	12.35	12.50	12.87	25.14	23.50	5.38
TracIn	11.96	11.82	17.40	13.39	13.94	19.12	28.91	37.25	5.07
TRAK	6.94	11.02	23.50	12.42	13.49	14.43	25.10	34.50	5.41
CHIPS (ours)	8.96	11.87	21.40	13.50	13.80	14.87	23.20	38.50	5.92
r=30%									
Random	7.93	12.07	22.10	15.32	10.76	11.33	30.60	17.25	6.42
Concept-Balance	8.40	12.07	23.60	16.53	12.38	11.92	33.11	16.75	6.48
Concept-Filter	7.52	12.02	25.20	17.38	13.72	13.57	36.16	17.00	6.81
CLIPScore	4.19	11.32	26.90	9.64	11.16	10.69	28.69	22.00	4.49
Dot	10.49	11.62	21.40	13.21	12.61	12.56	30.03	23.50	7.69
TracIn	17.75	11.97	18.00	14.57	14.57	16.74	33.68	28.75	5.69
TRAK	6.49	11.42	19.20	13.80	14.61	15.91	22.98	28.25	7.75
CHIPS (ours)	10.65	11.37	21.30	13.13	14.28	15.33	24.43	30.25	7.69
r=50%									
Random	9.86	11.82	22.60	16.86	12.66	12.90	35.31	17.25	10.06

Table 7. Full medical-domain results (Part 2/2) of the main experiment. Abbreviations: Derma = DermaMNIST, Oct = OctMNIST, OrganA = OrganAMNIST, OrganC = OrganCMNIST, OrganS = OrganSMNIST, Path = PathMNIST, Retina = RetinaMNIST, Tissue = TissueMNIST.

H.2. Generalization Experiment

The full results of the main experiment are detailed in Tab. 12 and Tab. 13 for the medical domain, and in Tab. 14 through Tab. 17 for the general domain.

H.3. Ablation Experiment

The full results of the main experiment are detailed in Tab. 18 and Tab. 19 for the medical domain, and in Tab. 20 through Tab. 23 for the general domain.

H.4. Analysis Experiment

- **Hyperparameter Analysis:** The analysis of evaluation set size, mixing hyperparameter α, and balance hyperparameter β is presented in Tab. 24 through Tab. 29.
- **End-point Subspace:** The analysis of the end-point subspace ϑ is shown in Tab. 30 and Tab. 31.
- **JL Random Projection:** The analysis of JL random projection is provided in Tab. 32 and Tab. 33.

I. Additional Results

We provide additional experiments on MedTrinity [56] in Tab. 34 and Tab. 35.

Model	Cars	Country211	FER	Aircraft	Food101	GTSRB	Imagenet -A	Imagenet -O	Imagenet 1k	Imagenet v2
PubMedCLIP	8.51	3.09	20.30	4.32	15.43	15.37	4.97	14.55	14.95	13.00
BioMedCLIP	0.37	0.37	13.33	1.02	0.99	2.49	0.72	0.75	0.07	0.07
BMCLIP	92.13	27.18	37.57	31.38	90.40	56.22	54.93	33.85	71.97	64.80
Vanilla	74.29	22.41	42.57	28.50	87.24	43.80	46.97	39.55	70.78	62.59
Full Dataset	66.11	20.41	45.00	20.37	81.00	36.94	42.24	37.60	66.28	58.90
r=10%										
Random	71.43	21.76	43.73	27.06	84.99	41.94	45.60	39.20	69.18	61.31
Concept-Balance	71.28	21.81	44.90	27.48	84.82	42.76	45.73	39.35	69.12	61.05
Concept-Filter	70.92	21.61	43.86	26.91	84.59	42.36	44.91	39.30	68.75	60.85
CLIPScore	74.93	22.42	42.09	28.77	86.90	44.25	47.16	39.65	70.60	62.63
Dot	65.32	17.74	38.12	19.98	76.25	29.78	37.57	31.10	60.91	53.14
TracIn	64.82	17.30	35.94	19.71	74.90	27.40	36.41	29.95	59.22	51.16
TRAK	65.10	17.69	37.46	20.58	76.27	28.89	37.61	30.15	60.54	52.58
CHIPS (ours)	65.61	17.80	38.72	20.43	75.94	29.65	38.13	30.25	61.07	53.01
r=20%										
Random	70.20	21.60	45.29	25.71	84.47	41.35	45.23	39.30	68.67	60.53
Concept-Balance	69.12	21.29	45.89	25.77	83.75	39.30	44.53	38.40	68.26	60.30
Concept-Filter	69.42	21.20	44.61	25.89	83.36	38.61	44.03	38.40	67.87	59.94
CLIPScore	71.88	20.70	41.32	26.79	85.41	41.09	45.59	35.95	68.13	60.13
Dot	70.84	18.83	39.24	22.86	81.28	29.65	40.45	31.75	63.04	55.01
TracIn	69.54	18.55	39.29	22.23	80.37	29.30	39.48	30.55	62.51	54.46
TRAK	69.44	18.79	37.98	23.34	80.05	31.58	41.32	31.85	63.56	55.44
CHIPS (ours)	69.32	19.54	39.40	24.36	80.05	31.06	40.33	32.10	62.97	55.01
r=30%										
Random	69.36	21.36	46.22	25.23	84.10	39.74	44.61	38.40	68.21	60.20
Concept-Balance	69.03	21.50	46.20	25.92	83.55	38.97	44.56	38.60	68.33	60.48
Concept-Filter	68.88	20.82	43.81	25.08	82.21	36.93	42.68	37.75	67.13	59.44
CLIPScore	72.11	20.48	42.05	26.46	84.78	38.42	44.92	35.40	67.68	59.30
Dot	69.36	18.16	37.94	22.47	79.58	27.87	39.15	30.40	61.90	53.53
TracIn	68.96	18.20	39.84	22.89	79.51	29.26	39.65	30.15	61.87	53.63
TRAK	68.82	18.36	38.63	23.91	79.20	30.37	39.47	30.75	62.41	54.55
CHIPS (ours)	68.30	18.39	37.89	23.46	79.27	30.36	38.96	31.40	62.33	54.45
r=50%										
Random	68.47	21.03	46.15	23.28	82.96	38.74	43.67	37.90	67.61	59.90

Table 8. Full general-domain results (Part 1/4) of the main experiment.

Model	MNIST	Rendered SST2	STL10	Sun397	Sun397 Official	VOC	Caltech101	CIFAR10	CIFAR100
PubMedCLIP	25.55	54.37	59.24	18.27	16.86	34.95	22.73	38.40	10.54
BioMedCLIP	9.58	50.08	9.70	0.36	0.24	3.55	0.51	10.31	1.19
BMCLIP	86.27	63.92	97.86	69.79	69.28	80.87	83.55	96.98	81.53
Vanilla	47.83	60.57	97.28	66.83	68.20	72.18	83.80	90.24	66.69
Full Dataset	27.61	55.24	97.29	65.17	65.64	66.12	83.73	91.43	66.79
r=10%									
Random	45.07	55.85	97.05	66.82	67.89	69.95	85.55	91.94	67.61
Concept-Balance	37.41	54.26	97.12	67.01	68.09	69.44	85.85	92.34	68.40
Concept-Filter	41.63	56.29	96.96	66.88	67.80	69.87	85.55	92.25	67.94
CLIPScore	47.82	60.57	97.28	66.82	68.45	72.70	84.40	90.83	67.90
Dot	38.78	55.63	95.54	58.91	61.10	68.78	81.79	85.03	58.04
TracIn	39.09	52.94	95.24	57.33	59.55	68.56	82.17	82.60	54.60
TRAK	44.99	56.84	95.80	58.33	60.34	66.35	83.22	82.17	54.39
CHIPS (ours)	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
r=20%									
Random	34.08	54.48	97.17	66.71	67.57	67.92	85.39	92.22	67.98
Concept-Balance	38.87	55.68	97.14	66.54	67.39	67.91	85.09	91.96	67.62
Concept-Filter	36.40	57.11	96.88	66.43	66.99	69.49	85.19	91.68	67.26
CLIPScore	47.78	58.54	97.04	62.08	64.06	71.39	83.63	88.03	61.89
Dot	38.50	57.77	95.36	56.67	59.19	66.22	81.61	84.61	57.69
TracIn	37.95	55.19	94.67	56.22	58.38	67.15	81.71	84.93	56.99
TRAK	43.87	51.62	95.14	55.84	58.63	68.10	82.04	85.28	58.14
CHIPS (ours)	42.91	50.63	95.42	55.94	58.51	65.89	81.38	85.72	59.00
r=30%									
Random	27.07	54.26	97.15	66.51	67.18	67.28	85.08	92.34	68.13
Concept-Balance	31.72	54.31	97.00	66.68	67.37	68.08	85.09	92.18	68.32
Concept-Filter	34.15	57.33	96.80	65.84	66.55	68.56	84.34	91.07	65.50
CLIPScore	46.31	57.06	97.05	62.11	63.81	69.97	83.32	88.35	62.38
Dot	37.90	54.64	95.31	55.82	57.89	66.75	80.82	83.64	55.78
TracIn	37.04	53.21	94.96	55.72	57.92	67.11	80.84	83.05	54.84
TRAK	32.68	50.08	95.30	55.68	57.78	64.52	80.81	84.56	57.45
CHIPS (ours)	34.25	49.92	95.36	55.69	57.91	64.36	80.79	84.61	57.64
r=50%									
Random	30.17	54.09	97.21	66.02	66.73	66.76	84.60	92.11	67.94

Table 9. Full general-domain results (Part 2/4) of the main experiment.

Model	CLEVR Closest	CLEVR Count	DMLAB	DTD	Eurosat	Flowers	KITTI
PubMedCLIP	22.28	16.23	18.85	10.59	19.70	14.54	31.36
BioMedCLIP	24.51	16.85	16.86	1.17	11.26	0.59	29.54
BMCLIP	15.79	33.98	13.80	55.64	63.02	76.11	26.30
Vanilla	22.47	29.11	16.11	56.22	55.96	73.57	24.47
Full Dataset	20.45	21.03	12.08	50.48	50.06	70.69	32.49
r=10%							
Random	22.57	29.39	12.28	53.56	53.31	72.68	31.08
Concept-Balance	22.59	31.12	12.02	52.93	52.39	72.94	28.83
Concept-Filter	22.41	31.32	11.99	52.66	48.26	72.68	29.54
CLIPScore	22.45	31.71	15.08	56.60	56.54	74.21	22.93
Dot	21.38	25.25	15.86	47.82	42.61	66.14	17.58
TracIn	21.39	22.96	14.83	45.80	40.17	66.08	17.02
TRAK	22.55	21.23	14.04	48.14	37.98	68.12	20.82
CHIPS (ours)	22.59	21.31	14.92	48.86	40.27	66.29	19.83
r=20%							
Random	22.16	26.30	11.98	52.66	51.19	73.05	30.66
Concept-Balance	21.53	29.22	11.80	52.07	52.26	72.48	29.68
Concept-Filter	21.51	30.64	11.92	51.91	50.33	71.74	31.50
CLIPScore	22.47	28.19	14.92	55.00	48.13	70.60	20.39
Dot	20.93	20.89	16.73	46.44	41.33	65.67	20.39
TracIn	18.71	21.03	15.45	43.88	40.69	64.55	16.60
TRAK	21.21	24.05	15.61	45.11	41.83	66.45	18.71
CHIPS (ours)	21.52	25.25	16.21	44.47	42.89	66.53	16.46
r=30%							
Random	21.86	26.93	11.85	52.34	52.33	72.56	32.07
Concept-Balance	21.23	27.33	11.83	51.44	52.69	72.56	30.38
Concept-Filter	21.63	27.87	11.84	51.65	48.91	71.59	31.50
CLIPScore	22.44	27.65	14.88	55.11	47.78	70.92	18.00
Dot	21.01	20.81	15.41	42.82	39.65	64.90	16.60
TracIn	18.77	20.52	15.44	43.14	37.85	65.05	17.02
TRAK	21.40	20.57	15.64	44.26	41.35	64.68	15.61
CHIPS (ours)	21.60	20.78	15.76	44.20	42.19	65.30	14.77
r=50%							
Random	21.29	24.69	11.74	51.38	50.54	71.85	32.77

Table 10. Full general-domain results (Part 3/4) of the main experiment.

Model	Pets	RESISC45	Smallnorb Azimuth	Smallnorb Elevation	SVHN
PubMedCLIP	23.36	14.22	5.51	10.97	7.43
BioMedCLIP	2.89	2.21	5.64	10.95	9.83
BMCLIP	91.39	63.38	6.29	10.54	50.43
Vanilla	90.54	66.08	5.42	10.92	24.34
Full Dataset	88.77	63.05	6.77	12.12	19.39
r=10%					
Random	90.52	64.08	5.93	10.88	18.41
Concept-Balance	90.11	63.76	5.82	10.47	19.17
Concept-Filter	89.83	62.35	5.82	10.53	19.56
CLIPScore	90.02	66.38	5.67	11.51	25.80
Dot	88.09	60.21	5.84	10.44	26.41
TracIn	88.23	60.14	5.47	11.37	24.96
TRAK	87.57	60.54	5.26	11.60	26.31
CHIPS (ours)	86.97	58.44	5.55	11.82	25.23
r=20%					
Random	90.16	63.87	6.11	10.97	18.19
Concept-Balance	89.45	63.49	6.07	11.05	18.78
Concept-Filter	89.32	61.79	5.90	11.12	20.29
CLIPScore	89.02	64.83	5.34	10.80	27.57
Dot	86.48	60.22	6.51	10.48	25.77
TracIn	88.25	59.75	5.37	10.96	25.78
TRAK	86.59	60.29	5.67	11.80	24.32
CHIPS (ours)	87.24	59.68	5.49	11.02	24.63
r=30%					
Random	89.92	63.79	6.27	10.74	17.19
Concept-Balance	89.29	64.56	6.41	10.78	18.74
Concept-Filter	89.23	62.60	5.97	11.74	20.28
CLIPScore	88.39	64.11	5.13	10.89	26.40
Dot	86.59	60.65	5.69	11.64	23.29
TracIn	87.76	59.51	5.54	10.67	23.28
TRAK	86.84	59.44	5.60	11.30	23.07
CHIPS (ours)	87.08	60.00	5.84	11.18	22.45
r=50%					
Random	89.64	63.71	6.19	11.04	17.09

Table 11. Full general-domain results (Part 4/4) of the main experiment.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
B32-400M								
Random	53.80	65.05	40.88	67.57	93.21	23.61	14.56	1.24
TracIn	57.38	55.25	37.12	50.47	96.33	12.32	21.66	2.16
CHIPS (ours)	59.94	55.22	36.27	69.83	98.35	13.97	16.22	1.56
B32-CC								
Random	73.20	62.22	21.31	32.24	1.54	87.94	35.52	3.53
TracIn	66.56	56.44	19.17	23.26	5.01	85.16	30.20	11.05
CHIPS (ours)	51.66	56.75	19.44	39.85	6.46	87.88	28.73	34.37
B16-400M								
Random	67.29	68.43	38.91	10.18	19.85	68.07	20.61	8.44
TracIn	65.40	56.75	40.03	26.34	90.73	12.84	16.69	3.69
CHIPS (ours)	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
B16-CC								
Random	72.85	59.01	20.61	16.66	49.21	87.68	16.90	2.33
TracIn	74.49	52.07	20.33	19.04	91.57	37.16	12.36	2.00
CHIPS (ours)	76.91	54.99	32.75	22.88	34.65	82.59	14.88	2.90
L14-400M								
Random	73.65	55.76	43.68	19.61	11.95	87.16	9.97	5.72
TracIn	65.62	60.57	40.61	16.03	18.51	51.40	19.47	5.10
CHIPS (ours)	65.34	61.27	39.65	19.48	44.86	73.60	11.14	5.84
L14-CC								
Random	47.71	54.50	24.16	68.13	11.29	86.15	18.06	41.76
TracIn	51.44	56.73	26.40	60.72	42.54	83.83	18.80	49.22
CHIPS (ours)	54.22	56.09	21.73	63.54	33.75	84.75	20.50	46.61
H14-CC								
Random	72.41	64.77	50.27	12.19	92.56	86.96	24.44	2.43
TracIn	73.09	65.87	50.75	12.70	76.15	83.24	14.24	2.92
CHIPS (ours)	78.64	65.30	55.12	11.50	82.87	86.85	20.94	8.00

Table 12. Full medical-domain results (Part 1/2) of the generalization experiment.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
B32-400M									
Random	6.05	13.97	25.00	10.77	6.86	5.30	34.50	9.75	5.31
TracIn	3.80	17.11	24.80	9.93	10.32	9.19	16.56	36.00	12.49
CHIPS (ours)	2.64	12.52	24.20	13.31	11.33	10.59	27.87	19.25	11.63
B32-CC									
Random	4.42	9.63	25.00	16.22	8.98	6.24	31.91	14.25	9.18
TracIn	6.71	16.06	24.30	18.37	11.37	11.03	27.80	19.00	10.78
CHIPS (ours)	9.14	13.82	24.70	23.67	14.28	13.67	29.82	13.00	10.36
B16-400M									
Random	6.57	11.27	23.00	13.80	10.35	10.49	25.75	17.75	4.68
TracIn	6.53	12.02	18.00	13.98	12.85	15.28	22.98	40.50	4.96
CHIPS (ours)	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
B16-CC									
Random	9.92	11.82	25.50	17.61	11.04	12.47	31.50	6.00	5.38
TracIn	9.12	12.82	23.30	14.47	14.05	15.37	41.14	8.50	5.94
CHIPS (ours)	10.10	12.37	24.20	13.61	13.56	14.42	43.15	9.75	6.22
L14-400M									
Random	9.81	19.10	23.20	26.17	16.99	17.37	36.39	38.50	23.02
TracIn	10.51	21.05	23.30	23.10	16.13	17.55	36.92	31.00	11.04
CHIPS (ours)	8.74	25.74	22.20	23.01	15.06	17.07	35.04	38.50	17.37
L14-CC									
Random	13.06	21.40	26.20	25.68	19.81	19.23	30.74	20.25	7.31
TracIn	15.82	27.83	23.60	16.71	15.68	11.29	21.88	20.25	6.11
CHIPS (ours)	16.39	28.99	23.30	19.39	16.38	15.26	25.42	16.75	7.74
H14-CC									
Random	8.97	62.19	24.80	25.52	16.69	14.35	38.29	8.00	4.83
TracIn	13.05	64.69	23.80	31.44	20.53	18.83	37.51	22.25	6.53
CHIPS (ours)	8.89	54.16	23.20	27.00	18.78	15.67	34.21	19.25	7.24

Table 13. Full medical-domain results (Part 2/2) of the generalization experiment.

Model	Cars	Country211	FER	Aircraft	Food101	GTSRB	Imagenet -A	Imagenet -O	Imagenet 1k	Imagenet v2
B32-400M										
Random	66.66	16.73	30.54	25.44	78.84	38.45	27.85	45.80	63.49	55.78
TracIn	66.00	15.50	33.74	22.41	77.07	27.69	27.07	40.75	61.35	52.69
CHIPS (ours)	65.97	15.85	31.51	24.42	77.74	30.26	28.40	42.10	62.46	53.85
B32-CC										
Random	74.99	17.20	43.52	23.34	80.65	39.71	28.37	46.35	65.90	57.95
TracIn	73.96	15.58	35.82	23.46	78.49	37.19	29.13	40.85	63.71	55.60
CHIPS (ours)	73.37	16.05	45.75	24.15	79.41	35.33	29.73	41.55	64.16	56.40
B16-400M										
Random	71.43	21.76	43.73	27.06	84.99	41.94	45.60	39.20	69.18	61.31
TracIn	70.63	19.09	37.11	23.91	81.94	31.37	40.48	31.20	63.62	55.34
CHIPS (ours)	68.40	21.47	40.69	25.32	82.60	31.02	42.36	32.50	66.88	56.22
B16-CC										
Random	81.12	22.35	48.75	30.42	86.97	51.00	48.59	40.85	71.06	64.12
TracIn	75.25	17.76	34.20	22.59	81.33	36.29	41.97	31.20	62.65	55.68
CHIPS (ours)	75.96	18.10	35.33	23.49	82.34	40.96	43.88	32.20	64.15	57.34
L14-400M										
Random	82.93	30.34	41.36	38.97	89.35	48.79	65.73	29.35	75.23	68.72
TracIn	78.54	26.27	36.57	35.25	87.81	39.90	61.49	24.30	69.84	63.35
CHIPS (ours)	80.26	26.73	39.45	34.71	88.02	41.71	62.24	25.20	70.88	64.61
L14-CC										
Random	88.02	33.86	54.99	42.18	93.31	56.33	71.47	29.05	78.28	71.32
TracIn	83.62	29.36	46.68	41.19	90.38	48.76	66.41	24.65	73.51	66.43
CHIPS (ours)	84.65	29.78	47.94	40.98	91.25	46.98	68.07	25.15	74.15	67.55
H14-CC										
Random	89.64	37.46	52.49	51.43	93.87	58.34	74.92	29.45	79.75	73.69
TracIn	88.73	35.35	53.19	45.15	92.81	57.78	73.16	24.05	76.19	69.91
CHIPS (ours)	88.30	35.64	55.46	47.58	93.09	57.86	74.25	24.85	76.63	70.39

Table 14. Full general-domain results (Part 1/4) of the generalization experiment.

Model	MNIST	Rendered SST2	STL10	Sun397	Sun397 Official	VOC	Caltech101	CIFAR10	CIFAR100
B32-400M									
Random	39.99	54.81	96.11	63.75	64.66	66.37	84.21	91.19	67.92
TracIn	35.04	53.76	94.42	59.56	60.19	68.32	82.86	88.24	63.02
CHIPS (ours)	38.52	55.13	95.08	59.43	60.70	66.67	82.70	88.68	63.70
B32-CC									
Random	43.11	53.87	96.42	66.50	66.95	76.30	85.80	95.17	77.92
TracIn	38.90	52.83	95.40	63.29	64.03	75.01	86.06	94.27	76.24
CHIPS (ours)	36.16	51.62	95.71	63.95	64.63	75.73	84.62	94.01	75.47
B16-400M									
Random	45.07	55.85	97.05	66.82	67.89	69.95	85.55	91.94	67.61
TracIn	39.09	52.94	95.24	57.33	59.55	68.56	82.17	82.60	54.60
CHIPS (ours)	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
B16-CC									
Random	59.17	54.75	98.35	68.43	69.29	78.17	84.54	96.09	79.97
TracIn	54.70	53.27	97.12	58.73	60.19	69.29	82.28	92.97	72.27
CHIPS (ours)	55.66	54.42	97.61	61.01	62.31	70.19	82.68	93.50	73.19
L14-400M									
Random	52.89	64.52	99.20	71.34	72.54	74.47	85.98	95.97	76.67
TracIn	58.40	58.26	97.34	62.08	63.28	62.04	82.28	89.52	70.96
CHIPS (ours)	59.23	62.93	97.74	61.34	62.34	62.95	82.83	91.16	70.80
L14-CC									
Random	60.93	68.26	99.25	72.03	73.77	80.58	88.15	97.60	84.84
TracIn	57.61	56.73	98.61	66.52	67.98	74.63	84.45	95.20	79.35
CHIPS (ours)	64.90	64.80	98.84	67.41	68.64	75.45	84.42	95.46	78.86
H14-CC									
Random	70.45	70.68	99.49	73.01	75.34	73.45	87.54	98.10	86.61
TracIn	33.99	66.89	99.19	68.30	69.37	67.22	83.43	97.15	84.05
CHIPS (ours)	20.47	63.87	99.20	68.98	70.47	69.35	85.18	97.09	84.10

Table 15. Full general-domain results (Part 2/4) of the generalization experiment.

Model	CLEVR Closest	CLEVR Count	DMLAB	DTD	Eurosat	Flowers	KITTI
B32-400M							
Random	21.01	23.33	19.06	50.32	50.06	70.06	32.63
TracIn	21.85	22.33	16.89	47.55	47.96	66.60	24.61
CHIPS (ours)	22.40	23.06	15.15	48.30	45.43	69.15	32.07
B32-CC							
Random	23.78	21.09	12.20	55.64	49.52	69.07	15.33
TracIn	22.59	19.37	12.34	53.03	45.76	63.77	15.61
CHIPS (ours)	22.55	20.12	13.17	54.52	43.00	66.11	19.69
B16-400M							
Random	22.57	29.39	12.28	53.56	53.31	72.68	31.08
TracIn	21.39	22.96	14.83	45.80	40.17	66.08	17.02
CHIPS (ours)	22.59	21.31	14.92	48.86	40.27	66.29	19.83
B16-CC							
Random	22.55	28.51	21.39	61.65	52.61	74.94	27.00
TracIn	22.69	25.08	19.25	51.60	53.44	62.06	22.22
CHIPS (ours)	22.72	26.79	19.38	52.87	55.43	65.07	22.08
L14-400M							
Random	21.07	34.80	20.12	59.84	62.15	77.90	28.41
TracIn	22.93	27.41	19.26	51.97	50.56	73.78	30.38
CHIPS (ours)	23.67	27.62	17.88	53.72	50.19	75.05	29.54
L14-CC							
Random	22.54	28.61	16.64	66.38	66.76	81.53	24.33
TracIn	22.54	27.45	20.95	59.89	56.39	78.09	25.04
CHIPS (ours)	22.54	28.74	18.50	60.37	53.33	78.60	32.77
H14-CC							
Random	10.25	20.78	14.86	68.99	69.37	83.72	27.14
TracIn	21.85	23.17	13.62	63.24	62.37	81.18	22.93
CHIPS (ours)	20.85	21.61	13.02	66.17	61.69	81.43	25.32

Table 16. Full general-domain results (Part 3/4) of the generalization experiment.

Model	Pets	RESISC45	Smallnorb Azimuth	Smallnorb Elevation	SVHN
B32-400M					
Random	85.75	57.11	5.35	12.03	23.40
TracIn	84.79	55.87	5.74	11.45	27.67
CHIPS (ours)	86.67	56.70	5.24	12.04	25.37
B32-CC					
Random	88.74	59.37	5.70	12.20	18.77
TracIn	87.84	56.25	5.60	11.65	22.40
CHIPS (ours)	88.47	58.83	6.09	12.47	21.03
B16-400M					
Random	90.52	64.08	5.93	10.88	18.41
TracIn	88.23	60.14	5.47	11.37	24.96
CHIPS (ours)	86.97	58.44	5.55	11.82	25.23
B16-CC					
Random	91.09	66.79	4.95	10.77	37.11
TracIn	87.57	59.35	5.33	11.01	36.07
CHIPS (ours)	87.82	60.21	5.28	10.77	36.69
L14-400M					
Random	92.89	68.68	5.51	11.04	22.41
TracIn	90.30	63.06	5.93	11.94	27.65
CHIPS (ours)	91.41	63.32	5.85	12.17	27.55
L14-CC					
Random	94.06	75.13	4.85	11.13	46.73
TracIn	90.81	66.89	5.40	11.11	49.62
CHIPS (ours)	92.37	66.44	5.21	11.56	50.27
H14-CC					
Random	95.53	72.56	6.30	12.30	44.79
TracIn	94.19	68.65	5.82	12.40	50.89
CHIPS (ours)	94.49	68.44	6.44	12.21	50.99

Table 17. Full general-domain results (Part 4/4) of the generalization experiment.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
r=10%								
Alignment-only	59.51	57.64	39.41	12.76	88.69	13.17	23.62	3.71
Alignment-Margin	59.56	57.65	40.21	11.69	87.51	13.94	23.56	3.26
CHIPS (ours)	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
r=20%								
Alignment-only	72.41	68.04	46.98	30.54	55.27	21.32	20.56	12.24
Alignment-Margin	72.02	68.32	46.87	30.10	58.24	21.65	21.34	12.34
CHIPS (ours)	72.50	68.67	48.07	30.36	53.58	23.50	22.49	13.97
r=30%								
Alignment-only	71.30	72.76	43.79	32.14	60.26	22.71	21.88	11.77
Alignment-Margin	71.24	70.43	48.16	32.42	61.79	25.12	23.44	13.19
CHIPS (ours)	73.42	73.66	49.88	36.34	67.90	25.15	25.02	14.39

Table 18. Full medical-domain results (Part 1/2) of the ablation experiment.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
r=10%									
Alignment-only	21.49	12.07	20.20	11.65	10.08	10.81	27.28	27.00	4.86
Alignment-Margin	21.02	12.17	19.60	12.10	10.00	10.28	26.70	29.25	4.99
CHIPS (ours)	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
r=20%									
Alignment-only	9.57	11.22	20.80	14.10	13.86	15.24	24.57	32.00	5.99
Alignment-Margin	9.07	11.12	20.80	14.53	14.17	15.04	24.55	35.75	5.39
CHIPS (ours)	8.96	11.87	21.40	13.50	13.80	14.87	23.20	38.50	5.92
r=30%									
Alignment-only	8.28	11.22	21.80	13.62	12.99	14.87	21.56	30.75	7.46
Alignment-Margin	10.49	11.32	20.50	14.35	13.38	12.69	24.08	30.50	6.40
CHIPS (ours)	10.65	11.37	21.30	13.13	14.28	15.33	24.43	30.25	7.69

Table 19. Full medical-domain results (Part 2/2) of the ablation experiment.

Model	Cars	Country211	FER	Aircraft	Food101	GSTRB	Imagenet -A	Imagenet -O	Imagenet 1k	Imagenet v2
r=10%										
Alignment-only	69.99	19.50	40.85	25.59	82.55	31.56	42.47	32.95	64.84	56.63
Alignment-Margin	70.12	21.67	40.81	25.92	82.60	31.62	42.17	32.45	65.01	56.78
CHIPS (ours)	68.40	21.47	40.69	25.32	82.60	31.02	42.36	32.50	66.88	56.22
r=20%										
Alignment-only	68.97	18.80	39.19	24.60	80.06	31.50	40.11	31.60	62.95	55.05
Alignment-Margin	68.95	19.59	39.23	24.51	80.03	31.92	40.15	31.60	62.89	54.99
CHIPS (ours)	69.32	19.54	39.40	24.36	80.05	31.06	40.33	32.10	62.97	55.01
r=30%										
Alignment-only	68.86	18.41	38.35	24.78	79.45	30.78	39.45	31.60	62.63	54.52
Alignment-Margin	68.74	18.48	42.28	24.18	79.93	30.43	39.83	31.40	62.56	54.74
CHIPS (ours)	68.30	18.39	37.89	23.46	79.27	30.36	38.96	31.40	62.33	54.45

Table 20. Full general-domain results (Part 1/4) of the ablation experiment.

Model	MNIST	Rendered SST2	STL10	Sun397	Sun397 Official	VOC	Caltech101	CIFAR10	CIFAR100
r=10%									
Alignment-only	45.04	56.78	95.42	58.44	60.47	66.18	83.27	82.55	55.19
Alignment-Margin	44.47	57.44	95.34	58.34	60.40	66.66	82.93	82.57	55.34
CHIPS (ours)	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
r=20%									
Alignment-only	43.42	50.74	95.19	56.52	58.73	66.31	81.38	84.85	57.73
Alignment-Margin	42.65	50.36	94.99	56.47	58.76	65.93	81.10	85.76	59.11
CHIPS (ours)	42.91	50.63	95.42	55.94	58.51	65.89	81.38	85.72	59.00
r=30%									
Alignment-only	38.92	50.14	95.21	56.16	58.28	65.77	80.12	85.44	57.97
Alignment-Margin	35.97	50.03	95.17	55.95	58.43	62.85	80.30	84.61	56.85
CHIPS (ours)	34.25	49.92	95.36	55.69	57.91	64.36	80.79	84.61	57.64

Table 21. Full general-domain results (Part 2/4) of the ablation experiment.

Model	CLEVR Closest	CLEVR Count	DMLAB	DTD	Eurosat	Flowers	KITTI
r=10%							
Alignment-only	22.62	21.30	13.90	48.67	38.98	68.11	23.49
Alignment-Margin	22.58	21.41	14.14	49.36	39.11	68.01	23.21
CHIPS (ours)	22.59	21.31	14.92	48.86	40.27	66.29	19.83
r=20%							
Alignment-only	21.27	24.39	15.90	44.79	41.89	66.22	17.86
Alignment-Margin	21.00	25.33	15.90	44.26	42.15	66.30	17.44
CHIPS (ours)	21.52	25.25	16.21	44.47	42.89	66.53	16.46
r=30%							
Alignment-only	21.24	21.56	15.40	45.21	42.76	65.03	15.61
Alignment-Margin	22.24	20.59	16.86	44.15	42.06	65.67	17.72
CHIPS (ours)	21.60	20.78	15.76	44.20	42.19	65.30	14.77

Table 22. Full general-domain results (Part 3/4) of the ablation experiment.

Model	Pets	RESISC45	Smallnorb Azimuth	Smallnorb Elevation	SVHN
r=10%					
Alignment-only	87.46	60.33	5.51	11.34	26.19
Alignment-Margin	87.35	60.21	5.47	11.32	26.04
CHIPS (ours)	86.97	58.44	5.55	11.82	25.23
r=20%					
Alignment-only	87.35	59.48	5.62	11.06	24.34
Alignment-Margin	87.35	59.49	5.32	10.74	23.96
CHIPS (ours)	87.24	59.68	5.49	11.02	24.63
r=30%					
Alignment-only	86.70	59.60	5.53	10.79	23.94
Alignment-Margin	87.00	59.68	5.82	10.99	23.79
CHIPS (ours)	87.08	60.00	5.84	11.18	22.45

Table 23. Full general-domain results (Part 4/4) of the ablation experiment.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
Evaluation Set Size								
50	67.33	57.66	39.68	18.10	87.22	12.69	21.51	2.79
100	57.57	58.45	39.39	14.90	82.92	13.02	24.76	2.50
150	54.62	56.38	40.61	11.38	85.64	13.07	24.06	4.61
200	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
250	68.07	66.49	40.55	22.20	87.99	12.63	23.91	2.14
α								
0.2	52.04	57.57	40.53	13.70	79.23	13.61	24.55	4.28
0.4	51.45	57.18	41.20	13.51	74.95	14.03	24.90	3.28
0.6	56.65	55.74	40.21	14.39	89.81	13.85	25.17	2.90
0.8	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
1.0	51.01	55.88	40.67	14.77	83.71	13.51	23.79	3.29
β								
0	71.38	59.27	44.43	27.03	43.02	25.05	18.36	3.13
0.25	54.72	56.48	39.89	11.75	90.43	12.74	22.42	3.08
0.50	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
0.75	52.53	57.27	38.93	13.83	86.47	13.79	23.21	4.25
1.0	71.53	59.11	41.89	26.40	54.35	23.21	15.76	2.71

Table 24. Full medical-domain results (Part 1/2) of the analysis on evaluation set size, α, and β.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
Evaluation Set Size									
50	16.07	12.07	21.50	8.66	8.70	10.04	26.56	35.25	5.20
100	21.20	12.17	20.70	9.23	8.87	9.49	26.24	32.75	5.87
150	28.48	12.32	21.00	11.42	10.32	10.59	30.21	35.00	4.92
200	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
250	16.49	12.07	19.50	12.48	10.13	9.88	24.38	30.25	5.02
α									
0.2	30.17	12.12	20.30	10.93	9.54	10.30	28.23	30.00	4.92
0.4	29.49	12.17	20.70	10.98	10.08	10.82	28.08	29.25	4.82
0.6	28.46	12.32	20.70	11.76	10.66	11.74	29.33	33.00	4.95
0.8	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
1.0	28.75	12.07	20.30	11.49	10.30	11.39	29.04	30.50	5.04
β									
0	7.56	11.32	19.90	10.15	11.19	12.91	24.43	37.00	5.80
0.25	19.36	12.07	20.60	11.63	10.02	10.11	26.38	35.50	5.18
0.50	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
0.75	22.38	11.82	19.90	11.30	10.70	11.05	29.28	32.75	5.45
1.0	6.39	10.97	19.50	10.09	11.51	12.71	24.32	38.75	5.68

Table 25. Full medical-domain results (Part 2/2) of the analysis on evaluation set size, α, and β.

Model	Cars	Country211	FER	Aircraft	Food101	GSTRB	Imagenet -A	Imagenet -O	Imagenet 1k	Imagenet v2
Evaluation Set Size										
50	70.15	19.38	40.58	25.65	82.56	31.46	42.68	33.05	65.04	57.01
100	69.79	19.40	41.70	25.26	82.52	33.02	42.25	32.10	65.15	57.00
150	69.48	19.48	41.07	24.75	82.70	32.54	42.09	33.40	64.97	56.96
200	68.40	21.47	40.69	25.32	82.60	31.02	42.36	32.50	66.88	56.22
250	67.38	19.41	40.35	25.08	82.91	30.02	42.45	32.25	64.91	56.80
α										
0.2	69.59	19.26	41.71	24.99	82.4	32.38	42.72	32.35	64.97	56.81
0.4	69.82	19.34	40.97	25.26	82.65	32.07	42.71	32.40	65.12	57.15
0.6	68.40	21.47	40.69	25.32	82.60	31.02	42.36	32.50	66.88	56.22
0.8	69.88	19.40	40.92	25.77	82.56	31.95	42.85	31.70	65.08	57.13
1.0	70.07	19.50	41.31	25.56	82.63	31.50	43.00	32.15	65.12	57.08
β										
0	71.00	19.63	42.77	25.77	82.05	32.64	42.29	31.70	64.59	56.71
0.25	69.98	19.42	40.14	25.26	82.51	31.57	41.99	32.30	64.99	56.68
0.50	68.40	21.47	40.69	25.32	82.60	31.02	42.36	32.50	66.88	56.22
0.75	70.54	19.26	41.39	25.26	82.66	33.66	42.52	32.10	65.11	57.02
1.0	70.39	19.41	41.84	26.76	81.91	32.07	41.99	32.25	64.63	56.26

Table 26. Full general-domain results (Part 1/4) of the analysis on evaluation set size, α, and β.

Model	MNIST	Rendered SST2	STL10	Sun397	Sun397 Official	VOC	Caltech101	CIFAR10	CIFAR100
Evaluation Set Size									
50	47.03	57.77	95.29	58.61	60.61	66.67	82.74	82.78	54.91
100	46.74	56.84	95.40	58.71	60.62	65.63	83.47	82.48	54.73
150	43.58	57.00	95.31	59.10	60.79	64.98	83.57	82.77	55.25
200	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
250	42.26	58.26	95.09	57.58	60.68	63.62	83.25	80.11	55.04
α									
0.2	45.55	57.44	95.31	58.91	60.49	66.91	83.06	82.61	55.31
0.4	45.44	57.83	95.40	58.78	60.65	67.05	83.11	83.08	55.68
0.6	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
0.8	45.82	57.33	95.39	58.92	60.55	66.19	83.27	82.96	55.15
1.0	47.11	57.66	95.51	58.73	60.53	66.57	83.27	82.95	55.21
β									
0	42.85	52.33	95.51	57.74	60.47	66.73	82.65	85.81	58.97
0.25	44.89	58.26	94.31	58.89	60.89	66.29	82.97	82.37	54.13
0.50	38.82	57.84	94.91	58.00	57.88	68.66	82.48	80.76	55.33
0.75	42.92	58.37	95.34	58.86	60.79	66.25	83.58	84.02	56.88
1.0	40.30	51.89	95.14	57.72	60.44	66.75	82.51	85.63	59.00

Table 27. Full general-domain results (Part 2/4) of the analysis on evaluation set size, α, and β.

Model	CLEVR Closest	CLEVR Count	DMLAB	DTD	Eurosat	Flowers	KITTI
Evaluation Set Size							
50	22.65	21.22	14.08	48.24	39.11	68.11	23.21
100	22.65	21.45	14.08	48.83	38.69	68.04	22.22
150	22.60	22.14	14.63	48.35	39.52	67.47	23.07
200	22.59	21.31	14.92	48.86	40.27	66.29	19.83
250	22.44	21.19	13.76	48.35	37.52	67.05	25.32
α							
0.2	22.59	21.78	14.44	47.93	38.48	68.21	22.22
0.4	22.59	21.87	14.42	48.51	38.28	68.14	22.78
0.6	22.59	21.31	14.92	48.86	40.27	66.29	19.83
0.8	22.61	21.41	14.03	48.40	38.52	67.99	21.94
1.0	22.63	21.91	14.18	48.62	38.28	67.99	21.10
β							
0	22.10	26.25	15.35	46.91	42.26	68.35	20.53
0.25	22.66	21.09	13.89	48.14	39.11	68.48	23.07
0.50	22.59	21.31	14.92	48.86	40.27	66.29	19.83
0.75	22.55	21.87	13.98	47.93	37.81	67.75	23.21
1.0	22.06	26.15	15.76	47.23	42.78	68.24	20.53

Table 28. Full general-domain results (Part 3/4) of the analysis on evaluation set size, α, and β.

Model	Pets	RESISC45	Smallnorb Azimuth	Smallnorb Elevation	SVHN
Evaluation Set Size					
50	87.63	60.05	5.44	11.27	25.64
100	88.06	60.27	5.51	11.51	26.43
150	88.01	60.24	5.56	11.77	26.67
200	86.97	58.44	5.55	11.82	25.23
250	87.49	60.38	5.42	11.32	26.01
α					
0.2	87.76	60.56	5.42	11.41	25.56
0.4	87.46	60.46	5.79	11.49	25.97
0.6	86.97	58.44	5.55	11.82	25.23
0.8	87.74	60.62	5.28	11.49	26.41
1.0	87.90	60.43	5.33	11.28	26.49
β					
0	87.93	62.27	5.77	10.63	25.60
0.25	87.74	60.14	5.57	11.54	25.85
0.50	86.97	58.44	5.55	11.82	25.23
0.75	87.44	59.79	5.72	11.55	26.83
1.0	87.27	61.63	5.73	11.06	25.86

Table 29. Full general-domain results (Part 4/4) of the analysis on evaluation set size, α, and β.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
All								
Dot	71.67	57.20	39.89	23.51	72.64	18.34	16.87	5.50
TracIn	65.40	56.75	40.03	26.34	90.73	12.84	16.69	3.69
TRAK	43.23	59.23	39.57	12.88	77.16	14.74	25.93	3.60
CHIPS (ours)	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
Logit-only								
Dot	62.45	59.27	31.28	25.96	13.48	63.18	20.90	9.60
TracIn	58.24	61.00	33.89	16.09	8.30	71.48	19.38	10.15
TRAK	64.29	62.24	35.12	26.02	14.73	59.51	17.36	11.74
CHIPS (ours)	59.94	61.30	35.36	20.93	13.72	70.29	17.89	21.50
Visual-only								
Dot	61.96	56.15	37.71	23.00	13.33	47.26	20.05	6.74
TracIn	59.11	53.51	34.99	18.16	28.44	31.75	18.15	13.78
TRAK	71.23	59.48	44.83	32.75	23.59	24.27	15.02	15.37
CHIPS (ours)	64.01	57.61	41.25	42.61	33.29	53.67	20.93	5.48
Text-only								
Dot	71.29	61.81	40.77	24.07	40.91	12.56	17.10	20.57
TracIn	61.86	60.90	39.23	25.71	77.58	13.00	18.65	11.51
TRAK	51.61	57.08	39.55	21.36	34.21	46.31	20.61	11.22
CHIPS (ours)	69.76	60.02	46.83	35.80	25.76	55.13	18.65	12.80

Table 30. Full medical-domain results (Part 1/2) of the analysis on end-point subspace.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
All									
Dot	6.17	12.17	20.60	9.48	10.36	11.19	25.42	33.75	4.16
TracIn	6.53	12.02	18.00	13.98	12.85	15.28	22.98	40.50	4.96
TRAK	25.84	12.17	20.90	10.74	10.43	11.22	28.68	26.75	4.46
CHIPS (ours)	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
Logit-only									
Dot	9.63	11.72	24.10	9.22	9.10	7.31	13.89	21.50	6.13
TracIn	10.82	11.77	22.40	10.54	8.82	7.50	15.08	22.25	5.21
TRAK	6.84	11.92	22.80	9.20	7.80	7.34	15.10	23.25	5.53
CHIPS (ours)	9.49	12.02	22.50	9.42	8.68	7.68	16.18	19.00	5.16
Visual-only									
Dot	3.66	11.62	21.30	11.04	9.30	10.72	24.21	22.75	4.72
TracIn	6.45	10.97	19.80	11.80	12.04	13.90	24.25	26.00	4.95
TRAK	6.20	11.77	18.90	11.51	9.90	10.42	17.19	40.00	5.56
CHIPS (ours)	19.32	12.02	19.60	10.91	10.69	12.22	23.69	26.50	5.31
Text-only									
Dot	11.54	11.97	22.50	11.72	11.81	12.30	23.79	33.00	4.55
TracIn	10.52	12.07	19.80	10.27	12.34	13.11	23.97	40.25	4.69
TRAK	18.09	12.27	22.30	12.52	11.82	12.82	25.97	27.25	6.31
CHIPS (ours)	19.51	11.42	19.70	10.09	9.31	9.67	17.66	39.25	5.09

Table 31. Full medical-domain results (Part 2/2) of the analysis on end-point subspace.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
CountSketch								
2k	65.13	59.60	48.05	21.43	35.20	32.39	16.69	4.16
4k	68.35	63.66	40.51	25.96	88.06	12.38	21.43	6.24
8k	63.37	57.97	39.76	17.03	73.36	14.87	22.45	2.14
16k	62.10	55.73	40.00	21.50	79.72	23.67	22.27	3.38
Sparse								
2k	66.86	61.11	40.91	25.27	18.14	68.09	17.45	5.81
4k	68.68	60.15	44.19	23.76	23.67	39.46	10.38	5.01
8k	65.29	65.76	43.33	27.03	41.30	12.49	14.53	4.12
16k	68.51	62.10	43.89	35.89	43.41	56.96	24.26	8.14
SRHT								
2k	32.69	60.39	47.09	33.19	68.64	39.88	18.91	6.90
4k	69.35	58.69	39.04	14.39	88.25	15.72	18.65	8.84
8k	41.47	56.59	39.31	27.66	88.67	20.62	14.21	1.82
16k	65.89	58.36	40.35	41.04	32.65	44.25	12.72	3.77

Table 32. Full medical-domain results (Part 1/2) of the analysis on JL random projection.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
CountSketch									
2k	10.10	11.32	17.80	14.33	15.17	16.28	21.18	28.00	5.63
4k	16.63	11.87	18.20	12.03	9.64	10.71	23.90	30.25	5.29
8k	21.02	12.77	17.40	14.20	14.00	13.61	27.45	30.50	5.88
16k	20.71	12.37	21.40	14.04	14.62	13.73	25.96	36.00	8.09
Sparse									
2k	11.39	11.12	23.90	14.60	13.47	13.90	26.13	35.75	6.81
4k	10.47	11.72	20.60	10.70	11.31	11.35	24.35	30.00	6.68
8k	7.93	12.17	15.10	11.33	12.20	15.33	23.16	24.50	5.33
16k	7.06	11.42	20.70	12.30	12.07	12.35	28.83	31.75	5.60
SRHT									
2k	13.51	11.57	22.50	10.75	10.47	10.39	23.23	27.00	5.27
4k	9.45	12.02	17.10	9.60	10.72	11.03	31.63	36.00	6.47
8k	22.18	11.87	17.40	13.21	13.21	13.23	28.70	27.00	5.91
16k	7.81	11.22	18.30	12.30	12.48	11.63	24.75	26.25	6.55

Table 33. Full medical-domain results (Part 2/2) of the analysis on JL random projection.

Model	Diabetic	PCAM	LC25000	Pollen	Amyloid CAA	Amyloid Diffuse	BloodMNIST	ChestMNIST
B32-400M								
Random 10%	63.65	50.20	32.51	72.91	97.26	12.82	14.88	6.86
Random 50%	34.53	52.70	25.68	72.91	97.76	12.50	9.97	28.41
B32-CC								
Random 10%	3.45	63.08	8.80	11.00	1.84	87.94	29.49	2.75
Random 50%	3.11	56.25	23.89	51.85	2.86	87.94	21.60	17.87
B16-400M								
Random 10%	10.74	56.77	39.28	7.23	88.27	59.82	14.44	9.08
Random 50%	17.08	66.73	32.91	8.30	96.46	14.03	7.28	1.24
B16-CC								
Random 10%	46.68	50.49	6.40	21.06	60.28	87.95	6.58	1.56
Random 50%	22.16	50.83	17.95	61.28	62.36	87.95	8.21	1.07

Table 34. Medical-domain results (Part 1/2) of MedTrinity [56] dataset.

Model	ChestXray14	Derma	Oct	OrganA	OrganC	OrganS	Path	Retina	Tissue
B32-400M									
Random 10%	6.05	13.97	25.00	10.77	6.86	5.30	34.50	9.75	5.31
Random 50%	11.29	11.32	22.60	10.68	6.69	6.30	35.72	8.50	5.02
B32-CC									
Random 10%	4.42	9.63	25.00	16.22	8.98	6.24	31.91	14.25	9.18
Random 50%	7.56	11.07	25.00	21.59	14.56	14.72	46.03	16.00	8.01
B16-400M									
Random 10%	4.78	10.62	12.40	13.98	8.57	7.73	27.73	12.75	9.46
Random 50%	2.86	10.62	16.20	13.17	11.93	10.32	30.28	7.50	9.70
B16-CC									
Random 10%	9.89	10.62	25.90	13.04	12.71	13.33	35.65	5.00	6.15
Random 50%	6.45	18.50	25.40	19.92	14.79	15.54	42.45	5.75	4.92

Table 35. Medical-domain results (Part 2/2) of MedTrinity [56] dataset.