BALANCING REPRESENTATIONS BY IDENTIFYING AND GENERATING UNDER-REPRESENTED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Self-supervised learning inherits the coverage of its source data. When an unlabeled corpus is long-tailed, the representation becomes dense around frequent modes and thin where coverage is weak. Objectives for long-tailed SSL reweight the images already in the corpus and add no new ones. We introduce BRIDGE (**B**alancing **R**epresentations by **Id**entifying and **Ge**nerating Under-represented Data), a label-free feedback loop around the source set. The current encoder scores every source image by its mean kNN distance, and a budgeted acquisition rule selects images from a band of that score distribution where support is thin and the generator still produces faithful variants. A frozen image-to-image diffusion model then adds local variants of the selected anchors, and pretraining continues on the enlarged source set. We formulate acquisition as recoverability-weighted allocation of local support and derive when a finite budget should favor a sparse band that excludes the extreme tail and when reallocation after each repair is optimal under a fixed objective. Across five long-tailed, web-scraped and synthetic source regimes, BRIDGE has the highest seven-task linear-probe and kNN average of the compared methods, and paired experiments extend the gain to ViT-S under contrastive and self-distillation objectives, to low-label fine-tuning and to semantic segmentation. The linear-probe gain is smaller under masked reconstruction. Generated repair matches an oracle that restores the withheld real images, and under a frozen encoder the inserted images shrink the neighborhoods of the selected anchors about twice as much as those of matched unselected images. BRIDGE needs no labels and leaves the pretraining objective unchanged, so it can be added to pretraining on uncurated web data, where the encoder decides which images to generate.

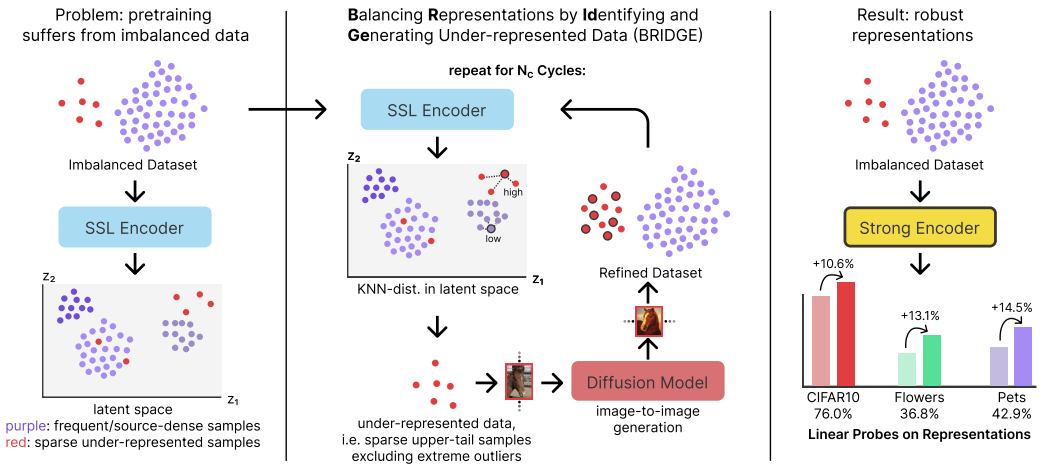

Figure 1: Overview of BRIDGE. Left, an imbalanced source set embedded by the SSL encoder, with sparse under-represented samples in red. Right, the loop. Distance to the k nearest neighbors in the current latent space selects sparse samples from the upper tail of the score distribution while the extreme tail is excluded, image-to-image generation adds variants of them, and the refined dataset is the source set of the next cycle.

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.

1 INTRODUCTION

Self-supervised learning produces transferable visual representations from unlabeled images (Chen et al., 2020; He et al., 2020; Caron et al., 2021; He et al., 2022; Oquab et al., 2024; Siméoni et al., 2025), and every objective learns from what its source corpus contains. When that corpus is long-tailed, the representation concentrates around frequent modes, covers rare regions weakly and transfers worse (Tian et al., 2021; Assran et al., 2023; Kukleva et al., 2023; Jiang et al., 2021). Prior work responds on the model side, with a temperature schedule (Kukleva et al., 2023), a pruned self-competitor (Jiang et al., 2021) or imported external data (Bai et al., 2023), leaving the source set fixed.

We ask whether the source set can be repaired while learning proceeds. BRIDGE, **B**alancing **R**epresentations by **Id**entifying and **Ge**nerating Under-represented Data, embeds the source set with the current encoder after each cycle and scores every image by its mean distance to its k nearest neighbors. It selects a sparse but semantically coherent band of that distribution, generates local variants of the selected images with a frozen image-to-image diffusion model and continues pretraining on the enlarged set (Figure 1), while no labels, class frequencies or downstream signals enter the loop.

We formalize acquisition as budgeted allocation over regions of representation space with diminishing returns to local support and repair fidelity that falls in the extreme tail. Under this model the budget belongs in a sparse band that excludes the extreme tail, and greedy reallocation after each repair is optimal under a fixed objective. In the experiments this band beats uniform acquisition by 6.6 to 6.8 points and top-tail acquisition by 7.8 to 8.0 points.

Across five long-tailed, web-scraped and synthetic sources, BRIDGE raises the seven-task linear-probe average of SimCLR on the same source by 2.4 to 10.6 points and has the best average of all compared methods under linear-probe and kNN evaluation. With a power-law class imbalance imposed on ImageNet-100, repair adds 11.1 points, while a class-balanced source of the same size adds 0.1. On ViT-S the linear-probe average rises by 2.0 to 2.8 points under SimCLR, MoCo v3 and DINO, and generated repair reaches the accuracy of an oracle that adds real images withheld from the source, 41.5 against 40.7.

Our contributions are as follows.

- A theory of label-free data repair as recoverability-weighted allocation of local support (Section 3), which shows that under diminishing returns and fidelity that decays with sparsity the budget belongs in a sparse band that excludes the extreme tail, and that greedy reallocation is optimal under a fixed separable objective.
- BRIDGE, a label-free feedback loop that instantiates this policy (Section 4) with mean kNN distance as the support estimate, a trimmed score band that holds sparse images the generator can still vary faithfully, farthest-point sampling against redundancy and a frozen image-to-image generator as the repair operator.
- An evaluation on five source regimes and seven transfer tasks under linear-probe and kNN protocols (Section 6), with paired ViT-S experiments across four SSL objectives, low-label fine-tuning, where accuracy rises by 2.0 and 3.2 points, and PASCAL VOC segmentation, where it rises by 2.4 mIoU.
- An account of where the effect comes from (Sections 6.1 and 6.2). The band beats uniform and top-tail acquisition as Section 3 predicts, and a fixed-encoder test shows the generated images shrinking the neighborhoods of their own anchors twice as much as those of matched unselected images.

2 RELATED WORK

Self-supervised learning under skewed sources. Contrastive, self-distillation and masked-reconstruction pretraining (Chen et al., 2020; He et al., 2020; Caron et al., 2021; Oquab et al., 2024; Siméoni et al., 2025; He et al., 2022) is sensitive to source imbalance, which has been addressed with clustered experts, the hidden uniform cluster prior and reweighting by estimated class frequency (Tian et al., 2021; Assran et al., 2023; Lin et al., 2023). The closest methods change

the objective on a fixed corpus, SDCLR with a pruned self-competitor (Jiang et al., 2021) and Temperature Schedules by emphasizing head and tail at different times (Kukleva et al., 2023), or import external images (Bai et al., 2023).

Rarity, curation and budgeted acquisition. Mean deep nearest-neighbor distance has been used as a nonparametric out-of-distribution score (Sun et al., 2022), and we use it as a local-support estimate on the training set itself. Curation by prototype distance or concept-cluster complexity removes examples (Sorscher et al., 2022; Abbas et al., 2024; Whang et al., 2023), SMOTE interpolates labeled minority examples (Chawla et al., 2002), and DOPING oversamples the low-density edge of a learned latent distribution for anomaly detection with no image generator and no recomputed target (Lim et al., 2018), which our top-tail selection rule reproduces under our generator and budget. Active learning spends a finite budget under a learned representation, its core-set form treating acquisition as k-center coverage under the greedy farthest-first traversal we use inside the mode window (Sener and Savarese, 2018; Gonzalez, 1985). ProbCover instead maximizes covered probability mass at a fixed radius (Yehuda et al., 2022), and Hacohen et al. (2022) find dense regions the better target at small budgets.

Generative augmentation and synthetic sources. SDEdit adds noise to an image and denoises it under a diffusion prior (Ho et al., 2020; Rombach et al., 2022; Meng et al., 2022), and that operator is our repair step. It has been applied to few-shot supervised recognition (Trabucco et al., 2024), class-conditional generation improves supervised ImageNet accuracy (Azizi et al., 2023), and synthetic-data utility depends on prompting, filtering, diversity and domain gap (He et al., 2023; Fan et al., 2024). TADA generates faithful variants of the examples a supervised classifier learns slowly (Nguyen et al., 2026), a criterion that requires labels. Synthetic imagery has also been a training source for self-supervised features since Ren and Lee (2018), from procedural noise and image programs (Baradad et al., 2021; 2022) to generator latents, synthetic ImageNet clones and text-to-image views (Jahanian et al., 2022; Sariyildiz et al., 2023; Tian et al., 2023), none allocating a budget over an observed skewed corpus. DiffAug jointly trains an encoder and a conditional diffusion generator so that generated positives improve contrastive learning (Zang et al., 2024), learning that generator from the small skewed source it is meant to repair, while BRIDGE holds a pretrained generator fixed and asks where it should intervene. Industrial data engines close a loop between a deployed model and its training data, diagnosing failures with vision-language and language models and then curating, auto-labeling and verifying (Liang et al., 2024; Worker and Pathak, 2022). They work from supervised task failures with annotation in the loop and acquire by semantic-cluster occupancy. A frozen generator also brings knowledge from its pretraining data (Li et al., 2023; Tang et al., 2023; Yu et al., 2025). Appendix A tabulates the closest methods.

3 Theoretical Framework

Targeted repair is modeled as budgeted allocation over coherent regions of representation space. A repair helps only where it adds support in a region that contributes to downstream error and the generator preserves the content of its anchor, so the policy weighs support deficit against repairability and redundancy. Let $r_k(z; Z)$ be the distance from z to its kth nearest neighbor in Z and A a finite set of valid variants generated from z. Appendix C gives complete proofs.

Proposition 1 (Monotone local support)**.** *For any finite A, $r_k(z; Z \cup A) \le r_k(z; Z)$. The inequality is strict when an added point enters the first k neighbors of z and displaces a neighbor at strictly larger distance.*

Adding points can only introduce smaller candidate distances, so the kth order statistic cannot increase.

Partition the representation into regions j with effective support n_j and model the regional contribution to downstream error by

$$\mathcal{B}(\mathbf{m}) = \sum_j p_j a_j (n_j + m_j + \kappa)^{-\gamma}, \tag{1}$$

where p_j is the transfer-relevant mass of region j, a_j its difficulty, m_j the number of valid repairs allocated to it, and $\kappa, \gamma > 0$. Write $\Delta_j(m_j) = p_j a_j[(n_j + m_j + \kappa)^{-\gamma} - (n_j + m_j + 1 + \kappa)^{-\gamma}]$ for the reduction from one more valid repair. None of p_j, a_j, κ, γ or the regions is instantiated or

estimated in the experiments, and Equation 1 is a qualitative model in the diminishing-returns form of empirical scaling laws (Sorscher et al., 2022).

Proposition 2 (Sparse-region marginal value)**.** *$\Delta_j(m_j)$ is positive and strictly decreasing in $n_j + m_j$. Between two regions of equal transfer mass and difficulty, the region with less effective support has the larger marginal value.*

The forward difference of the convex decreasing map $x \mapsto x^{-\gamma}$ is positive and decreasing, which gives the claim.

Sparsity alone is not sufficient, because generation fails on corrupt or incoherent anchors. Let d be a sparsity score, $b(d)$ the marginal value of a valid repair at that score, $\rho(d)$ the probability of a faithful variant and $\lambda \geq 0$ the cost of an unfaithful one, so the expected repair utility is

$$u(d) = \rho(d)\, b(d) - \big(1 - \rho(d)\big)\lambda. \tag{2}$$

Proposition 3 (Interior recoverable-support optimum)**.** *Suppose b and ρ are positive and continuously differentiable on $[d_{\min}, d_{\max}]$ with $b'(d) > 0$ and $\rho'(d) < 0$, and define $g(d) = b'(d)/(b(d) + \lambda) + \rho'(d)/\rho(d)$. If g is strictly decreasing with $g(d_{\min}) > 0$ and $g(d_{\max}) < 0$, then u has a unique maximizer $d^\star$ in the interior, characterized by $g(d^\star) = 0$.*

The sign of u' equals the sign of g, so the unique sign change of g is the unique maximum.

The monotonicity assumption on g holds for $b(d) = b_0 d^\beta$ with $\beta \in (0, 1]$ and $\rho(d) = \rho_0 e^{-\mu d}$ (Proposition 6), where g is strictly decreasing for every $\lambda \geq 0$ and $d^\star = \beta/\mu$ at $\lambda = 0$. The optimum moves outward when support value is closer to linear and inward when fidelity decays faster. BRIDGE uses a trimmed score band and estimates neither parameter. For the repeated case, consider a separable objective $U(\mathbf{m}) = \sum_j U_j(n_j + m_j)$ with each U_j increasing and discretely concave and a budget of R repairs, $\sum_j m_j = R$.

Proposition 4 (Adaptive allocation under diminishing returns)**.** *For integer allocations, repeatedly assigning one repair to the region with the largest current marginal gain maximizes $U(\mathbf{m})$. A fixed allocation may become suboptimal once a region's marginal gain falls below that of another region.*

Because every marginal sequence is nonincreasing, the greedy picks are the largest feasible marginals overall. In BRIDGE the encoder changes after every repair, and the regions and their marginals with it, so the benefit of re-estimation in the non-stationary case is an empirical claim.

Proposition 5 (Single-stage insensitivity)**.** *Let $\mathbf{m}$ and $\mathbf{m}'$ be any two allocations of the same budget of R repairs under Equation 1, computed from one representation and applied without re-estimation. Then*

$$\big|\mathcal{B}(\mathbf{m}) - \mathcal{B}(\mathbf{m}')\big| \;\leq\; R\,\gamma \max_j p_j a_j \,(n_j + \kappa)^{-\gamma-1}, \tag{3}$$

which tends to zero as $R/n_{\min} \to 0$ for the least-supported region $n_{\min}$.

Both allocations subtract at most R marginals from the same baseline, and the mean value theorem bounds every marginal.

A single allocation can therefore differ from any other only in proportion to the budget it adds, and the effect of the acquisition rule accumulates when the allocation is re-estimated over several cycles.

4 BRIDGE: Balancing Representations by Identifying and Generating Under-represented Data

4.1 Loop and scoring

Let $D^{(0)}$ be an unlabeled source set and f_θ an SSL encoder. BRIDGE alternates pretraining and repair over C cycles, with encoder $f_{\theta^{(c)}}$, source set $D^{(c)}$ and embedding set $Z^{(c)}$ at cycle c. The encoder is the only learned component, so the loop applies to any objective that yields embeddings. Every cycle continues from the previous encoder with a fresh optimizer schedule and retains all earlier images.

After training on $D^{(c)}$ we embed each image, $z_i^{(c)} = f_{\theta^{(c)}}(x_i)$, normalize to unit length, and compute

$$d_i^{(c)} = \frac{1}{k} \sum_{z \in \mathcal{N}_k(z_i^{(c)}; Z^{(c)} \setminus \{z_i^{(c)}\})} \|z_i^{(c)} - z\|_2^2, \tag{4}$$

where $\mathcal{N}_k(z; Z)$ is the set of k nearest neighbors of z in Z, so a large score marks an image in a sparse part of the current representation. We use $k = 100$.

4.2 Sparse and recoverable anchor selection

The extreme tail holds artifacts, degenerate crops and images the generator cannot vary faithfully, while uniform sampling spends most of its budget in regions that are already dense. In three steps (Figure 2) the selector restricts attention to the band $U^{(c)} = \{i \mid q_{0.75}^{(c)} \le d_i^{(c)} \le q_{0.99}^{(c)}\}$, locates the center $\bar{d}^{(c)}$ of the modal bin of a histogram over that band, forms the pool $P^{(c)} \subseteq U^{(c)}$ of size αB that minimizes $\sum_{i \in P^{(c)}} |d_i^{(c)} - \bar{d}^{(c)}|$, and selects the anchor set $S^{(c)}$ of size B from that pool by greedy farthest-point sampling in embedding space. In every run the modal bin of the right-skewed band sits at its lower edge, so the candidate pool is the $4B$ scores just above the 75th percentile and the histogram step changes nothing (Figure 2). We use $\alpha = 4$ and $B = 500$.

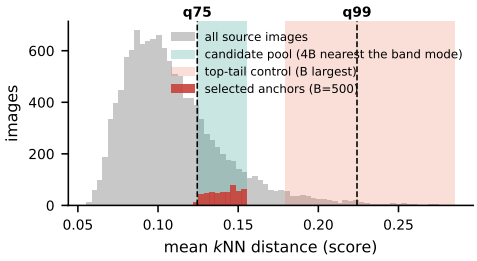

Figure 2: Score histogram of one ImageNet-100-LT run at cycle 0 with the 75th and 99th percentiles marked. The green pool holds the $4B$ scores nearest the modal bin of the band, which lies at its lower edge, the red region is the top-tail selection rule, and the 500 selected anchors are overlaid.

4.3 Local generative repair and continued pretraining

For each anchor we generate $G = 5$ variants with a frozen image-to-image diffusion model g_ϕ, conditioning on the anchor image under an empty text prompt so the repair signal comes from the anchor alone, and set $D^{(c+1)} = D^{(c)} \cup \{g_\phi(x_i, \epsilon_j) \mid i \in S^{(c)}, j \in [G]\}$. The noise ϵ_j indexes independent translations of one anchor, so variants keep coarse structure while changing texture, pose or color. We use Stable Diffusion 3 Medium with 20 denoising steps and strength 0.6 at the training crop resolution, and Algorithm 1 gives the loop.

5 Experimental Setup

Sources, targets and protocol. We pretrain on five unlabeled source regimes, the long-tailed ImageNet-100-LT (Liu et al., 2019), CIFAR-10-LT and CIFAR-100-LT (Krizhevsky, 2009), and 10,000-image subsets of the web corpora PASS (Asano et al., 2021) and DiffusionDB (Wang et al., 2023). Transfer is measured on seven benchmarks, CIFAR-10-LT, CIFAR-100-LT, Stanford Cars (Krause et al., 2013), FGVC Aircraft (Maji et al., 2013), Oxford Flowers (Nilsback and Zisserman, 2008), Oxford-IIIT Pets (Parkhi et al., 2012) and ImageNet-100-LT, by linear probe and by kNN classification on frozen features. Every number is a three-seed mean and standard deviation, paired comparisons share seeds and checkpoints, and ResNet-50 is the default backbone. The baselines train five cycles of 100 epochs on a source that does not grow, while BRIDGE trains the same epochs on a source that grows from about 10,000 to about 20,000 images, so it receives about 1.5 times the gradient updates in Tables 2 and 3. The cycle ablation and the paired ViT-S experiments match updates exactly. The seed standard deviation of a seven-task average is 0.6 to 1.5 points, so differences below about one point are inside seed variation. Appendix D gives implementation details.

Long-tailed construction. ImageNet-100 is class-balanced, so we construct its long-tailed version. A Pareto profile with exponent $\alpha = 2.5$ over the 100 class ranks is rescaled to per-class retention counts between 1,280 and 5 images and permuted over the classes. Each image is kept with the

Table 1: Imposed imbalance on ImageNet-100 over three seeds. Balancing the classes changes SimCLR by 0.1 points, while BRIDGE on the long-tailed source adds 11.1 linear-probe and 7.3 kNN points.

Protocol	Source	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Linear probe	Balanced	65.3 ± 1.2	37.0 ± 0.8	8.0 ± 0.5	15.1 ± 0.7	22.4 ± 1.6	31.2 ± 2.0	39.0 ± 1.0	31.1 ± 1.1
	Purposefully imbalanced	65.4 ± 4.2	38.6 ± 1.4	7.6 ± 1.1	15.3 ± 0.9	23.7 ± 1.0	28.4 ± 1.3	39.6 ± 0.3	31.2 ± 1.4
	BRIDGE on the imbalanced source	**76.0 ± 0.3**	**49.1 ± 0.2**	**15.2 ± 1.2**	**22.4 ± 0.4**	**36.8 ± 1.9**	**42.9 ± 1.0**	**53.8 ± 0.1**	**42.3 ± 0.7**
kNN	Balanced	63.6 ± 0.7	32.8 ± 0.8	4.4 ± 0.4	9.0 ± 0.4	22.3 ± 0.1	17.2 ± 0.1	34.0 ± 1.1	26.2 ± 0.5
	Purposefully imbalanced	63.9 ± 0.2	33.2 ± 0.3	4.5 ± 0.6	9.0 ± 0.3	22.5 ± 0.3	16.8 ± 0.2	34.3 ± 0.3	26.3 ± 0.3
	BRIDGE on the imbalanced source	**69.0 ± 0.5**	**37.8 ± 0.3**	**6.7 ± 0.3**	**11.7 ± 0.2**	**36.5 ± 0.4**	**25.3 ± 1.0**	**48.1 ± 0.6**	**33.6 ± 0.5**

Table 2: Linear-probe transfer from five sources to seven tasks, mean ± standard deviation over three seeds. BRIDGE has the best average on every source, 2.4 to 10.6 points above SimCLR. Bold marks the best value in a column.

Source	Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ImageNet-100-LT	SimCLR	73.9 ± 0.5	46.8 ± 0.4	13.7 ± 0.3	19.6 ± 0.2	35.9 ± 7.2	38.6 ± 1.5	50.4 ± 0.5	39.8 ± 1.5
	TS	72.5 ± 0.4	44.0 ± 0.5	12.5 ± 1.2	15.0 ± 0.1	35.7 ± 1.8	40.6 ± 2.2	**56.7 ± 1.2**	39.6 ± 1.1
	SDCLR	41.1 ± 1.7	18.0 ± 0.6	1.4 ± 0.3	2.3 ± 0.5	10.8 ± 1.7	10.4 ± 0.3	21.9 ± 0.6	15.1 ± 0.8
	BRIDGE (ours)	**75.8 ± 0.2**	**49.3 ± 0.2**	**16.4 ± 1.2**	**22.4 ± 0.4**	**36.8 ± 1.9**	**42.9 ± 1.0**	53.8 ± 0.1	**42.5 ± 0.7**
CIFAR-10-LT	SimCLR	67.8 ± 0.2	38.0 ± 1.0	6.5 ± 0.4	16.0 ± 0.9	**19.2 ± 1.6**	24.0 ± 1.3	32.0 ± 0.7	29.1 ± 0.9
	TS	68.5 ± 1.3	36.0 ± 1.7	4.3 ± 0.2	10.3 ± 0.1	13.1 ± 2.0	19.7 ± 2.3	31.5 ± 1.2	26.2 ± 1.3
	SDCLR	39.6 ± 1.1	15.6 ± 0.2	1.1 ± 0.1	2.4 ± 0.9	12.0 ± 1.6	9.1 ± 1.8	16.5 ± 1.6	13.8 ± 1.1
	BRIDGE (ours)	**78.1 ± 0.3**	**49.8 ± 0.4**	**9.2 ± 0.8**	**17.7 ± 1.0**	18.2 ± 1.1	**27.4 ± 2.7**	**39.0 ± 0.6**	**34.2 ± 1.0**
CIFAR-100-LT	SimCLR	59.0 ± 1.5	31.1 ± 1.6	4.1 ± 0.3	11.2 ± 1.0	12.6 ± 3.4	20.1 ± 1.0	28.7 ± 1.3	23.8 ± 1.4
	TS	56.3 ± 0.2	28.0 ± 0.1	2.8 ± 0.6	8.5 ± 0.2	15.5 ± 2.5	16.2 ± 3.1	26.2 ± 1.8	21.9 ± 1.2
	SDCLR	37.1 ± 0.8	15.7 ± 0.2	1.9 ± 0.1	2.4 ± 0.7	9.4 ± 1.0	9.9 ± 1.7	16.2 ± 0.3	13.2 ± 0.7
	BRIDGE (ours)	**74.2 ± 0.5**	**48.1 ± 0.4**	**8.3 ± 1.0**	**18.2 ± 0.7**	**24.4 ± 2.3**	**27.6 ± 0.7**	**39.8 ± 1.1**	**34.4 ± 1.0**
PASS-10k	SimCLR	72.3 ± 0.2	45.5 ± 0.5	11.0 ± 0.2	17.5 ± 1.4	33.3 ± 1.3	33.3 ± 0.9	44.7 ± 0.9	36.8 ± 0.8
	TS	71.5 ± 0.2	43.5 ± 0.4	10.2 ± 1.2	17.2 ± 0.2	33.1 ± 2.1	36.4 ± 1.8	**49.4 ± 0.4**	37.3 ± 0.9
	SDCLR	42.2 ± 1.1	17.7 ± 0.8	1.8 ± 0.3	2.2 ± 0.8	10.6 ± 1.3	8.2 ± 1.6	20.8 ± 0.4	14.8 ± 0.9
	BRIDGE (ours)	**73.9 ± 0.3**	**47.1 ± 0.4**	**13.6 ± 0.7**	**20.5 ± 0.7**	**37.7 ± 4.0**	**37.1 ± 0.7**	49.1 ± 0.5	**39.9 ± 1.0**
DiffusionDB-10k	SimCLR	74.4 ± 0.5	48.6 ± 0.2	10.9 ± 0.7	17.6 ± 0.8	31.2 ± 2.4	36.1 ± 1.3	45.3 ± 0.5	37.7 ± 0.9
	TS	72.3 ± 0.3	44.4 ± 0.3	10.7 ± 0.6	15.4 ± 0.6	33.1 ± 1.1	**39.9 ± 2.7**	48.8 ± 0.2	37.8 ± 0.8
	SDCLR	42.4 ± 0.2	18.4 ± 0.2	1.1 ± 0.2	2.4 ± 0.1	13.0 ± 3.7	8.2 ± 1.4	21.0 ± 0.6	15.2 ± 0.9
	BRIDGE (ours)	**75.9 ± 0.4**	**49.9 ± 0.5**	**12.8 ± 1.0**	**20.0 ± 0.4**	**35.7 ± 1.1**	37.7 ± 0.8	**49.1 ± 0.5**	**40.1 ± 0.7**

retention probability of its class. The balanced source of Table 1 uses the average of those probabilities for every class, so the two sources have the same expected size and a different class distribution. CIFAR-10-LT and CIFAR-100-LT use their standard configurations.

Compared methods and baselines. We compare with two groups of methods. TS and SDCLR change the loss on a fixed long-tailed source and are trained on each of the five sources, with hybrids in the appendix. TADA, AIDE and DiffAug add generated images as BRIDGE does, so they start from the same ImageNet-100-LT checkpoint per seed and follow the same five cycles, and only the choice or generation of images differs. All ablations use ImageNet-100-LT pretraining, the same downstream suite and three seeds under the five-cycle protocol. The generator ablation includes an oracle that restores 2,500 withheld real images per repair, the same count as the generated variants, placing true support at a deficit no label-free method can see.

6 Results

Imposed imbalance on ImageNet-100. ImageNet is class-balanced by construction, so the first question is what a class imbalance costs once it is imposed on purpose. Two sources are drawn from the same pool, one with the power-law profile of Section 5 and one subsampled uniformly at its average retention rate, so they differ only in how images are spread over classes. The two SimCLR runs differ by 0.1 points, 31.1 against 31.2 on the linear-probe seven-task average and 26.2 against 26.3 under kNN. BRIDGE on the imbalanced source reaches 42.3 and 33.6, which is 11.1 and 7.3 points above SimCLR on the same source and a similar margin above the balanced one. The remaining experiments use sources that are long-tailed without our intervention.

Transfer across five source regimes. BRIDGE has the best seven-task average in every source regime under both protocols (Tables 2 and 3), and against SimCLR and TS it is best in 31 of 35 linear-probe cells and 33 of 35 kNN cells. The gain over source-matched SimCLR is above ten linear-probe points under CIFAR-100-LT and above five under CIFAR-10-LT, the two smallest and

Table 3: kNN counterpart of Table 2. BRIDGE again has the best average on every source.

Source	Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ImageNet-100-LT	SimCLR	68.6 ± 0.4	37.3 ± 0.2	6.0 ± 0.7	11.2 ± 0.4	32.2 ± 0.7	23.6 ± 0.3	43.9 ± 0.9	31.8 ± 0.5
	TS	68.4 ± 0.3	36.5 ± 0.1	4.7 ± 0.2	8.3 ± 0.5	26.6 ± 1.0	23.6 ± 1.7	**52.2 ± 0.5**	31.5 ± 0.6
	SDCLR	32.5 ± 0.4	10.9 ± 0.2	1.0 ± 0.3	2.1 ± 0.1	10.7 ± 0.7	4.7 ± 0.2	11.2 ± 0.2	10.4 ± 0.3
	BRIDGE (ours)	**69.0 ± 0.5**	**37.8 ± 0.3**	**7.0 ± 0.2**	**11.7 ± 0.2**	**36.5 ± 0.4**	**25.3 ± 1.0**	48.1 ± 0.6	**33.6 ± 0.4**
CIFAR-10-LT	SimCLR	66.9 ± 0.4	33.4 ± 0.3	3.7 ± 0.4	9.8 ± 0.7	16.5 ± 0.8	13.9 ± 0.8	24.0 ± 0.8	24.0 ± 0.6
	TS	66.3 ± 0.6	27.8 ± 0.4	2.4 ± 0.6	6.4 ± 0.2	10.5 ± 0.7	11.1 ± 0.4	22.0 ± 0.9	20.9 ± 0.5
	SDCLR	30.1 ± 0.3	10.0 ± 0.3	1.7 ± 0.3	2.2 ± 0.2	10.1 ± 0.7	4.4 ± 0.4	9.8 ± 0.3	9.7 ± 0.4
	BRIDGE (ours)	**74.8 ± 0.4**	**41.0 ± 0.4**	**4.8 ± 0.5**	**10.0 ± 0.2**	**24.2 ± 0.6**	**16.2 ± 0.7**	**29.7 ± 0.3**	**28.7 ± 0.4**
CIFAR-100-LT	SimCLR	59.3 ± 0.8	30.1 ± 0.9	3.0 ± 0.3	7.4 ± 0.4	13.7 ± 0.2	10.8 ± 0.4	21.8 ± 1.1	20.9 ± 0.6
	TS	53.1 ± 0.6	24.6 ± 0.4	2.4 ± 0.4	5.5 ± 0.3	9.3 ± 0.4	8.3 ± 0.2	18.7 ± 1.0	17.4 ± 0.5
	SDCLR	25.7 ± 1.2	8.5 ± 0.6	0.9 ± 0.4	1.9 ± 0.2	7.3 ± 0.3	3.9 ± 0.3	8.3 ± 1.1	8.1 ± 0.6
	BRIDGE (ours)	**69.3 ± 0.5**	**39.5 ± 0.1**	**4.8 ± 0.6**	**8.9 ± 0.1**	**21.9 ± 0.1**	**15.5 ± 0.1**	**29.1 ± 0.7**	**27.0 ± 0.3**
PASS-10k	SimCLR	67.2 ± 0.2	**37.0 ± 0.3**	5.4 ± 0.3	10.3 ± 0.5	30.4 ± 0.0	16.7 ± 0.5	38.3 ± 0.3	29.3 ± 0.3
	TS	65.2 ± 0.7	34.6 ± 0.3	4.4 ± 0.2	7.9 ± 0.2	24.5 ± 0.1	14.4 ± 0.4	39.3 ± 0.4	27.2 ± 0.3
	SDCLR	32.4 ± 1.5	11.5 ± 0.8	1.3 ± 0.2	2.0 ± 0.1	8.6 ± 0.7	5.2 ± 0.2	11.1 ± 0.8	10.3 ± 0.6
	BRIDGE (ours)	**67.5 ± 0.3**	36.9 ± 0.3	**5.8 ± 0.5**	**10.7 ± 0.4**	**33.7 ± 0.2**	**17.8 ± 0.6**	**41.4 ± 0.8**	**30.6 ± 0.4**
DiffusionDB-10k	SimCLR	67.3 ± 0.1	36.2 ± 0.4	4.3 ± 0.2	8.8 ± 0.1	26.8 ± 0.4	17.3 ± 0.1	36.7 ± 0.8	28.2 ± 0.3
	TS	64.7 ± 0.4	33.6 ± 0.3	3.1 ± 0.5	7.3 ± 0.3	23.2 ± 0.4	17.0 ± 0.0	37.8 ± 0.6	26.7 ± 0.4
	SDCLR	32.6 ± 1.2	10.0 ± 0.3	1.3 ± 0.1	2.2 ± 0.2	6.1 ± 0.7	4.8 ± 0.4	10.6 ± 0.5	9.7 ± 0.5
	BRIDGE (ours)	**68.3 ± 0.2**	**37.4 ± 0.4**	**5.4 ± 0.3**	**9.8 ± 0.1**	**30.4 ± 0.2**	**18.6 ± 0.2**	**39.4 ± 1.0**	**29.9 ± 0.3**

Table 4: Paired runs with and without BRIDGE, three seeds. (A) Repair raises the linear-probe average of all four SSL objectives on ViT-S. (B) The gain persists under fine-tuning and segmentation. Bold marks the better run of each pair.

A. Frozen ViT-S transfer under four objectives

Objective	Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg. linear	Avg. kNN
SimCLR	SSL baseline	60.5 ± 0.5	35.0 ± 0.2	7.2 ± 0.2	10.9 ± 0.4	32.9 ± 2.7	25.0 ± 1.2	42.0 ± 1.4	30.5	30.4
	with BRIDGE	**60.9 ± 0.3**	**36.3 ± 0.1**	**8.2 ± 0.4**	**11.8 ± 0.4**	**36.6 ± 2.9**	**29.8 ± 3.0**	**43.9 ± 1.1**	**32.5**	**31.1**
MoCo v3	SSL baseline	60.0 ± 0.4	34.5 ± 0.0	7.6 ± 0.2	11.1 ± 1.3	30.3 ± 1.4	29.3 ± 3.1	44.8 ± 1.2	31.1	33.0
	with BRIDGE	**62.4 ± 0.5**	**36.3 ± 0.1**	**9.6 ± 0.3**	**13.4 ± 1.0**	**36.8 ± 6.1**	**31.2 ± 4.0**	**47.5 ± 0.9**	**33.9**	32.8
DINO	SSL baseline	58.9 ± 0.9	33.6 ± 1.8	8.1 ± 0.2	11.4 ± 1.0	**38.1 ± 5.2**	27.5 ± 2.3	44.6 ± 1.5	31.8	30.5
	with BRIDGE	**61.3 ± 0.1**	**35.2 ± 0.3**	**9.5 ± 0.9**	**13.3 ± 0.6**	37.2 ± 2.6	**34.6 ± 3.6**	**46.8 ± 0.8**	**34.0**	**32.0**
MAE	SSL baseline	37.7 ± 3.7	13.2 ± 1.9	1.5 ± 0.2	2.4 ± 1.0	**10.5 ± 5.9**	**10.1 ± 3.0**	19.9 ± 2.1	13.6	9.6
	with BRIDGE	**41.5 ± 2.2**	**15.0 ± 2.4**	**1.6 ± 0.4**	**3.2 ± 0.7**	8.5 ± 2.3	8.6 ± 2.2	**22.2 ± 2.4**	**14.4**	**11.6**

B. End-to-end downstream adaptation, SimCLR ResNet-50

Evaluation	Setting	SSL baseline	With BRIDGE	Paired difference
Full fine-tuning	IN100-LT, 1% labels	14.31 ± 0.33	**16.29 ± 1.11**	+1.97 ± 0.78
Full fine-tuning	IN100-LT, 10% labels	21.56 ± 0.75	**24.72 ± 1.75**	+3.17 ± 1.66
Segmentation	PASCAL VOC, mIoU	12.78 ± 0.79	**15.20 ± 0.86**	+2.42 ± 0.27

most skewed sources, and 2.4 to 3.1 points under the three larger sources. TS wins the in-domain target under the ImageNet-100-LT and PASS sources, and SDCLR's pruning schedule was tuned for larger sources, so at 10,000 images its averages fall to 13 to 15.

Pre-training objectives and end-to-end adaptation. Four objectives run paired on ViT-S under their reference recipes and one repair budget, and BRIDGE raises the seven-task linear-probe average of each, by 2.0 points for SimCLR, 2.8 for MoCo v3 and 2.2 for DINO, and by 0.8 for MAE, inside seed variation. Under kNN the MoCo v3 average is unchanged to slightly lower, 33.0 against 32.8, while the MAE average rises 2.0 points. Of the 14 measurements per objective, 12 or 13 improve, and fine-grained targets gain most, up to 7.1 points on Pets for DINO. The same ResNet-50 representations improve after end-to-end adaptation, by 2.0 and 3.2 points of fine-tuning accuracy with 1 and 10 percent of the labels and by 2.4 mIoU on PASCAL VOC segmentation.

Prior methods. Table 5 compares BRIDGE with three methods that also add generated images. BRIDGE has the highest average under both probes, 2.8 linear-probe points and 1.5 kNN points above the best prior method, and the highest accuracy on six of the seven targets. The exception is Flowers, where TADA reaches 39.2 against 36.8. No prior method raises either average by more than 0.3 points over SimCLR. TADA generates variants of the examples a supervised classifier learns slowly. Without labels, we select the 500 images with the highest InfoNCE loss under the current encoder and repair them with the image-to-image step of BRIDGE. AIDE samples images inversely to the size of their cluster in CLIP ViT-B/32 feature space, captions them with BLIP and generates new images from the caption alone. Both use the same generator and image budget as BRIDGE, so the gap comes from which images are chosen and how they are generated. DiffAug trains its own diffusion generator jointly with the encoder and reaches 11.3, and adding BRIDGE repair to it changes the average by less than half a point.

Table 5: Prior methods on ImageNet-100-LT with one source checkpoint per seed and three seeds. Averages cover both probes, and the SimCLR and BRIDGE rows repeat Tables 2 and 3.

Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.	kNN Avg.
SimCLR	73.9 ± 0.5	46.8 ± 0.4	13.7 ± 0.3	19.6 ± 0.2	35.9 ± 7.2	38.6 ± 1.5	50.4 ± 0.5	39.8 ± 1.5	31.8 ± 0.5
DiffAug	35.5 ± 5.7	10.2 ± 3.3	1.1 ± 0.2	2.7 ± 0.1	3.1 ± 2.1	8.2 ± 0.7	18.6 ± 1.5	11.3 ± 1.6	22.0 ± 0.2
DiffAug + BRIDGE	36.1 ± 4.0	10.3 ± 2.4	1.2 ± 0.1	2.6 ± 0.2	1.1 ± 1.4	7.5 ± 2.1	19.0 ± 1.1	11.1 ± 1.4	22.3 ± 0.2
TADA (SSL loss)	73.9 ± 1.5	45.0 ± 2.2	14.4 ± 0.8	15.3 ± 1.4	**39.2 ± 2.4**	40.1 ± 1.4	50.2 ± 1.0	39.7 ± 1.0	31.7 ± 0.8
AIDE (CLIP clusters, BLIP captions)	72.9 ± 0.4	44.8 ± 1.6	14.8 ± 0.0	17.0 ± 0.9	37.7 ± 2.5	40.8 ± 3.3	49.7 ± 0.9	39.7 ± 0.5	32.1 ± 0.3
BRIDGE (ours)	**75.8 ± 0.2**	**49.3 ± 0.2**	**16.4 ± 1.2**	**22.4 ± 0.4**	36.8 ± 1.9	**42.9 ± 1.0**	**53.8 ± 0.1**	**42.5 ± 0.7**	**33.6 ± 0.4**

Table 6: Ablations on ImageNet-100-LT, linear probe, three seeds, defaults shaded. Generated repair matches the real-image oracle, every band window beats uniform and top-tail selection, and five cycles beat two and ten.

Group	Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Generator	Stable Diffusion 3 (default)	**75.0 ± 0.3**	**48.5 ± 0.2**	**16.3 ± 0.8**	**20.7 ± 0.4**	35.7 ± 0.9	41.8 ± 0.8	**52.4 ± 0.7**	**41.5 ± 0.6**
	FLUX	70.7 ± 0.4	42.1 ± 1.4	10.3 ± 1.3	17.5 ± 0.9	29.4 ± 4.6	31.5 ± 2.9	43.7 ± 1.4	35.0 ± 1.9
	Oracle (withheld real images restored)	74.8 ± 0.4	47.8 ± 0.7	13.1 ± 0.1	20.5 ± 0.5	40.3 ± 6.1	38.1 ± 1.5	50.3 ± 0.4	40.7 ± 1.4
	BLIP caption to SD3 text-to-image	72.6 ± 1.0	45.1 ± 1.2	14.3 ± 0.3	17.4 ± 0.6	**35.9 ± 3.7**	**43.3 ± 2.5**	49.2 ± 0.0	39.7 ± 0.1
	SD3 image-to-image with BLIP caption as prompt	70.7 ± 1.0	41.5 ± 1.2	12.7 ± 0.1	14.5 ± 1.1	**35.9 ± 5.2**	39.7 ± 2.4	47.4 ± 0.4	37.5 ± 0.8
Distillation	SD3 latent regression, no generated images	**70.7 ± 1.1**	**42.3 ± 1.1**	**13.2 ± 0.9**	**14.6 ± 1.6**	33.8 ± 4.8	**39.9 ± 2.1**	47.1 ± 0.5	**37.4 ± 1.5**
	SD3 latent regression and BRIDGE	69.6 ± 0.4	40.9 ± 0.6	13.0 ± 0.1	14.1 ± 0.1	**35.3 ± 1.8**	38.9 ± 6.5	**47.8 ± 0.1**	37.1 ± 1.1
Selection	Mode window, lower cutoff p1	75.6 ± 0.1	49.2 ± 0.2	**16.9 ± 0.7**	21.9 ± 0.4	34.6 ± 2.9	43.5 ± 0.7	52.8 ± 0.1	42.1 ± 0.7
	Mode window, lower cutoff p50	75.5 ± 0.1	**49.3 ± 0.4**	15.2 ± 0.3	**22.4 ± 0.6**	35.5 ± 1.1	**43.8 ± 3.1**	53.4 ± 0.9	42.2 ± 0.9
	Mode window, lower cutoff p75 (default)	**76.0 ± 0.3**	49.1 ± 0.2	15.2 ± 1.2	**22.4 ± 0.4**	**36.8 ± 1.9**	42.9 ± 1.0	**53.8 ± 0.1**	**42.3 ± 0.7**
	Top tail	70.8 ± 0.5	43.1 ± 1.0	10.0 ± 1.7	15.9 ± 1.3	27.7 ± 1.0	30.5 ± 1.2	42.2 ± 1.4	34.3 ± 1.2
	Uniform	71.5 ± 0.2	43.8 ± 0.7	9.3 ± 0.9	17.5 ± 1.2	32.9 ± 0.4	31.7 ± 3.9	42.0 ± 0.9	35.5 ± 1.2
	Inverse cluster occupancy	69.9 ± 1.0	40.9 ± 1.9	15.3 ± 0.6	16.5 ± 0.3	33.6 ± 8.4	39.5 ± 2.7	49.4 ± 0.6	37.9 ± 0.6
Cycles	2 cycles	68.9 ± 0.6	41.6 ± 1.3	10.6 ± 1.2	16.4 ± 0.4	30.7 ± 4.5	33.2 ± 0.9	43.3 ± 0.6	35.0 ± 1.4
	5 cycles (default)	**76.0 ± 0.3**	**49.1 ± 0.2**	**15.2 ± 1.2**	**22.4 ± 0.4**	**36.8 ± 1.9**	**42.9 ± 1.0**	**53.8 ± 0.1**	**42.3 ± 0.7**
	10 cycles	70.5 ± 1.2	43.8 ± 0.6	9.2 ± 0.9	17.2 ± 1.6	24.8 ± 5.2	31.3 ± 1.7	41.9 ± 0.6	34.1 ± 1.7
Architecture	ResNet-18	68.8 ± 0.1	41.7 ± 0.2	13.5 ± 0.3	19.7 ± 0.6	38.7 ± 3.4	38.4 ± 2.6	50.2 ± 0.2	38.7 ± 1.1
	ResNet-50 (default)	**76.0 ± 0.3**	**49.1 ± 0.2**	**15.2 ± 1.2**	**22.4 ± 0.4**	36.8 ± 1.9	**42.9 ± 1.0**	**53.8 ± 0.1**	**42.3 ± 0.7**
	ViT-S	62.7 ± 0.4	37.6 ± 0.6	8.8 ± 0.5	12.2 ± 0.2	35.1 ± 4.2	30.2 ± 1.6	42.5 ± 1.2	32.7 ± 1.3
	ViT-B	66.0 ± 2.3	40.6 ± 2.3	8.3 ± 0.5	14.3 ± 0.8	**38.8 ± 2.6**	31.3 ± 2.1	44.5 ± 1.7	34.8 ± 1.8

6.1 Ablations

Generator. Generated repair with Stable Diffusion 3 matches the oracle of Section 5, 41.5 against 40.7. With FLUX at six sampling steps and guidance 2.5, the average falls to 35.0, 6.5 points below Stable Diffusion 3. With the BLIP captions of AIDE at the anchors of BRIDGE, generating from the caption alone reaches 39.7 and passing it as the prompt of the image-to-image step reaches 37.5.

Selection. The three mode windows with lower cutoffs at the 1st, 50th and 75th percentile reach 42.1 to 42.3 and are not separated, while uniform acquisition reaches 35.5 ± 1.2 and the top tail 34.3 ± 1.2 (Welch t-tests with Holm adjustment, $p \leq .011$ for every window against both). This agrees with Proposition 3, where excluding the extreme tail raises the objective, and leaves the lower boundary of the band unresolved. Each window changes trimming and diversification at once, and the 6.6 to 8.0 points include both effects. The inverse-occupancy rule of AIDE on $\sqrt{N}$ spherical k-means clusters of the encoder features reaches 37.9.

Distillation. A linear head on the encoder regresses the averaged Stable Diffusion 3 VAE latent of each training image, with a cosine loss of weight 0.1 added to the SimCLR loss. The average is 37.4 without generated images and 37.1 with BRIDGE repair, 5.2 points below BRIDGE alone, so the gain of BRIDGE comes from the generated images at the selected anchors.

Cycles and backbone. At fixed total repair volume and optimizer budget, five cycles improve the seven-task average over two by 7.3 points and remain 8.2 points ahead of ten. Two cycles spend most of the budget before the representation has changed, and ten leave only 50 epochs of training after each repair. Among the tested backbones, ResNet-50 has the highest average at this source size.

6.2 Representation geometry

Under the frozen cycle-0 encoder (Figure 3), inserting the generated variants reduces the radius of the treated anchors by 4.7 percent against 2.4 percent for unselected images matched within class on initial radius, a paired difference of -2.3 points. Purity falls by 0.6 points at the anchors and by 2.0

Table 7: kNN counterpart of Table 6.

Group	Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Generator	Stable Diffusion 3 (default)	68.6 ± 0.1	37.2 ± 0.2	6.4 ± 0.6	11.5 ± 0.2	**35.2 ± 0.3**	**24.3 ± 0.6**	**46.8 ± 0.7**	**32.9 ± 0.4**
	FLUX	65.8 ± 1.0	35.4 ± 0.8	5.7 ± 0.7	9.4 ± 0.9	27.7 ± 1.0	18.8 ± 1.3	38.7 ± 1.2	28.8 ± 1.0
	Oracle (withheld real images restored)	68.8 ± 0.2	37.7 ± 0.3	6.6 ± 0.5	11.4 ± 0.5	32.7 ± 0.1	23.6 ± 0.6	44.4 ± 0.4	32.2 ± 0.4
	BLIP caption to SD3 text-to-image	**70.1 ± 0.4**	**39.6 ± 0.8**	7.2 ± 0.5	**11.7 ± 0.4**	31.9 ± 1.6	21.1 ± 0.9	41.6 ± 0.9	31.9 ± 0.5
	SD3 image-to-image with BLIP caption as prompt	69.6 ± 0.1	39.1 ± 0.6	**7.4 ± 0.3**	**11.7 ± 0.2**	31.5 ± 0.8	20.8 ± 0.4	41.2 ± 0.6	31.6 ± 0.3
Distillation	SD3 latent regression, no generated images	**70.3 ± 0.2**	**39.3 ± 0.2**	7.2 ± 0.3	11.3 ± 0.4	30.5 ± 0.6	**21.8 ± 0.7**	**41.1 ± 0.3**	**31.7 ± 0.1**
	SD3 latent regression and BRIDGE	68.6 ± 0.7	38.6 ± 0.4	**7.5 ± 0.6**	**11.6 ± 0.3**	**31.0 ± 0.9**	20.8 ± 1.0	40.4 ± 0.2	31.2 ± 0.5
Selection	Mode window, lower cutoff p1	**69.2 ± 0.3**	37.7 ± 0.2	6.6 ± 0.2	11.4 ± 0.2	35.9 ± 0.2	25.0 ± 0.5	48.3 ± 0.3	33.4 ± 0.3
	Mode window, lower cutoff p50	68.7 ± 0.1	38.0 ± 0.2	7.3 ± 0.2	11.3 ± 0.2	**36.5 ± 0.1**	24.8 ± 0.8	**48.5 ± 0.4**	**33.6 ± 0.3**
	Mode window, lower cutoff p75 (default)	69.0 ± 0.5	37.8 ± 0.3	6.7 ± 0.3	11.7 ± 0.2	**36.5 ± 0.4**	**25.3 ± 1.0**	48.1 ± 0.6	**33.6 ± 0.5**
	Top tail	65.7 ± 0.8	34.4 ± 0.2	4.1 ± 0.4	9.6 ± 0.2	23.9 ± 0.9	18.1 ± 0.8	35.4 ± 1.2	27.3 ± 0.6
	Uniform	66.1 ± 1.0	35.7 ± 0.3	4.9 ± 0.1	10.2 ± 0.7	25.3 ± 0.4	19.0 ± 0.8	37.4 ± 0.6	28.4 ± 0.6
	Inverse cluster occupancy	68.7 ± 0.3	**38.6 ± 0.2**	**7.7 ± 0.2**	**11.8 ± 0.3**	32.1 ± 0.9	21.4 ± 0.6	41.2 ± 0.4	31.6 ± 0.2
Cycles	2 cycles	65.3 ± 0.4	34.2 ± 0.5	5.0 ± 0.7	10.0 ± 0.2	26.1 ± 0.8	18.7 ± 0.6	35.9 ± 0.3	27.9 ± 0.5
	5 cycles (default)	**69.0 ± 0.5**	**37.8 ± 0.3**	**6.7 ± 0.3**	**11.7 ± 0.2**	**36.5 ± 0.4**	**25.3 ± 1.0**	**48.1 ± 0.6**	**33.6 ± 0.5**
	10 cycles	66.4 ± 0.5	35.4 ± 0.3	4.8 ± 0.3	10.4 ± 0.6	25.2 ± 0.8	18.6 ± 0.9	36.0 ± 0.7	28.1 ± 0.6
Architecture	ResNet-18	66.3 ± 0.2	35.7 ± 0.3	5.9 ± 0.2	11.1 ± 0.4	33.5 ± 0.1	23.3 ± 0.5	44.5 ± 0.7	31.5 ± 0.4
	ResNet-50 (default)	69.0 ± 0.5	37.8 ± 0.3	**6.7 ± 0.3**	**11.7 ± 0.2**	36.5 ± 0.4	**25.3 ± 1.0**	**48.1 ± 0.6**	**33.6 ± 0.5**
	ViT-S	67.3 ± 0.4	37.5 ± 0.3	5.3 ± 0.2	7.6 ± 0.1	36.2 ± 0.7	20.1 ± 1.0	42.1 ± 1.0	30.9 ± 0.5
	ViT-B	**69.3 ± 0.8**	**40.1 ± 0.8**	6.0 ± 0.2	8.3 ± 0.2	**39.8 ± 0.8**	21.9 ± 1.4	45.1 ± 0.7	32.9 ± 0.7

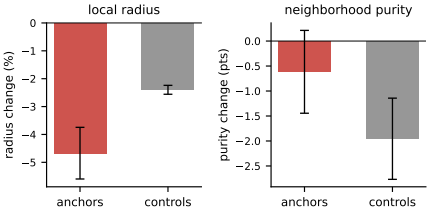

Figure 3: With the encoder frozen, generated images shrink the neighborhoods of their own anchors about twice as much as those of matched unselected images. Seed means with one standard deviation.

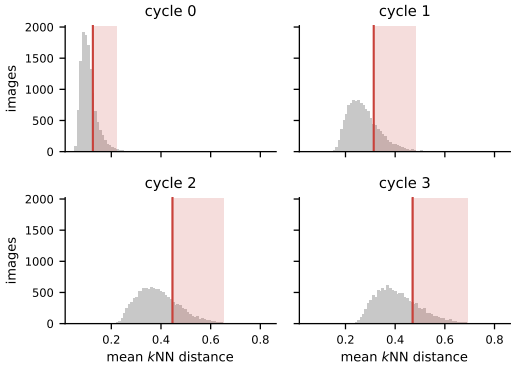

Figure 4: Scores of one ImageNet-100-LT run over four repair cycles, 75th to 99th percentile band shaded. The distribution shifts at every cycle, so the band is re-derived each time.

at the matched images, so the variants inserted near the anchors mostly share the query's class. The diagnostic uses $k = 20$, and per-seed values are in Appendix B.

7 Limitations

BRIDGE depends on informative encoder neighborhoods in the first cycle and on a frozen generator that imports the biases of its own pretraining data. Repair fidelity, the ρ of Section 3, is not measured in pixel or semantic space, and the deficit of top-tail selection is consistent with low fidelity there. Measured fidelity could set the band edges in place of fixed percentiles. Later cycles select generated images as anchors, 7 to 13 percent at cycle 1 and 34 to 45 percent at cycle 4, with no filter applied. The band, neighborhood size and budget were held at their defaults across all five sources. The loop adds 30 to 45 percent wall-clock time (Appendix E).

8 Conclusion

In this paper we introduce BRIDGE, which lets a self-supervised learner fill the sparse regions of its own representation and improves transfer from every long-tailed source we tested. Web-scale pretraining corpora are long-tailed in the same way as our sources, and as generators improve, the representation itself could decide at each stage of large-scale pretraining which regions to fill, making data curation part of training. The same loop applies to any modality with a pretrained generator, such as audio, video or molecules, which should be explored in further research.

AI USE STATEMENT

In this work we used generative AI tools for writing and debugging the data generators, the training and evaluation scripts and the analysis code, for tabulating results, and for editing the text of the manuscript under the direction of the authors. We also used them to check the bibliography, and every entry was subsequently verified against its publisher, proceedings or arXiv record by the authors. We have not used generative AI tools for generating research ideas. The generated images that enter pretraining come from the frozen image generator described in Section 4, and no assistant-written text or code is part of that data. All AI-assisted code was reviewed and tested by the authors, every number in the paper was checked against the evaluation outputs, and all text was reviewed and revised by the authors. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Section 4 and Algorithm 1 specify the loop, Section 5 and Appendix D give every operating parameter, schedule, generator setting and data split, and Appendix C contains complete proofs with their assumptions stated in Section 3. Appendix H reports every per-task number behind the main-text averages. An anonymous code release with configuration files for every experiment is linked in Appendix D.

ETHICS STATEMENT

The work uses public datasets under their release terms and a frozen pretrained image generator, releases no new generator, scraped collection or generated corpus, and involves no human subjects. Generative repair inherits the biases and memorization risks of the generator's pretraining data, and a deployment would require a task-specific review of fairness, privacy and content safety, as Appendix G discusses.

REFERENCES

Amro Kamal Mohamed Abbas, Evgenia Rusak, Kushal Tirumala, Wieland Brendel, Kamalika Chaudhuri, and Ari S. Morcos. Effective pruning of web-scale datasets based on complexity of concept clusters. In *International Conference on Learning Representations*, 2024. doi: 10.48550/arXiv.2401.04578.

Yuki Asano, Christian Rupprecht, Andrew Zisserman, and Andrea Vedaldi. PASS: An ImageNet replacement for self-supervised pretraining without humans. In Joaquin Vanschoren and Serena Yeung, editors, *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. doi: 10.48550/arXiv.2109.13228.

Mahmoud Assran, Randall Balestriero, Quentin Duval, Florian Bordes, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, and Nicolas Ballas. The hidden uniform cluster prior in self-supervised learning. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2210.07277.

Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J. Fleet. Synthetic data from diffusion models improves ImageNet classification. *Transactions on Machine Learning Research*, 2023. doi: 10.48550/arXiv.2304.08466.

Jianhong Bai, Zuozhu Liu, Hualiang Wang, Jin Hao, Yang Feng, Huanpeng Chu, and Haoji Hu. On the effectiveness of out-of-distribution data in self-supervised long-tail learning. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2306.04934.

Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 2556–2569. Curran Associates, Inc., 2021. doi: 10.48550/arXiv.2106.05963.

Manel Baradad, Chun-Fu (Richard) Chen, Jonas Wulff, Tongzhou Wang, Rogerio Feris, Antonio Torralba, and Phillip Isola. Procedural image programs for representation learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 6450–6462. Curran Associates, Inc., 2022. doi: 10.52202/068431-0467.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9630–9640. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00951.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16: 321–357, 2002. doi: 10.1613/jair.953.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 13–18 Jul 2020. doi: 10.48550/arXiv.2002.05709.

Lijie Fan, Kaifeng Chen, Dilip Krishnan, Dina Katabi, Phillip Isola, and Yonglong Tian. Scaling laws of synthetic images for model training ... for now. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7382–7392. IEEE, 2024. doi: 10.1109/CVPR52733.2024.00705.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. doi: 10.1016/0304-3975(85)90224-5.

Guy Hacohen, Avihu Dekel, and Daphna Weinshall. Active learning on a budget: Opposite strategies suit high and low budgets. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8175–8195. PMLR, 17–23 Jul 2022. doi: 10.48550/arXiv.2202.02794.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735. IEEE, 2020. doi: 10.1109/CVPR42600.2020.00975.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01553.

Ruifei He, Shuyang Sun, Xin Yu, Chuhui Xue, Wenqing Zhang, Philip Torr, Song Bai, and Xiaojuan Qi. Is synthetic data from generative models ready for image recognition? In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2210.07574.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. doi: 10.48550/arXiv.2006.11239.

Ali Jahanian, Xavier Puig, Yonglong Tian, and Phillip Isola. Generative models as a data source for multiview representation learning. In *International Conference on Learning Representations*, 2022. doi: 10.48550/arXiv.2106.05258.

Ziyu Jiang, Tianlong Chen, Bobak J. Mortazavi, and Zhangyang Wang. Self-damaging contrastive learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4927–4939. PMLR, 18–24 Jul 2021. doi: 10.48550/arXiv.2106.02990.

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *2013 IEEE International Conference on Computer Vision Workshops*, pages 554–561. IEEE, 2013. doi: 10.1109/ICCVW.2013.77.

Alex Krizhevsky. Learning multiple layers of features from Tiny Images. Technical report, University of Toronto, 2009. URL `https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf`.

Anna Kukleva, Moritz Böhle, Bernt Schiele, Hilde Kuehne, and Christian Rupprecht. Temperature schedules for self-supervised contrastive methods on long-tail data. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2303.13664.

Daiqing Li, Huan Ling, Amlan Kar, David Acuna, Seung Wook Kim, Karsten Kreis, Antonio Torralba, and Sanja Fidler. DreamTeacher: Pretraining image backbones with deep generative models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16652–16662. IEEE, 2023. doi: 10.1109/ICCV51070.2023.01531.

Mingfu Liang, Jong-Chyi Su, Samuel Schulter, Sparsh Garg, Shiyu Zhao, Ying Wu, and Manmohan Chandraker. AIDE: An automatic data engine for object detection in autonomous driving. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14695–14706. IEEE, 2024. doi: 10.1109/CVPR52733.2024.01392.

Swee Kiat Lim, Yi Loo, Ngoc-Trung Tran, Ngai-Man Cheung, Gemma Roig, and Yuval Elovici. DOPING: Generative data augmentation for unsupervised anomaly detection with GAN. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1122–1127. IEEE, 2018. doi: 10.1109/ICDM.2018.00146.

Ci-Siang Lin, Min-Hung Chen, and Yu-Chiang Frank Wang. Frequency-aware self-supervised long-tailed learning. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 963–972. IEEE, 2023. doi: 10.1109/ICCVW60793.2023.00103.

Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X. Yu. Large-scale long-tailed recognition in an open world. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2532–2541. IEEE, 2019. doi: 10.1109/CVPR.2019.00264.

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew B. Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. doi: 10.48550/arXiv.1306.5151.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. doi: 10.48550/arXiv.2108.01073.

Dang Nguyen, Jiping Li, Jinghao Zheng, and Baharan Mirzasoleiman. Do we need all the synthetic data? Targeted image augmentation via diffusion models. In *International Conference on Learning Representations*, 2026. doi: 10.48550/arXiv.2505.21574.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, pages 722–729. IEEE, 2008. doi: 10.1109/ICVGIP.2008.47.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. doi: 10.48550/arXiv.2304.07193.

Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3498–3505. IEEE, 2012. doi: 10.1109/CVPR.2012.6248092.

Zhongzheng Ren and Yong Jae Lee. Cross-domain self-supervised multi-task feature learning using synthetic imagery. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 762–771. IEEE, 2018. doi: 10.1109/CVPR.2018.00086.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01042.

Mert Bulent Sariyildiz, Karteek Alahari, Diane Larlus, and Yannis Kalantidis. Fake it till you make it: Learning transferable representations from synthetic ImageNet clones. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8011–8021. IEEE, 2023. doi: 10.1109/CVPR52729.2023.00774.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018. doi: 10.48550/arXiv.1708.00489.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025. doi: 10.48550/arXiv.2508.10104.

Ben Sorscher, Robert Geirhos, Shashank Shekhar, Surya Ganguli, and Ari S. Morcos. Beyond neural scaling laws: Beating power law scaling via data pruning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 19523–19536. Curran Associates, Inc., 2022. doi: 10.52202/068431-1419.

Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 20827–20840. PMLR, 17–23 Jul 2022. doi: 10.48550/arXiv.2204.06507.

Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 1363–1389. Curran Associates, Inc., 2023. doi: 10.52202/075280-0068.

Yonglong Tian, Olivier J. Hénaff, and Aäron van den Oord. Divide and contrast: Self-supervised learning from uncurated data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10043–10054. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00991.

Yonglong Tian, Lijie Fan, Phillip Isola, Huiwen Chang, and Dilip Krishnan. StableRep: Synthetic images from text-to-image models make strong visual representation learners. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 48382–48402. Curran Associates, Inc., 2023. doi: 10.52202/075280-2098.

Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective data augmentation with diffusion models. In *International Conference on Learning Representations*, 2024. doi: 10.48550/arXiv.2302.07944.

Zijie J. Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. DiffusionDB: A large-scale prompt gallery dataset for text-to-image generative models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 893–911, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.51.

Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. Data collection and quality challenges in deep learning: A data-centric AI perspective. *The VLDB Journal*, 32(4):791–813, 2023. doi: 10.1007/s00778-022-00775-9.

Nyla Worker and Bhumin Pathak. Build custom synthetic data generation pipelines with Omniverse Replicator. NVIDIA Technical Blog, June 2022. URL https://developer.nvidia.com/blog/build-custom-synthetic-data-generation-pipelines-with-omniverse-replicator/.

Ofer Yehuda, Avihu Dekel, Guy Hacohen, and Daphna Weinshall. Active learning through a covering lens. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 22354–22367. Curran Associates, Inc., 2022. doi: 10.52202/068431-1624.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In *International Conference on Learning Representations*, 2025. doi: 10.48550/arXiv.2410.06940.

Zelin Zang, Hao Luo, Kai Wang, Panpan Zhang, Fan Wang, Stan Z. Li, and Yang You. DiffAug: Enhance unsupervised contrastive learning with domain-knowledge-free diffusion-based data augmentation. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 58174–58196. PMLR, 21–27 Jul 2024. doi: 10.48550/arXiv.2309.07909.

A Comparison with Related Acquisition Methods

Classical oversampling and TADA receive labels or supervised learning signals (Table 8). DOPING uses latent rarity, targets anomaly detection and does not close the loop through a changing SSL representation. GenRep and StableRep make synthetic images or latent neighborhoods the source of contrastive views without allocating a repair budget over an observed skewed corpus. DiffAug is the prior SSL method most similar to ours, since it also iterates between encoder and generator, and it jointly learns a generator for generic positive views without asking where a frozen generator should intervene. AIDE and industrial data engines retrain a model on data collected where it fails a supervised task, using VLM and LLM reasoning, annotation and domain-specific verification. BRIDGE studies the label-free case in which local support in the current SSL representation is both the diagnostic and the acquisition signal.

Table 8: Closest methods and the dimensions relevant to the acquisition problem. Feedback means that the learner or task model changes the next acquisition decision, and the methods are cited in Section 2.

Method	Labels or task loss	Acquisition signal	Data action	Feedback
SMOTE	class labels	class minority	interpolate examples	no
DOPING	no labels	latent edge density	latent oversampling	no
Concept pruning	no task labels	cluster complexity	remove examples	no
DA-Fusion	labels or few-shot set	class and exemplar	image-conditioned diffusion	no
TADA	supervised loss dynamics	slow-learned examples	faithful diffusion variants	no
GenRep and StableRep	prompts or generator latent	global sampling	synthetic SSL views	no
DiffAug	no labels	learned positive-view objective	jointly learned diffusion views	yes
AIDE	task labels and failures	VLM and LLM failure analysis	curate, generate, annotate	yes
BRIDGE	no labels	local support in current SSL space	budgeted local variants	yes

B Representation Diagnostics

The fixed-encoder intervention test freezes the cycle-0 encoder and queries the same original images before and after inserting all generated samples, so any change reflects the inserted points alone. First-cycle anchors are matched one-to-one within class to original images with the closest initial kNN radius that are never selected in any cycle, and Table 9 reports the per-seed values behind Figure 3. Figure 5 draws every paired ViT-S measurement of Table 4 as a difference. The learned-space comparison retrains the encoder after the intervention and evaluates the same original-image

Table 9: Fixed-encoder intervention test behind Figure 3, cycle-0 encoder, $k = 20$. Generated images shrink the radius of their own anchors about twice as much as that of matched unselected images, and lower anchor purity less. The last two columns are anchor-minus-matched differences.

	Radius change (%)		Purity change (pts)		Difference	
Seed	Anchors	Matched	Anchors	Matched	Radius	Purity
0	-5.38	-2.27	-1.57	-2.79	-3.11	$+1.22$
1	-5.01	-2.35	-0.20	-1.91	-2.66	$+1.70$
2	-3.62	-2.58	-0.07	-1.17	-1.05	$+1.09$
Mean	-4.67	-2.40	-0.61	-1.96	-2.27	$+1.34$

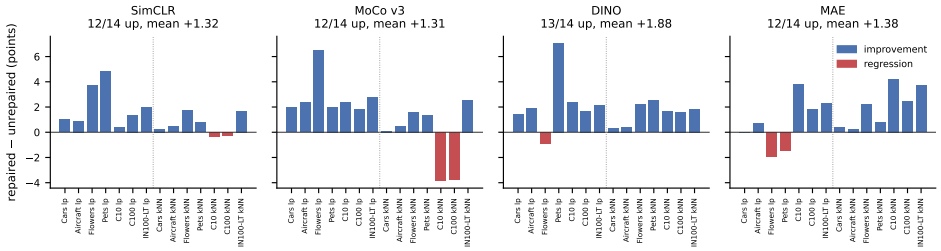

Figure 5: Repair against no repair for four objectives on ViT-S, one bar per downstream measurement (linear probe and kNN on each of seven datasets), from the paired runs of Table 4. Blue bars are improvements. Red bars are regressions, two for SimCLR and MoCo v3 (CIFAR-10 and CIFAR-100 under kNN), one for DINO (Flowers) and two for MAE (Flowers and Pets under the linear probe).

panel. On that panel MoCo v3 on ViT-S gains 21.97 ± 1.58 in effective rank, 0.296 ± 0.019 in spectral entropy and 3.80 ± 0.80 points of 1-NN accuracy after repair, and the first two survive Holm correction, whereas SimCLR improves transfer while its effective rank falls and its local effect is mixed. We therefore use the fixed-encoder result to locate where the variants land in representation space and the learned-space statistics as objective-dependent diagnostics.

Where the temperature-schedule hybrid helps. Combining BRIDGE with TS lowers the seven-task average in every source regime, and it wins 7 of the 70 individual cells (Appendix Tables 10 and 11). Six of those seven fall on CIFAR-10 and ImageNet-100-LT, the two long-tailed downstream targets, and the largest are $+3.3$ and $+2.8$ points on ImageNet-100-LT and CIFAR-10 under a PASS and a CIFAR-10-LT source. The fine-grained targets lose in almost every case. TS reweights the contrastive temperature toward rare samples, which sharpens the head-to-tail boundary that the long-tailed targets score, and the hybrid gains on those targets. BRIDGE alone gains more on the four fine-grained targets.

C Proofs for the Support-Allocation Framework

Proposition 6 (Closed-form interior optimum)**.** *Let $b(d) = b_0 d^{\beta}$ with $b_0 > 0$ and $\beta \in (0, 1]$, and let $\rho(d) = \rho_0 e^{-\mu d}$ with $\mu > 0$. Then g is strictly decreasing on $(0, \infty)$ for every $\lambda \geq 0$, so Proposition 3 applies whenever g changes sign on the interval. If $\lambda = 0$ the maximizer is $d^{\star} = \beta/\mu$, independent of b_0 and ρ_0.*

Proof of Proposition 1. Sort the distances from z to points in Z and denote their kth order statistic by $r_k(z; Z)$. Adding A preserves every old candidate distance and may introduce smaller ones. The kth order statistic can therefore only decrease. It decreases strictly when an added distance enters the first k positions and displaces a strictly larger distance. □

Proof of Proposition 2. Let $x = n_j + m_j + \kappa$. Since $f(x) = x^{-\gamma}$ is positive, strictly decreasing and strictly convex for $x > 0$, the forward difference $f(x) - f(x + 1)$ is positive and strictly decreases with x. Multiplication by $p_j a_j > 0$ preserves both properties. Equal $p_j a_j$ therefore leaves effective support as the sole determinant of the marginal ordering. □

Proof of Proposition 3. Rewrite the utility as $u(d) = \rho(d)(b(d) + \lambda) - \lambda$. Its derivative is

$$u'(d) = \rho(d)(b(d) + \lambda) \left[\frac{b'(d)}{b(d) + \lambda} + \frac{\rho'(d)}{\rho(d)} \right] = \rho(d)(b(d) + \lambda)\, g(d).$$

The factor preceding $g(d)$ is positive. Since g is strictly decreasing and changes sign across the interval, it has exactly one root $d^\star$ by continuity. Hence u increases before $d^\star$ and decreases after it, making $d^\star$ the unique global maximizer. □

Proof of Proposition 6. With $b(d) = b_0 d^\beta$ and $\rho(d) = \rho_0 e^{-\mu d}$ we have $\rho'(d)/\rho(d) = -\mu$, so

$$g(d) = \frac{b_0 \beta d^{\beta-1}}{b_0 d^\beta + \lambda} - \mu = \frac{\beta d^{\beta-1}}{d^\beta + c} - \mu, \qquad c = \lambda / b_0 \geq 0.$$

Write $h(d) = \beta d^{\beta-1}(d^\beta + c)^{-1}$. Differentiating,

$$h'(d) = \frac{\beta d^{\beta-2}\big[(\beta-1)(d^\beta + c) - \beta d^\beta\big]}{(d^\beta + c)^2} = \frac{\beta d^{\beta-2}\big[(\beta-1)c - d^\beta\big]}{(d^\beta + c)^2}.$$

For $\beta \in (0, 1]$ we have $(\beta - 1)c \leq 0 < d^\beta$ for all $d > 0$, so the bracket is strictly negative and $h'(d) < 0$. Hence $g = h - \mu$ is strictly decreasing for every $\lambda \geq 0$. When $\lambda = 0$ we have $c = 0$ and $g(d) = \beta/d - \mu$, with unique root is $d^\star = \beta/\mu$. Since g is positive below it and negative above it, $d^\star$ is the maximizer by the argument of Proposition 3. □

Proof of Proposition 4. For region j define the marginal sequence $\delta_{j,t} = U_j(n_j + t) - U_j(n_j + t - 1)$ for $t \geq 1$, which discrete concavity makes nonincreasing, so every feasible allocation of R repairs selects a prefix of length m_j from each sequence and its utility is the baseline plus the selected prefix marginals. After t greedy steps the selected set is the t largest prefix-feasible marginals, and the next greedy pick is the largest remaining prefix-feasible marginal, so the set stays the top $t + 1$ and the claim follows by induction on t. No feasible allocation can therefore exceed the greedy total. For the second sentence take two regions with marginal sequences $(3, 1, 1)$ and $(2, 2, 2)$ and a budget of three, where the fixed allocation $(3, 0)$ built from the initial ordering yields 5 while greedy yields $3 + 2 + 2 = 7$. □

Proof of Proposition 5. Write $f(x) = x^{-\gamma}$. For $x > 0$ the mean value theorem gives $f(x) - f(x + 1) = \gamma \xi^{-\gamma-1}$ for some $\xi \in (x, x + 1)$, so

$$0 < \Delta_j(m_j) = p_j a_j \big[f(n_j + m_j + \kappa) - f(n_j + m_j + 1 + \kappa) \big] \leq p_j a_j \gamma (n_j + \kappa)^{-\gamma-1},$$

using $\xi > n_j + m_j + \kappa \geq n_j + \kappa$ and $m_j \geq 0$. Both $\mathcal{B}(\mathbf{m})$ and $\mathcal{B}(\mathbf{m}')$ are obtained from the common baseline $\mathcal{B}(\mathbf{0})$ by subtracting sums of at most R such marginals, since $\sum_j m_j = \sum_j m'_j = R$. Each sum lies in $[0,\, R\gamma \max_j p_j a_j (n_j + \kappa)^{-\gamma-1}]$, and the difference of two numbers drawn from an interval is bounded by its length. Monotonicity of $(n + \kappa)^{-\gamma-1}$ in n gives the $n_{\min}$ form. □

D Implementation and Reproducibility

Training schedule. The encoder is warm-started from the previous cycle and is never retrained from scratch. Each cycle uses a fresh optimizer and learning-rate schedule. All earlier real and generated images remain in the source set, so the setting differs from sequential-task learning in which previous data disappear. The cycle ablation holds the total optimizer update budget and the total number of generated images fixed and compares fewer repairs with more learning between them against more frequent repairs with less.

Core configuration. The main experiments use ResNet-50 encoders, five cycles of 100 epochs, mean kNN distance with $k = 100$, 500 anchors per cycle and five variants per anchor. The optimizer is AdamW with weight decay 10^{-4}, and method-specific learning rates and temperatures follow the settings of each baseline. The paired ViT-S experiments use the reference recipe of each objective. DINO uses two global and eight local crops at 96×96 with a bottleneck projection head, and MAE runs five cycles of 80 epochs.

Scoring and selection. Scoring uses the mean squared normalized ℓ_2 distance of Equation 4. For unit vectors the squared distance is an affine function of the cosine, so cosine scoring selects the identical anchor set. The mode-window selector restricts the score distribution to the 75th to 99th percentile range, builds a histogram with the automatic bin heuristic, forms a mode-centered candidate pool of size $4B$ and applies greedy farthest-point sampling on normalized features to select the final B anchors.

Generators. The main method uses Stable Diffusion 3 Medium in image-to-image mode, where an empty prompt makes the guidance scale inert. The generation ablation replaces it with FLUX or with the oracle, which adds the withheld real images and runs no generator. In the PASS and DiffusionDB experiments the SD3 batch size is 12 to fit the 24-hour wall-time limit. The DiffusionDB source is itself synthetic and shows the stylistic bias of the generator that produced it, so the added variants there are synthetic images derived from synthetic anchors.

Source data. PASS and DiffusionDB use `train[:10000]` with pretraining split override `[1.0, 0.0, 0.0]`. The long-tailed ImageNet-100, CIFAR-10 and CIFAR-100 sources use their standard long-tailed configurations without the 10,000-image restriction. The ImageNet-100 long tail is built from the `clane9/imagenet-100` train and validation split. Class ranks 1 to 100 receive Pareto weights $r^{-1/\alpha}$ with $\alpha = 2.5$, the weights are min-max normalized and mapped linearly onto retention counts from 1,280 down to 5 images, the resulting per-class ratios are permuted uniformly at random over the classes, and each image is retained by an independent Bernoulli draw with the ratio of its class. The balanced source of Table 1 reads the same split and replaces the per-class ratios by their mean, so it removes the same expected fraction of images uniformly over classes.

Seeds and reporting. All claims rest on three-seed averages reported as mean and standard deviation. Paired comparisons share seeds and, where stated, source checkpoints. Multiple comparisons within one table are corrected by the Holm procedure. No two conditions within a table share an encoder fingerprint.

Code availability. An anonymous code release is available at `https://anonymous.4open.science/status/FOMO-0488`.

E Compute Resources and Runtime

Most experiments use one node with two NVIDIA A100 80GB GPUs under a 24-hour wall-time limit. The DINO runs fit on a single 48GB card at the batch size used. The 454 completed runs retained for this paper total approximately 5,471 wall-clock hours, of which the baselines and ablations that appear only in the appendix account for 2,987. Runs on two-GPU nodes consume roughly twice that in GPU hours, while the paired ViT-S runs hold a single card each. Failed runs, timeouts, cache migration and pilot jobs are excluded from these totals.

Mean wall-clock times over three seeds are 16.9 hours for SimCLR and 21.8 for BRIDGE on ImageNet-100-LT, 10.2 and 14.9 on PASS, and 11.6 and 16.4 on DiffusionDB. The loop therefore raises wall-clock time by roughly 30 to 45 percent over the corresponding no-repair run, of which the embedding pass is a small fraction and diffusion sampling the remainder. Four repairs add 10,000 images, which is 10 percent of a 100,000-image source and 1 percent of a one-million-image source. Whether a fixed budget still helps at 100,000 or one million images was not tested.

F Assets and Terms of Use

ImageNet-100-LT is derived from ImageNet. The ImageNet access terms restrict the images to non-commercial research and educational use. PASS is released under Creative Commons Attribution 4.0 and DiffusionDB under CC0-1.0. CIFAR-10, CIFAR-100, Stanford Cars, FGVC Aircraft, Oxford Flowers and Oxford-IIIT Pets are used from their official academic releases under their original terms. Stable Diffusion 3 Medium is used under the Stability AI Community License. FLUX.1-dev is used only in the ablation under the FLUX.1-dev Non-Commercial License. We do not redistribute model weights, third-party datasets or generated image corpora.

G Broader Impact

A method that improves representation quality under biased source data can reduce the need for manual relabeling and can make SSL more useful where balanced data are difficult to collect. The method also relies on large image generators and web-sourced datasets, which raise risks of license violations, privacy and unintended memorization or harmful content. Our experiments stay within existing dataset and model release terms, release no new generators or scraped collections, and any deployment would need a review of these risks for its own task.

H Full Results

Table 10: Full linear-probe transfer results across all source datasets and comparison methods.

Source	Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ImageNet-100-LT	SimCLR	73.9 ± 0.5	46.8 ± 0.4	13.7 ± 0.3	19.6 ± 0.2	35.9 ± 7.2	38.6 ± 1.5	50.4 ± 0.5	39.8 ± 1.5
	TS	72.5 ± 0.4	44.0 ± 0.5	12.5 ± 1.2	15.0 ± 0.1	35.7 ± 1.8	40.6 ± 2.2	**56.7 ± 1.2**	39.6 ± 1.1
	SDCLR	41.1 ± 1.7	18.0 ± 0.6	1.4 ± 0.3	2.3 ± 0.5	10.8 ± 1.7	10.4 ± 0.3	21.9 ± 0.6	15.1 ± 0.8
	BRIDGE (ours)	**75.8 ± 0.2**	**49.3 ± 0.2**	**16.4 ± 1.2**	**22.4 ± 0.4**	**36.8 ± 1.9**	**42.9 ± 1.0**	53.8 ± 0.1	**42.5 ± 0.7**
	BRIDGE+TS (ours)	74.2 ± 0.5	45.6 ± 0.8	9.4 ± 0.5	15.0 ± 0.2	25.5 ± 2.5	33.1 ± 2.8	48.3 ± 0.4	35.9 ± 1.1
	BRIDGE+SDCLR (ours)	43.8 ± 1.7	18.4 ± 0.7	2.0 ± 0.3	2.5 ± 0.8	10.0 ± 0.8	12.2 ± 3.1	21.8 ± 0.5	15.8 ± 1.1
CIFAR-10-LT	SimCLR	67.8 ± 0.2	38.0 ± 1.0	6.5 ± 0.4	16.0 ± 0.9	**19.2 ± 1.6**	24.0 ± 1.3	32.0 ± 0.7	29.1 ± 0.9
	TS	68.5 ± 1.3	36.0 ± 1.7	4.3 ± 0.2	10.3 ± 0.1	13.1 ± 2.0	19.7 ± 2.3	31.5 ± 1.2	26.2 ± 1.3
	SDCLR	39.6 ± 1.1	15.6 ± 0.2	1.1 ± 0.1	2.4 ± 0.9	12.0 ± 1.6	9.1 ± 1.8	16.5 ± 1.6	13.8 ± 1.1
	BRIDGE (ours)	78.1 ± 0.3	**49.8 ± 0.4**	**9.2 ± 0.8**	**17.7 ± 1.0**	18.2 ± 1.1	**27.4 ± 2.7**	39.0 ± 0.6	**34.2 ± 1.0**
	BRIDGE+TS (ours)	**78.6 ± 0.3**	46.1 ± 0.3	7.5 ± 0.9	11.1 ± 0.7	14.5 ± 0.6	25.1 ± 2.7	**39.6 ± 0.8**	31.8 ± 0.9
	BRIDGE+SDCLR (ours)	40.9 ± 0.7	17.9 ± 0.5	1.4 ± 0.1	2.2 ± 0.6	11.0 ± 2.4	10.0 ± 2.7	21.0 ± 0.6	14.9 ± 1.1
CIFAR-100-LT	SimCLR	59.0 ± 1.5	31.1 ± 1.6	4.1 ± 0.3	11.2 ± 1.0	12.6 ± 3.4	20.1 ± 1.0	28.7 ± 1.3	23.8 ± 1.4
	TS	56.3 ± 0.2	28.0 ± 0.1	2.8 ± 0.6	8.5 ± 0.2	15.5 ± 2.5	16.2 ± 3.1	26.2 ± 1.8	21.9 ± 1.2
	SDCLR	37.1 ± 0.8	15.7 ± 0.2	1.9 ± 0.1	2.4 ± 0.7	9.4 ± 1.0	9.9 ± 1.7	16.2 ± 0.3	13.2 ± 0.7
	BRIDGE (ours)	**74.2 ± 0.5**	**48.1 ± 0.4**	**8.3 ± 1.0**	**18.2 ± 0.7**	**24.4 ± 2.3**	27.6 ± 0.7	39.8 ± 1.1	**34.4 ± 1.0**
	BRIDGE+TS (ours)	73.7 ± 0.1	47.3 ± 0.5	7.9 ± 0.6	14.1 ± 0.5	18.2 ± 3.2	**28.4 ± 0.4**	**41.3 ± 0.3**	33.0 ± 0.8
	BRIDGE+SDCLR (ours)	42.0 ± 1.5	16.9 ± 0.9	1.3 ± 0.1	2.2 ± 0.6	10.6 ± 0.8	8.3 ± 0.8	20.2 ± 0.9	14.5 ± 0.8
PASS-10k	SimCLR	72.3 ± 0.2	45.5 ± 0.5	11.0 ± 0.2	17.5 ± 1.4	33.3 ± 1.3	33.3 ± 0.9	44.7 ± 0.9	36.8 ± 0.8
	TS	71.5 ± 0.2	43.5 ± 0.4	10.2 ± 1.2	17.2 ± 0.2	33.1 ± 2.1	36.4 ± 1.8	49.4 ± 0.4	37.3 ± 0.9
	SDCLR	42.2 ± 1.1	17.7 ± 0.8	1.8 ± 0.3	2.2 ± 0.8	10.6 ± 1.3	8.2 ± 1.6	20.8 ± 0.4	14.8 ± 0.9
	BRIDGE (ours)	**73.9 ± 0.3**	**47.1 ± 0.4**	**13.6 ± 0.7**	**20.5 ± 0.7**	**37.7 ± 4.0**	**37.1 ± 0.7**	49.1 ± 0.5	**39.9 ± 1.0**
	BRIDGE+TS (ours)	68.9 ± 0.6	40.6 ± 0.7	12.1 ± 0.4	14.1 ± 0.5	34.6 ± 2.9	31.6 ± 1.5	**52.4 ± 0.6**	36.4 ± 1.0
	BRIDGE+SDCLR (ours)	44.0 ± 0.9	18.8 ± 0.9	1.4 ± 0.3	1.8 ± 0.9	12.6 ± 2.6	12.4 ± 1.3	21.2 ± 1.1	16.0 ± 1.1
DiffusionDB-10k	SimCLR	74.4 ± 0.5	48.6 ± 0.2	10.9 ± 0.7	17.6 ± 0.8	31.2 ± 2.4	36.1 ± 1.3	45.3 ± 0.5	37.7 ± 0.9
	TS	72.3 ± 0.3	44.4 ± 0.3	10.7 ± 0.6	15.4 ± 0.6	33.1 ± 1.1	**39.9 ± 2.7**	48.8 ± 0.2	37.8 ± 0.8
	SDCLR	42.4 ± 0.2	18.4 ± 0.2	1.1 ± 0.2	2.4 ± 0.1	13.0 ± 3.7	8.2 ± 1.4	21.0 ± 0.6	15.2 ± 0.9
	BRIDGE (ours)	**75.9 ± 0.4**	**49.9 ± 0.5**	**12.8 ± 1.0**	**20.0 ± 0.4**	**35.7 ± 1.1**	37.7 ± 0.8	49.1 ± 0.5	**40.1 ± 0.7**
	BRIDGE+TS (ours)	67.8 ± 0.6	39.1 ± 0.5	9.4 ± 1.2	12.4 ± 0.4	30.3 ± 3.0	33.7 ± 1.9	**50.8 ± 0.6**	34.8 ± 1.2
	BRIDGE+SDCLR (ours)	43.8 ± 0.9	18.5 ± 0.3	1.3 ± 0.4	2.5 ± 0.0	14.3 ± 1.6	11.4 ± 0.6	21.9 ± 0.6	16.2 ± 0.6

Table 11: Full kNN transfer results across all source datasets and comparison methods.

Source	Method	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ImageNet-100-LT	SimCLR	68.6 ± 0.4	37.3 ± 0.2	6.0 ± 0.7	11.2 ± 0.4	32.2 ± 0.7	23.6 ± 0.3	43.9 ± 0.9	31.8 ± 0.5
	TS	68.4 ± 0.3	36.5 ± 0.1	4.7 ± 0.2	8.3 ± 0.5	26.6 ± 1.0	23.6 ± 1.7	**52.2 ± 0.5**	31.5 ± 0.6
	SDCLR	32.5 ± 0.4	10.9 ± 0.2	1.0 ± 0.3	2.1 ± 0.1	10.7 ± 0.7	4.7 ± 0.2	11.2 ± 0.2	10.4 ± 0.3
	BRIDGE (ours)	**69.0 ± 0.5**	**37.8 ± 0.3**	**7.0 ± 0.2**	**11.7 ± 0.2**	**36.5 ± 0.4**	**25.3 ± 1.0**	48.1 ± 0.6	**33.6 ± 0.4**
	BRIDGE+TS (ours)	65.4 ± 0.2	33.6 ± 0.3	3.1 ± 0.1	7.1 ± 0.3	18.2 ± 0.2	15.6 ± 0.7	40.8 ± 0.5	26.3 ± 0.3
	BRIDGE+SDCLR (ours)	36.3 ± 0.8	13.0 ± 0.5	1.1 ± 0.3	2.4 ± 0.2	9.7 ± 0.3	5.4 ± 0.4	13.9 ± 1.0	11.7 ± 0.5
CIFAR-10-LT	SimCLR	66.9 ± 0.4	33.4 ± 0.3	3.7 ± 0.4	9.8 ± 0.7	16.5 ± 0.8	13.9 ± 0.8	24.0 ± 0.8	24.0 ± 0.6
	TS	66.3 ± 0.6	27.8 ± 0.4	2.4 ± 0.6	6.4 ± 0.2	10.5 ± 0.7	11.1 ± 0.4	22.0 ± 0.9	20.9 ± 0.5
	SDCLR	30.1 ± 0.3	10.0 ± 0.3	1.7 ± 0.3	2.2 ± 0.2	10.1 ± 0.7	4.4 ± 0.4	9.8 ± 0.3	9.7 ± 0.4
	BRIDGE (ours)	74.8 ± 0.4	**41.0 ± 0.4**	**4.8 ± 0.5**	**10.0 ± 0.2**	**24.2 ± 0.6**	**16.2 ± 0.7**	**29.7 ± 0.3**	**28.7 ± 0.4**
	BRIDGE+TS (ours)	**77.6 ± 0.7**	35.3 ± 0.5	3.1 ± 0.3	6.6 ± 0.4	16.3 ± 0.6	12.5 ± 1.0	26.7 ± 0.1	25.5 ± 0.5
	BRIDGE+SDCLR (ours)	34.4 ± 1.2	11.4 ± 0.6	1.2 ± 0.2	2.2 ± 0.2	8.8 ± 1.5	5.0 ± 0.2	11.7 ± 1.1	10.7 ± 0.7
CIFAR-100-LT	SimCLR	59.3 ± 0.8	30.1 ± 0.9	3.0 ± 0.3	7.4 ± 0.4	13.7 ± 0.2	10.8 ± 0.4	21.8 ± 1.1	20.9 ± 0.6
	TS	53.1 ± 0.6	24.6 ± 0.4	2.4 ± 0.4	5.5 ± 0.3	9.3 ± 0.4	8.3 ± 0.2	18.7 ± 1.0	17.4 ± 0.5
	SDCLR	25.7 ± 1.2	8.5 ± 0.6	0.9 ± 0.4	1.9 ± 0.2	7.3 ± 0.3	3.9 ± 0.3	8.3 ± 1.1	8.1 ± 0.6
	BRIDGE (ours)	**69.3 ± 0.5**	**39.5 ± 0.1**	**4.8 ± 0.6**	**8.9 ± 0.1**	**21.9 ± 0.1**	**15.5 ± 0.1**	**29.1 ± 0.7**	**27.0 ± 0.3**
	BRIDGE+TS (ours)	64.3 ± 0.4	36.6 ± 0.4	2.9 ± 0.3	6.3 ± 0.1	16.4 ± 0.5	12.3 ± 0.0	28.2 ± 0.6	23.9 ± 0.3
	BRIDGE+SDCLR (ours)	31.9 ± 0.4	10.4 ± 0.5	0.8 ± 0.1	2.1 ± 0.5	7.5 ± 1.6	4.5 ± 0.1	9.1 ± 0.7	9.5 ± 0.6
PASS-10k	SimCLR	67.2 ± 0.2	**37.0 ± 0.3**	5.4 ± 0.3	10.3 ± 0.5	30.4 ± 0.0	16.7 ± 0.5	38.3 ± 0.3	29.3 ± 0.3
	TS	65.2 ± 0.7	34.6 ± 0.3	4.4 ± 0.2	7.9 ± 0.2	24.5 ± 0.1	14.4 ± 0.4	39.3 ± 0.4	27.2 ± 0.3
	SDCLR	32.4 ± 1.5	11.5 ± 0.8	1.3 ± 0.2	2.0 ± 0.1	8.6 ± 0.7	5.2 ± 0.2	11.1 ± 0.8	10.3 ± 0.6
	BRIDGE (ours)	**67.5 ± 0.3**	36.9 ± 0.3	**5.8 ± 0.5**	**10.7 ± 0.4**	**33.7 ± 0.2**	**17.8 ± 0.6**	**41.4 ± 0.8**	**30.6 ± 0.4**
	BRIDGE+TS (ours)	58.8 ± 0.5	27.9 ± 0.5	3.0 ± 0.2	6.6 ± 0.1	23.0 ± 0.3	13.4 ± 0.4	38.8 ± 0.8	24.5 ± 0.4
	BRIDGE+SDCLR (ours)	37.1 ± 2.0	13.6 ± 1.0	1.6 ± 0.2	2.4 ± 0.2	10.1 ± 0.5	6.3 ± 0.3	13.1 ± 0.6	12.0 ± 0.7
DiffusionDB-10k	SimCLR	67.3 ± 0.1	36.2 ± 0.4	4.3 ± 0.2	8.8 ± 0.1	26.8 ± 0.4	17.3 ± 0.1	36.7 ± 0.8	28.2 ± 0.3
	TS	64.7 ± 0.4	33.6 ± 0.3	3.1 ± 0.5	7.3 ± 0.3	23.2 ± 0.4	17.0 ± 0.0	37.8 ± 0.6	26.7 ± 0.4
	SDCLR	32.6 ± 1.2	10.0 ± 0.3	1.3 ± 0.1	2.2 ± 0.2	6.1 ± 0.7	4.8 ± 0.4	10.6 ± 0.5	9.7 ± 0.5
	BRIDGE (ours)	**68.3 ± 0.2**	**37.4 ± 0.4**	**5.4 ± 0.3**	**9.8 ± 0.1**	**30.4 ± 0.2**	**18.6 ± 0.2**	**39.4 ± 1.0**	**29.9 ± 0.3**
	BRIDGE+TS (ours)	57.5 ± 0.8	26.6 ± 0.3	3.0 ± 0.4	6.5 ± 0.1	21.7 ± 0.8	15.5 ± 0.4	37.0 ± 0.4	24.0 ± 0.5
	BRIDGE+SDCLR (ours)	35.2 ± 0.7	11.1 ± 0.3	1.0 ± 0.1	2.2 ± 0.1	8.1 ± 1.0	5.0 ± 0.3	11.3 ± 0.7	10.6 ± 0.5

Table 12: Augmentation mechanism ablation with full linear-probe results. The oracle row restores the withheld real images.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Stable Diffusion 3 (default)	**75.0 ± 0.3**	**48.5 ± 0.2**	**16.3 ± 0.8**	**20.7 ± 0.4**	35.7 ± 0.9	**41.8 ± 0.8**	**52.4 ± 0.7**	**41.5 ± 0.6**
FLUX	70.7 ± 0.4	42.1 ± 1.4	10.3 ± 1.3	17.5 ± 0.9	29.4 ± 4.6	31.5 ± 2.9	43.7 ± 1.4	35.0 ± 1.9
Oracle (withheld real images restored)	74.8 ± 0.4	47.8 ± 0.7	13.1 ± 0.1	20.5 ± 0.5	**40.3 ± 6.1**	38.1 ± 1.5	50.3 ± 0.4	40.7 ± 1.4

Table 13: Augmentation mechanism ablation with full kNN results. The oracle row restores the withheld real images.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Stable Diffusion 3 (default)	68.6 ± 0.1	37.2 ± 0.2	6.4 ± 0.6	**11.5 ± 0.2**	**35.2 ± 0.3**	**24.3 ± 0.6**	**46.8 ± 0.7**	**32.9 ± 0.4**
FLUX	65.8 ± 1.0	35.4 ± 0.8	5.7 ± 0.7	9.4 ± 0.9	27.7 ± 1.0	18.8 ± 1.3	38.7 ± 1.2	28.8 ± 1.0
Oracle (withheld real images restored)	**68.8 ± 0.2**	**37.7 ± 0.3**	**6.6 ± 0.5**	11.4 ± 0.5	32.7 ± 0.1	23.6 ± 0.6	44.4 ± 0.4	32.2 ± 0.4

Table 14: Selection rule ablation with full linear-probe results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Mode-window q1-99 + diversity	75.6 ± 0.1	49.2 ± 0.2	**16.9 ± 0.7**	21.9 ± 0.4	34.6 ± 2.9	43.5 ± 0.7	52.8 ± 0.1	42.1 ± 0.7
Mode-window q50-99 + diversity	75.5 ± 0.1	**49.3 ± 0.4**	15.2 ± 0.3	**22.4 ± 0.6**	35.5 ± 1.1	**43.8 ± 3.1**	53.4 ± 0.9	42.2 ± 0.9
Mode-window q75-99 + diversity (default)	**76.0 ± 0.3**	49.1 ± 0.2	15.2 ± 1.2	22.4 ± 0.4	**36.8 ± 1.9**	42.9 ± 1.0	**53.8 ± 0.1**	**42.3 ± 0.7**
Top-tail	70.8 ± 0.5	43.1 ± 1.0	10.0 ± 1.7	15.9 ± 1.3	27.7 ± 1.0	30.5 ± 1.2	42.2 ± 1.4	34.3 ± 1.2
Uniform	71.5 ± 0.2	43.8 ± 0.7	9.3 ± 0.9	17.5 ± 1.2	32.9 ± 0.4	31.7 ± 3.9	42.0 ± 0.9	35.5 ± 1.2

Table 15: Selection rule ablation with full kNN results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
Mode-window q1-99 + diversity	**69.2 ± 0.3**	37.7 ± 0.2	6.6 ± 0.2	11.4 ± 0.2	35.9 ± 0.2	25.0 ± 0.5	48.3 ± 0.3	33.4 ± 0.3
Mode-window q50-99 + diversity	68.7 ± 0.1	**38.0 ± 0.2**	**7.3 ± 0.2**	11.3 ± 0.2	**36.5 ± 0.1**	24.8 ± 0.8	**48.5 ± 0.4**	33.6 ± 0.3
Mode-window q75-99 + diversity (default)	69.0 ± 0.5	37.8 ± 0.3	6.7 ± 0.3	**11.7 ± 0.2**	36.5 ± 0.4	**25.3 ± 1.0**	48.1 ± 0.6	**33.6 ± 0.5**
Top-tail	65.7 ± 0.8	34.4 ± 0.2	4.1 ± 0.4	9.6 ± 0.2	23.9 ± 0.9	18.1 ± 0.8	35.4 ± 1.2	27.3 ± 0.6
Uniform	66.1 ± 1.0	35.7 ± 0.3	4.9 ± 0.1	10.2 ± 0.7	25.3 ± 0.4	19.0 ± 0.8	37.4 ± 0.6	28.4 ± 0.6

Table 16: Cycle count ablation with full linear-probe results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
2 cycles	68.9 ± 0.6	41.6 ± 1.3	10.6 ± 1.2	16.4 ± 0.4	30.7 ± 4.5	33.2 ± 0.9	43.3 ± 0.6	35.0 ± 1.4
5 cycles (default)	**76.0 ± 0.3**	**49.1 ± 0.2**	**15.2 ± 1.2**	**22.4 ± 0.4**	**36.8 ± 1.9**	**42.9 ± 1.0**	**53.8 ± 0.1**	**42.3 ± 0.7**
10 cycles	70.5 ± 1.2	43.8 ± 0.6	9.2 ± 0.9	17.2 ± 1.6	24.8 ± 5.2	31.3 ± 1.7	41.9 ± 0.6	34.1 ± 1.7

Table 20: Linear-probe accuracy of the paired ViT-S runs of Table 4 on every task, mean ± standard deviation over three seeds. MAE runs five cycles of 80 epochs, the other objectives five of 100.

Method	Cars	Aircraft	Flowers	Pets	CIFAR-10	CIFAR-100	IN100-LT
SimCLR, SSL baseline	7.15 ± 0.19	10.92 ± 0.40	32.90 ± 2.72	25.00 ± 1.25	60.46 ± 0.46	34.96 ± 0.20	41.95 ± 1.43
SimCLR, with BRIDGE	8.18 ± 0.42	11.82 ± 0.43	36.60 ± 2.85	29.80 ± 2.98	60.85 ± 0.35	36.33 ± 0.14	43.93 ± 1.09
MoCo v3, SSL baseline	7.57 ± 0.23	11.05 ± 1.31	30.28 ± 1.36	29.26 ± 3.13	59.96 ± 0.39	34.47 ± 0.01	44.79 ± 1.17
MoCo v3, with BRIDGE	9.55 ± 0.34	13.42 ± 0.98	36.82 ± 6.07	31.25 ± 4.00	62.36 ± 0.46	36.30 ± 0.13	47.55 ± 0.86
DINO, SSL baseline	8.10 ± 0.25	11.40 ± 1.03	38.13 ± 5.24	27.54 ± 2.34	58.93 ± 0.94	33.56 ± 1.75	44.63 ± 1.47
DINO, with BRIDGE	9.49 ± 0.88	13.28 ± 0.58	37.25 ± 2.61	34.60 ± 3.62	61.30 ± 0.07	35.23 ± 0.31	46.78 ± 0.77
MAE, SSL baseline	1.54 ± 0.19	2.44 ± 1.00	10.46 ± 5.88	10.05 ± 3.03	37.69 ± 3.73	13.20 ± 1.94	19.93 ± 2.13
MAE, with BRIDGE	1.56 ± 0.43	3.18 ± 0.70	8.50 ± 2.26	8.61 ± 2.20	41.46 ± 2.20	15.00 ± 2.40	22.19 ± 2.35

Table 21: kNN accuracy of the paired ViT-S runs of Table 4 on every task, mean ± standard deviation over three seeds. MAE runs five cycles of 80 epochs, the other objectives five of 100.

Method	Cars	Aircraft	Flowers	Pets	CIFAR-10	CIFAR-100	IN100-LT
SimCLR, SSL baseline	6.26 ± 0.17	7.09 ± 0.17	36.53 ± 0.89	19.16 ± 0.64	66.05 ± 0.48	36.72 ± 0.36	41.21 ± 0.72
SimCLR, with BRIDGE	6.52 ± 0.22	7.56 ± 0.23	38.29 ± 0.83	19.93 ± 0.53	65.70 ± 0.85	36.46 ± 0.33	42.91 ± 0.78
MoCo v3, SSL baseline	6.65 ± 0.38	7.71 ± 0.16	38.91 ± 1.03	23.56 ± 1.12	69.29 ± 0.25	39.98 ± 0.44	45.14 ± 0.46
MoCo v3, with BRIDGE	6.74 ± 0.26	8.21 ± 0.18	40.51 ± 1.45	24.91 ± 0.45	65.41 ± 0.21	36.24 ± 0.46	47.71 ± 1.53
DINO, SSL baseline	6.05 ± 0.27	7.73 ± 0.52	38.52 ± 1.19	19.26 ± 0.16	62.69 ± 2.42	33.98 ± 2.80	45.28 ± 0.98
DINO, with BRIDGE	6.39 ± 0.22	8.16 ± 0.26	40.72 ± 0.31	21.79 ± 1.15	64.38 ± 3.51	35.59 ± 2.51	47.14 ± 0.29
MAE, SSL baseline	1.00 ± 0.14	2.24 ± 0.29	5.30 ± 0.52	4.58 ± 0.17	31.34 ± 1.65	10.96 ± 0.44	11.75 ± 0.38
MAE, with BRIDGE	1.38 ± 0.37	2.50 ± 0.70	7.54 ± 2.13	5.35 ± 0.76	35.58 ± 7.14	13.44 ± 3.46	15.46 ± 2.68

Table 17: Cycle count ablation with full kNN results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
2 cycles	65.3 ± 0.4	34.2 ± 0.5	5.0 ± 0.7	10.0 ± 0.2	26.1 ± 0.8	18.7 ± 0.6	35.9 ± 0.3	27.9 ± 0.5
5 cycles (default)	$\mathbf{69.0 \pm 0.5}$	$\mathbf{37.8 \pm 0.3}$	$\mathbf{6.7 \pm 0.3}$	$\mathbf{11.7 \pm 0.2}$	$\mathbf{36.5 \pm 0.4}$	$\mathbf{25.3 \pm 1.0}$	$\mathbf{48.1 \pm 0.6}$	$\mathbf{33.6 \pm 0.5}$
10 cycles	66.4 ± 0.5	35.4 ± 0.3	4.8 ± 0.3	10.4 ± 0.6	25.2 ± 0.8	18.6 ± 0.9	36.0 ± 0.7	28.1 ± 0.6

Table 18: Backbone architecture ablation with full linear-probe results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ResNet-18	68.8 ± 0.1	41.7 ± 0.2	13.5 ± 0.3	19.7 ± 0.6	38.7 ± 3.4	38.4 ± 2.6	50.2 ± 0.2	38.7 ± 1.1
ResNet-50 (default)	$\mathbf{76.0 \pm 0.3}$	$\mathbf{49.1 \pm 0.2}$	$\mathbf{15.2 \pm 1.2}$	$\mathbf{22.4 \pm 0.4}$	36.8 ± 1.9	$\mathbf{42.9 \pm 1.0}$	$\mathbf{53.8 \pm 0.1}$	$\mathbf{42.3 \pm 0.7}$
ViT-S	62.7 ± 0.4	37.6 ± 0.6	8.8 ± 0.5	12.2 ± 0.2	35.1 ± 4.2	30.2 ± 1.6	42.5 ± 1.2	32.7 ± 1.3
ViT-B	66.0 ± 2.3	40.6 ± 2.3	8.3 ± 0.5	14.3 ± 0.8	$\mathbf{38.8 \pm 2.6}$	31.3 ± 2.1	44.5 ± 1.7	34.8 ± 1.8

Table 19: Backbone architecture ablation with full kNN results.

Setting	C10	C100	Cars	Aircraft	Flowers	Pets	IN100-LT	Avg.
ResNet-18	66.3 ± 0.2	35.7 ± 0.3	5.9 ± 0.2	11.1 ± 0.4	33.5 ± 0.1	23.3 ± 0.5	44.5 ± 0.7	31.5 ± 0.4
ResNet-50 (default)	69.0 ± 0.5	37.8 ± 0.3	$\mathbf{6.7 \pm 0.3}$	$\mathbf{11.7 \pm 0.2}$	36.5 ± 0.4	$\mathbf{25.3 \pm 1.0}$	$\mathbf{48.1 \pm 0.6}$	$\mathbf{33.6 \pm 0.5}$
ViT-S	67.3 ± 0.4	37.5 ± 0.3	5.3 ± 0.2	7.6 ± 0.1	36.2 ± 0.7	20.1 ± 1.0	42.1 ± 1.0	30.9 ± 0.5
ViT-B	$\mathbf{69.3 \pm 0.8}$	$\mathbf{40.1 \pm 0.8}$	6.0 ± 0.2	8.3 ± 0.2	$\mathbf{39.8 \pm 0.8}$	21.9 ± 1.4	45.1 ± 0.7	32.9 ± 0.7

I ALGORITHM

Algorithm 1 BRIDGE training loop with mode-window selection

Initialize source set $D^{(0)}$, encoder $f_{\theta^{(0)}}$, frozen generator g_ϕ, cycle lengths $\{E_c\}_{c=0}^{C-1}$
for $c = 0, \ldots, C-1$ **do**
  Continue training $f_{\theta^{(c)}}$ on $D^{(c)}$ for E_c epochs with a fresh optimizer schedule
  if $c < C-1$ **then**
    Compute normalized embeddings $Z^{(c)} = \{f_{\theta^{(c)}}(x_i) \mid x_i \in D^{(c)}\}$
    Compute mean kNN scores $\{d_i^{(c)}\}$ from $Z^{(c)}$
    Restrict to the 75th to 99th percentile band and locate the modal bin center $\bar{d}^{(c)}$
    Form the candidate pool $P^{(c)}$ of the αB images with scores closest to $\bar{d}^{(c)}$
    Select $S^{(c)} \subseteq P^{(c)}$ with $|S^{(c)}| = B$ by greedy farthest-point sampling in embedding space
    Form $\widetilde{D}^{(c)} = \{g_\phi(x_i, \epsilon_j) \mid i \in S^{(c)},\ j \in [G]\}$ and set $D^{(c+1)} = D^{(c)} \cup \widetilde{D}^{(c)}$
  end if
end for
Return the final encoder $f_{\theta^{(C-1)}}$