A. Supplementary Materials for PALM

A.1. Related Work

Active learning (AL) has been widely explored as a means to reduce annotation costs by querying the most informative samples. AL strategies are commonly categorized by their core selection criteria, including uncertainty, diversity, representation, and hybrid approaches. Uncertainty-based methods prioritize samples where the model exhibits the lowest confidence. A classical approach is uncertainty sampling [21], which selects data points with low predicted class confidence. Margin sampling [30] targets instances where the difference between the top two predicted class probabilities is small, indicating ambiguity. Entropy-based sampling [32] captures total predictive uncertainty by selecting samples with high entropy in the output distribution.

Diversity-based methods seek to avoid redundancy by selecting a set of samples that spans the data distribution. These techniques often rely on geometric or statistical distance metrics. For example, the k-Center Greedy algorithm [31] minimizes the maximum distance between selected points and the remaining pool. BADGE [2] combines uncertainty and diversity by clustering in gradient space using a k-means++ scheme. Diversity plays a crucial role in early-stage selection to ensure broader coverage of the input space. Moreover, representation-based methods utilize structure in the feature space to guide sampling. These often rely on clustering or geometric criteria to identify representative or central points. Examples include k-means-based sampling [37], medoid selection [1], and median-based heuristics [34]. TypiClust [14] extends this idea by combining sample typicality and cluster centrality, favoring samples that are both generalizable and diverse.

Hybrid methods integrate multiple selection criteria, often combining uncertainty with diversity. DBAL [10] leverages Bayesian dropout to estimate uncertainty, while promoting diversity among queried samples. These methods are particularly valuable in deep neural networks, where relying solely on uncertainty can lead to redundant or misleading selections. To facilitate the empirical comparison of AL strategies, LabelBench [40] was proposed as a modular and extensible benchmarking suite. It enables evaluation of AL, semi-supervised learning, and transfer learning under consistent conditions, including different model architectures and labeling budgets. LabelBench places strong emphasis on reproducibility and explores the synergy between AL and pretrained models, particularly vision transformers. Its findings suggest that combining AL with SSL can yield notable improvements in label efficiency.

While LabelBench provides a valuable empirical benchmarking framework, our work contributes a complementary modeling perspective. PALM introduces a predictive and interpretable parametric model for characterizing AL behavior. By estimating three key descriptors, i.e., initial performance, growth rate, and asymptotic accuracy, PALM enables quantitative comparison of AL methods and forecasting of future performance based on partial observations. Although both approaches evaluate common AL strategies, LabelBench emphasizes empirical performance across tasks and architectures, while PALM focuses on modeling and interpretability of AL dynamics.

A.2. PALM Proofs and Corollary

In this section, we provide the proofs and a corollary corresponding to the methods described in the main text.

A.2.1. Definition 1: Coverage Probability

Proof. By the complement rule in probability theory, the probability of an event occurring is equal to one minus the probability of its complement. Let A represent the event that a point x is covered by at least one object, and let A^c represent the complement event, where x is not covered by any object. According to the complement rule, we have:

$$P(A) + P(A^c) = 1. \tag{17}$$

Substituting $P(A) = P_{\text{C}}$ and $P(A^c) = P_{\text{NC}}$, the equation becomes:

$$P_{\text{C}} + P_{\text{NC}} = 1, \tag{18}$$

which completes the proof.

A.2.2. Definition 2: Coverage Probability with s Independent Objects

Proof. Let p represent the probability that a single randomly placed object covers point x. The probability that a single object does not cover x is $1 - p$. Now, consider s objects placed independently in the space $\mathbb{X}$. Since the objects are independent, the probability that none of them covers x is the product of their individual non-coverage probabilities:

$$P_{\text{UC}} = (1 - p)^s. \tag{19}$$

Thus, the probability that x is covered by at least one object is:

$$P_{\text{C}} = 1 - (1 - p)^s. \tag{20}$$

This completes the proof.

A.2.3. Corollary 2: Asymptotic Behavior of Accuracy as a Function of Coverage Probability

The test generalization accuracy function is given by:

$$\text{A} = \text{A}_C \left(1 - (1-\delta)^B\right) + \text{A}_{UC} \, (1-\delta)^B. \tag{21}$$

This function exhibits the following asymptotic behaviors:

Case 1. No Labeled Samples ($B = 0$): When no labeled samples are available, the coverage fraction is:

$$P_{\text{C}} = 1 - (1 - \delta)^0 = 0. \tag{22}$$

Substituting this into the accuracy function gives $\mathrm{A} = \mathrm{A}_{\mathrm{UC}}$, indicating that without labeled data, the model's accuracy depends on its performance in the uncovered regions.

Case 2. Infinite Labeled Samples ($B \to \infty$): As the number of labeled samples increases, the coverage probability approaches one:

$$\lim_{B\to\infty}(1-\delta)^B = 0. \quad (23)$$

Substituting this into the accuracy function yields:

$$\lim_{B\to\infty}\mathrm{A} = \mathrm{A}_{\mathrm{C}}, \quad (24)$$

which implies that with full coverage, the model achieves its maximum accuracy in the covered regions.

Case 3. Small δ Approximation for Large B: For small values of δ, the coverage term $(1-\delta)^B$ can be approximated using the first-order Taylor expansion:

$$(1-\delta)^B \approx e^{-B\delta}, \quad (25)$$

since $\lim_{x\to 0^+}(1-x) \approx e^{-x}$. Thus, for sufficiently large B, the accuracy function approximates:

$$\mathrm{A} \approx \mathrm{A}_{\mathrm{C}}\left(1-e^{-B\delta}\right) + \mathrm{A}_{\mathrm{UC}}\, e^{-B\delta}, \quad (26)$$

which shows that the accuracy converges exponentially towards A_{C}, with the rate of convergence governed by δ.

A.2.4. Definition 6: Generalized Accuracy as a Function of Coverage with Exponential Adjustment

Proof. We aim to derive the generalized accuracy function, which incorporates the parameters α and β. The function is defined as:

$$\mathrm{A} = \mathrm{A}_{\max}\left(1-(1-\delta)^{(B+\alpha)^\beta}\right), \quad (27)$$

where $\mathrm{A}_{\max}$ is the maximum achievable accuracy under full coverage, B is the cumulative number of labeled samples, δ represents the expected fraction of the space covered by a single labeled sample, α accounts for initial learning effects and prior knowledge, allowing non-zero accuracy when $B = 0$, and β controls the scaling of accuracy growth as B increases.

We start with the test accuracy function given by:

$$\mathrm{A} = \mathrm{A}_{\mathrm{C}}\left(1-(1-\delta)^B\right) + \mathrm{A}_{\mathrm{UC}}\,(1-\delta)^B. \quad (28)$$

Rearranging the terms:

$$\mathrm{A} = \mathrm{A}_{\mathrm{C}} - (\mathrm{A}_{\mathrm{C}} - \mathrm{A}_{\mathrm{UC}})(1-\delta)^B. \quad (29)$$

Assuming that $\mathrm{A}_{\max} = \mathrm{A}_{\mathrm{C}}$, we rewrite the expression as:

$$\mathrm{A} = \mathrm{A}_{\max} - (\mathrm{A}_{\max} - \mathrm{A}_{\mathrm{UC}})(1-\delta)^B. \quad (30)$$

As B increases, the second term vanishes, since $(1-\delta)^B \to 0$, which ensures that $\mathrm{A} \to \mathrm{A}_{\max}$, as expected under full coverage. To generalize this formulation and account for variations in early learning dynamics and growth rates, we replace B with the adjusted term $(B+\alpha)^\beta$, where $\alpha > 0$ allows the model to exhibit non-zero accuracy even when $B = 0$, representing prior knowledge or inherent generalization, and $\beta > 0$ modulates the rate of accuracy increase with B. Substituting this adjustment, the generalized accuracy function becomes:

$$\mathrm{A} = \mathrm{A}_{\max} - (\mathrm{A}_{\max} - \mathrm{A}_{\mathrm{UC}})(1-\delta)^{(B+\alpha)^\beta}. \quad (31)$$

Finally, assuming that the uncovered regions contribute negligible accuracy ($\mathrm{A}_{\mathrm{UC}} \approx 0$), the expression simplifies to:

$$\mathrm{A} = \mathrm{A}_{\max}\left(1-(1-\delta)^{(B+\alpha)^\beta}\right). \quad (32)$$

Thus, the generalized accuracy function models the influence of labeled sample coverage, prior knowledge, and learning dynamics on active learning performance.

A.2.5. Corollary 3: Asymptotic Behavior of Generalized Accuracy with Exponential Adjustment

The generalized test accuracy function exhibits the following asymptotic behavior:

Case 1. No Labeled Samples ($B = 0$): When no labeled samples are available, the accuracy simplifies to:

$$\mathrm{A} = \mathrm{A}_{\max}\left(1-(1-\delta)^{\alpha^\beta}\right), \quad (33)$$

which leads to two characteristic scenarios:

- If $\alpha > 0$, then $\mathrm{A} > 0$, indicating that the model achieves non-zero accuracy even without labeled data. This reflects the model's ability to generalize from uncovered regions or prior knowledge.
- If $\alpha = 0$, we recover the classical case where no coverage implies zero accuracy, i.e., $\mathrm{A} = 0$.

Case 2. Infinite Labeled Samples ($B \to \infty$): As the number of labeled samples grows to infinity:

$$\lim_{B\to\infty}(B+\alpha)^\beta = \infty. \quad (34)$$

Since $0 < 1-\delta < 1$, we have:

$$\lim_{B\to\infty}(1-\delta)^{(B+\alpha)^\beta} = 0. \quad (35)$$

Therefore, the accuracy converges to its theoretical maximum:

$$\lim_{B\to\infty}\mathrm{A} = \mathrm{A}_{\max}. \quad (36)$$

This confirms that, with full coverage, the model achieves optimal performance.

Case 3. Small δ Approximation for Large B: For small δ, the coverage term can be approximated using the first-order Taylor expansion:

$$(1-\delta)^{(B+\alpha)^\beta} \approx e^{-(B+\alpha)^\beta \delta}. \tag{37}$$

Thus, for sufficiently large B, the accuracy function approximates:

$$\mathrm{A} \approx \mathrm{A}_{\max}\left(1 - e^{-(B+\alpha)^\beta \delta}\right), \tag{38}$$

which shows that the accuracy converges exponentially toward $\mathrm{A}_{\max}$, with a rate of convergence determined by α, β, and δ.

A.2.6. Lemma 1: Parameter Estimation for Learning Dynamics in Active Learning Without Normalization

Let B denote the total number of labeled samples collected during an active learning (AL) process. Given the observed accuracy values from the AL process, we aim to estimate the parameters $\mathrm{A}_{\max}$, δ, α, and β in the model:

$$\mathrm{A} = \mathrm{A}_{\max}\left(1 - (1-\delta)^{(B+\alpha)^\beta}\right). \tag{39}$$

These parameters can be empirically estimated from accuracy measurements collected over multiple AL iterations without the need to normalize B.

Suppose accuracy is observed for at least four different cumulative budgets B_1, B_2, B_3, B_4, with corresponding accuracies A_1, A_2, A_3, and A_4. The parameters $\mathrm{A}_{\max}$, δ, α, and β can then be estimated by solving the following system of equations using nonlinear regression techniques:

$$\mathrm{A}_i = \mathrm{A}_{\max}\left(1 - (1-\delta)^{(B_i+\alpha)^\beta}\right), \quad i = 1,2,3,4. \tag{40}$$

The complexity of estimating the parameters primarily depends on evaluating the exponentiation term $(1-\delta)^{(B+\alpha)^\beta}$. Using optimized exponentiation algorithms, the complexity is approximately $\mathcal{O}(\log(B))$. However, in naive implementations, the complexity can approach $\mathcal{O}(B^\beta)$, especially for large B. For large values of B, the following challenges arise:

- Computational Overhead: The term $(B+\alpha)^\beta$ grows rapidly, increasing computation time.
- Numerical Instability: Large exponents may lead to floating-point precision errors.
- Diminishing Accuracy Gains: As B increases, the marginal contribution of additional labeled samples decreases due to saturation effects.

To mitigate these issues, normalizing B by the mean budget per iteration (b) reduces the computational cost from $\mathcal{O}(B^\beta)$ to $\mathcal{O}((B/b)^\beta)$, improves numerical stability during exponentiation, and ensures smoother convergence behavior of the accuracy function. The normalized generalized accuracy function is then given by:

$$\mathrm{A} = \mathrm{A}_{\max}\left(1 - (1-\delta)^{\left(\frac{B}{b}+\alpha\right)^\beta}\right), \tag{41}$$

where normalization aligns the function with the number of AL iterations rather than the absolute number of labeled samples.

A.2.7. Theorem 1: Comparing Two Active Learning Methods Using the Normalized Accuracy Function

Proof. Consider two active learning methods 1 and 2, with normalized accuracy functions defined as:

$$\mathrm{A}_1 = \mathrm{A}_{\max,1}\left(1 - (1-\delta_1)^{\left(\frac{B}{b_1}+\alpha_1\right)^{\beta_1}}\right), \tag{42}$$

$$\mathrm{A}_2 = \mathrm{A}_{\max,2}\left(1 - (1-\delta_2)^{\left(\frac{B}{b_2}+\alpha_2\right)^{\beta_2}}\right). \tag{43}$$

To compare their performance for a given budget B, we define the ratio of accuracies. Method 1 outperforms Method 2 at budget B if $\mathrm{A}_1/\mathrm{A}_2 > 1$.

Full Coverage Limit ($B \to \infty$). As the number of labeled samples approaches infinity:

$$\lim_{B\to\infty} \mathrm{A}_1 = \mathrm{A}_{\max,1}, \quad \lim_{B\to\infty} \mathrm{A}_2 = \mathrm{A}_{\max,2}. \tag{44}$$

Thus, in the limit of infinite budget, the method with the higher $\mathrm{A}_{\max}$ dominates:

$$\mathrm{A}_{\max,1} > \mathrm{A}_{\max,2} \quad \Rightarrow \quad \lim_{B\to\infty} \frac{\mathrm{A}_1}{\mathrm{A}_2} > 1. \tag{45}$$

Early-Stage Learning (small B). For small budget B, applying the first-order Taylor approximation results in:

$$(1-\delta)^x \approx e^{-x\delta}. \tag{46}$$

Therefore, we can approximate the accuracy functions as:

$$\mathrm{A}_1 \approx \mathrm{A}_{\max,1}\left(1 - e^{-\delta_1\left(\frac{B}{b_1}+\alpha_1\right)^{\beta_1}}\right), \tag{47}$$

$$\mathrm{A}_2 \approx \mathrm{A}_{\max,2}\left(1 - e^{-\delta_2\left(\frac{B}{b_2}+\alpha_2\right)^{\beta_2}}\right). \tag{48}$$

In this regime, faster accuracy growth occurs for the method with larger δ, higher α, smaller b, and larger β.

General Comparison Criterion. To compare the accuracy growth rates, differentiate the accuracy functions with respect to B. Method 1 improves faster than Method 2 if:

$$\frac{d\,\mathrm{A}_1}{dB} > \frac{d\,\mathrm{A}_2}{dB}. \tag{49}$$

This condition holds when:

$$\delta_1\left(\frac{1}{b_1} + \frac{\alpha_1}{B}\right)^{\beta_1} > \delta_2\left(\frac{1}{b_2} + \frac{\alpha_2}{B}\right)^{\beta_2}. \tag{50}$$

In summary, Method 1 outperforms Method 2 when it exhibits higher coverage efficiency (δ), smaller batch size (b), greater initial accuracy boost (α), or faster accuracy scaling (β). Additionally, the asymptotic accuracy $\mathrm{A}_{\mathrm{max}}$ determines long-term dominance as B increases. Together, these parameters provide a comprehensive framework for quantitatively comparing active learning strategies across different budget regimes.

A.3. Quantitative Results

Table 1. PALM parameter estimates for CIFAR-10 without pretrained embeddings, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the mean values from 5 repetitions. The table reports the maximum achievable accuracy ($\mathrm{A_{max}}$), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). In the absence of pretrained embeddings, methods show slower learning dynamics and lower δ values, with α increasing over time, indicating delayed accuracy gains. TypiClust demonstrates relatively higher δ values throughout, reflecting strong sample efficiency. In contrast, methods like Margin and Entropy show increasing α and β, indicating slower convergence in later stages.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β
Random	32.6	0.373	2.519	0.688	100	0.168	1.611	0.303	100	0.144	2.466	0.365	100	0.139	2.745	0.376	100	0.124	3.696	0.409	93.8	0.116	4.951	0.446	90.7	0.108	6.263	0.475
Uncertainty	100	0.105	1.714	0.467	36.3	0.138	3.365	0.932	49.0	0.147	2.716	0.701	65.2	0.161	1.624	0.510	91.5	0.135	1.092	0.416	99.6	0.102	2.614	0.456	93.7	0.090	4.644	0.506
Margin	39.1	0.457	1.289	0.390	100	0.167	1.672	0.316	100	0.150	2.227	0.360	100	0.142	2.626	0.377	100	0.127	3.587	0.409	93.6	0.097	7.353	0.497	92.6	0.094	7.812	0.507
Entropy	56.7	0.317	0.355	0.282	100	0.176	0.397	0.264	100	0.165	0.611	0.304	100	0.152	0.935	0.333	100	0.126	2.204	0.389	99.4	0.083	7.303	0.490	93.3	0.070	10.00	0.546
TypiClust	35.4	0.489	1.274	0.617	42.3	0.474	0.858	0.412	52.3	0.410	0.618	0.308	100	0.180	1.726	0.307	100	0.165	2.451	0.333	-	-	-	-	-	-	-	-

Table 2. PALM parameter estimates for CIFAR-10 without pretrained embeddings, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the minimum values from 5 repetitions. The table reports the maximum achievable accuracy ($\mathrm{A_{max}}$), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). In the absence of pretrained embeddings, methods show slower learning dynamics and lower δ values, with α increasing over time, indicating delayed accuracy gains. TypiClust demonstrates relatively higher δ values throughout, reflecting strong sample efficiency. In contrast, methods like Margin and Entropy show increasing α and β, indicating slower convergence in later stages.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β
Random	100	0.144	2.369	0.319	100	0.085	4.819	0.534	100	0.093	4.654	0.493	87.1	0.143	2.824	0.407	100	0.114	3.635	0.418	93.4	0.107	4.930	0.458	90.0	0.098	6.350	0.490
Uncertainty	73.7	0.000	38.91	4.559	57.9	0.000	56.46	4.146	50.7	0.148	1.790	0.658	57.7	0.154	1.355	0.571	81.6	0.138	0.732	0.442	100	0.095	2.140	0.462	93.7	0.083	4.104	0.512
Margin	70.8	0.000	39.14	4.522	100	0.188	0.446	0.246	100	0.164	0.955	0.319	100	0.142	1.733	0.369	100	0.125	2.669	0.406	97.3	0.104	4.834	0.460	92.8	0.093	6.879	0.502
Entropy	72.5	0.000	40.98	4.510	55.6	0.000	53.58	4.138	100	0.146	0.380	0.327	100	0.136	0.599	0.356	100	0.122	1.093	0.389	100	0.077	6.406	0.495	93.2	0.063	10.31	0.559
TypiClust	63.5	0.000	35.52	4.594	100	0.197	0.508	0.262	100	0.203	0.409	0.242	48.2	0.343	1.120	0.425	49.9	0.353	0.878	0.386	-	-	-	-	-	-	-	-

Table 3. PALM parameter estimates for CIFAR-10 without pretrained embeddings, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the maximum values from 5 repetitions. The table reports the maximum achievable accuracy ($\mathrm{A_{max}}$), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). In the absence of pretrained embeddings, methods show slower learning dynamics and lower δ values, with α increasing over time, indicating delayed accuracy gains. TypiClust demonstrates relatively higher δ values throughout, reflecting strong sample efficiency. In contrast, methods like Margin and Entropy show increasing α and β, indicating slower convergence in later stages.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β
Random	29.6	0.001	16.87	5.348	34.8	0.423	2.032	0.639	100	0.180	1.608	0.301	100	0.145	3.133	0.372	100	0.126	4.355	0.412	100	0.126	4.355	0.412	91.6	0.120	5.799	0.454
Uncertainty	100	0.008	11.58	1.265	33.1	0.000	28.16	4.467	40.4	0.022	9.550	1.423	71.9	0.150	2.545	0.495	99.5	0.123	2.053	0.420	94.8	0.101	4.184	0.486	93.8	0.098	4.613	0.496
Margin	100	0.092	6.633	0.520	100	0.000	44.99	2.612	41.2	0.007	16.22	1.686	100	0.139	3.940	0.389	100	0.123	5.139	0.421	91.1	0.088	10.16	0.534	92.6	0.097	8.479	0.506
Entropy	47.9	0.000	81.58	3.796	100	0.155	1.531	0.346	100	0.159	1.406	0.336	80.2	0.200	1.328	0.348	100	0.133	2.932	0.384	91.8	0.067	12.81	0.569	92.7	0.072	11.69	0.553
TypiClust	33.8	0.222	2.706	1.290	36.2	0.456	1.481	0.749	48.3	0.481	0.550	0.324	100	0.165	3.205	0.345	100	0.142	4.792	0.388	-	-	-	-	-	-	-	-

Table 4. PALM parameter estimates for CIFAR-10 using SimCLR embeddings for feature extraction, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the mean values from 5 repetitions. The table reports the maximum achievable accuracy ($\mathrm{A_{max}}$), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). The results highlight the acceleration of learning dynamics with pretrained embeddings, where TypiClust and Margin benefit from high δ and low α, indicating efficient early-stage learning. Conversely, methods like Entropy and DBAL exhibit delayed improvements at small budgets but show recovery and better performance as the annotation grows.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β	$\mathrm{A_{max}}$	δ	α	β
Random	100	0.443	0.536	0.341	100	0.451	0.473	0.317	100	0.394	0.912	0.413	84.2	0.290	2.314	0.735	85.0	0.355	1.690	0.633	86.6	0.496	0.609	0.434	87.1	0.536	0.392	0.381
Uncertainty	100	0.278	0.123	0.147	74.1	0.000	26.35	4.534	75.2	0.000	22.50	4.536	79.8	0.005	8.341	1.940	82.7	0.061	3.390	1.142	86.6	0.220	0.708	0.640	87.6	0.270	0.397	0.546
Margin	68.1	0.000	12.57	5.699	80.1	0.250	1.853	0.966	82.5	0.286	1.596	0.857	84.3	0.357	1.133	0.705	85.5	0.409	0.819	0.605	87.5	0.511	0.326	0.428	87.9	0.535	0.241	0.389
Entropy	34.8	0.000	13.34	5.462	39.1	0.299	1.822	0.856	81.6	0.000	31.60	4.337	76.6	0.000	31.40	4.427	80.2	0.008	8.037	1.643	86.2	0.183	0.581	0.632	87.5	0.227	0.305	0.542
TypiClust	78.1	0.859	0.536	0.347	88.2	0.788	0.188	0.147	96.7	0.721	0.153	0.116	85.5	0.803	0.298	0.188	85.3	0.803	0.312	0.192	-	-	-	-	-	-	-	-
DBAL	33.3	0.000	12.11	5.595	100	0.185	0.862	0.369	79.6	0.000	30.30	4.377	76.1	0.000	29.55	4.413	80.0	0.014	6.508	1.505	86.3	0.197	0.488	0.608	87.5	0.236	0.274	0.528

Table 5. PALM parameter estimates for CIFAR-10 using SimCLR embeddings for feature extraction, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the minimum values from 5 repetitions. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). The results highlight the acceleration of learning dynamics with pretrained embeddings, where TypiClust and Margin benefit from high δ and low α, indicating efficient early-stage learning. Conversely, methods like Entropy and DBAL exhibit delayed improvements at small budgets but show recovery and better performance as the annotation grows.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
Random	60.4	0.000	12.60	5.593	66.7	0.510	0.808	0.588	100	0.258	1.699	0.562	83.5	0.117	4.241	1.009	84.6	0.211	2.595	0.789	86.2	0.357	1.004	0.551	86.8	0.411	0.615	0.476
Uncertainty	57.3	0.000	66.30	3.987	70.1	0.000	19.46	4.726	71.4	0.000	17.86	4.808	77.9	0.010	5.007	1.777	82.0	0.075	1.750	1.069	86.4	0.190	0.379	0.665	87.4	0.226	0.237	0.584
Margin	62.4	0.267	0.894	1.361	94.7	0.362	0.115	0.462	100	0.333	0.137	0.479	85.4	0.348	0.276	0.618	85.5	0.350	0.266	0.612	87.3	0.398	0.119	0.508	87.7	0.416	0.087	0.475
Entropy	72.0	0.000	41.39	4.450	57.4	0.000	57.33	4.091	69.0	0.000	21.11	4.556	73.4	0.000	20.41	3.828	78.6	0.016	3.898	1.433	86.0	0.134	0.188	0.690	87.3	0.161	0.105	0.614
TypiClust	72.7	0.201	4.324	1.445	100	0.647	0.529	0.140	82.3	0.734	1.142	0.291	82.8	0.752	0.892	0.256	84.3	0.760	0.643	0.212	-	-	-	-	-	-	-	-
DBAL	61.5	0.000	61.75	4.078	55.4	0.000	51.67	4.223	70.4	0.000	23.07	4.499	71.9	0.000	25.99	4.604	78.5	0.037	2.457	1.183	86.2	0.149	0.162	0.647	87.4	0.172	0.101	0.588

Table 6. PALM parameter estimates for CIFAR-10 using SimCLR embeddings for feature extraction, evaluated across various AL strategies and different numbers of labeled points used for curve fitting based on the maximum values from 5 repetitions. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). The results highlight the acceleration of learning dynamics with pretrained embeddings, where TypiClust and Margin benefit from high δ and low α, indicating efficient early-stage learning. Conversely, methods like Entropy and DBAL exhibit delayed improvements at small budgets but show recovery and better performance as the annotation grows.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				500 Points				1000 Points			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
Random	100	0.509	0.559	0.321	100	0.518	0.494	0.306	96.5	0.531	0.535	0.327	84.9	0.480	1.317	0.556	85.3	0.511	1.096	0.511	86.9	0.626	0.362	0.342	87.3	0.652	0.238	0.303
Uncertainty	39.5	0.000	6.601	5.378	85.5	0.000	36.72	4.202	80.0	0.000	29.09	4.311	81.3	0.000	23.79	3.425	83.1	0.009	9.458	1.690	86.7	0.246	1.170	0.625	87.7	0.322	0.564	0.503
Margin	100	0.026	5.817	1.720	79.0	0.000	14.39	4.978	81.4	0.117	3.350	1.421	84.3	0.461	1.076	0.665	85.7	0.546	0.641	0.508	87.7	0.634	0.232	0.338	88.1	0.652	0.172	0.308
Entropy	45.9	0.000	14.25	5.363	43.3	0.000	12.78	5.494	88.4	0.000	38.05	4.203	78.6	0.000	32.07	4.160	80.9	0.000	19.07	2.554	86.4	0.268	0.739	0.541	87.7	0.320	0.404	0.456
TypiClust	78.3	0.000	10.77	6.109	78.3	0.000	10.73	6.118	100	0.741	0.017	0.078	93.5	0.793	0.023	0.092	88.2	0.836	0.048	0.124	-	-	-	-	-	-	-	-
DBAL	46.5	0.000	15.57	5.270	41.8	0.000	12.80	5.467	87.1	0.000	36.19	4.217	78.5	0.000	31.29	4.163	81.1	0.000	20.40	2.700	86.3	0.243	0.901	0.576	87.5	0.304	0.454	0.474

Table 7. PALM parameter estimates for CIFAR-100 across different AL methods and varying numbers of labeled points used for estimation. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). Higher values of δ and lower values of α indicate more efficient early-stage learning, while β reflects the scalability of the method as the number of labeled points increases.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				448 Points			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
Random	59.0	0.000	80.98	3.804	52.3	0.000	68.27	3.935	55.5	0.000	85.53	3.707	36.4	0.081	1.376	0.675	79.4	0.050	0.720	0.518	79.8	0.048	0.955	0.526
Uncertainty	55.6	0.000	86.03	3.771	43.7	0.000	85.64	3.781	46.1	0.000	85.11	3.715	93.8	0.035	0.410	0.531	100	0.030	0.626	0.551	69.9	0.029	3.469	0.656
Margin	44.8	0.000	85.36	3.772	56.6	0.000	85.26	3.748	46.6	0.000	86.15	3.728	25.8	0.000	73.40	3.936	100	0.042	1.044	0.482	58.0	0.025	10.64	0.751
Entropy	56.1	0.000	84.65	3.783	40.2	0.000	97.83	3.668	56.1	0.000	83.88	3.746	23.2	0.000	70.11	4.037	100	0.025	0.772	0.568	61.3	0.018	7.519	0.777
TypiClust	52.0	0.000	90.99	3.712	68.1	0.000	67.31	3.953	54.5	0.000	86.69	3.737	-	-	-	-	-	-	-	-	-	-	-	-
DBAL	51.3	0.000	88.04	3.764	52.6	0.000	88.59	3.740	54.7	0.000	84.73	3.736	23.3	0.000	54.62	4.185	47.5	0.031	1.284	0.774	100	0.024	0.452	0.595

Table 8. PALM parameter estimates for CIFAR-100 using SimCLR embeddings, evaluated across different AL strategies and varying numbers of labeled points used for estimation. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β). Higher δ values and lower α indicate more efficient early-stage learning, while β captures the rate of accuracy growth as the number of labeled points increases.

AL Method	6 Points				10 Points				20 Points				50 Points				100 Points				448 Points			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
Random	41.7	0.201	0.718	1.133	47.7	0.205	0.570	0.975	49.9	0.218	0.458	0.887	53.9	0.256	0.227	0.709	55.7	0.278	0.138	0.626	57.6	0.318	0.057	0.516
Uncertainty	51.0	0.000	88.128	3.777	57.3	0.000	83.024	3.805	38.4	0.037	0.615	1.347	46.6	0.059	0.000	1.033	51.3	0.073	0.000	0.881	58.2	0.116	0.001	0.634
Margin	56.0	0.000	65.254	4.041	39.9	0.032	2.265	1.822	46.9	0.104	0.969	1.157	52.6	0.163	0.366	0.840	55.3	0.195	0.175	0.707	58.0	0.237	0.068	0.573
Entropy	59.6	0.000	73.883	3.920	60.1	0.000	77.654	3.865	48.1	0.000	87.174	3.733	42.4	0.030	0.000	1.227	47.3	0.047	0.000	0.991	58.6	0.089	0.001	0.651
TypiClust	45.9	0.433	1.128	0.737	48.3	0.451	0.948	0.633	48.8	0.464	0.860	0.597	-	-	-	-	-	-	-	-	-	-	-	-
DBAL	59.6	0.000	73.29	3.937	60.0	0.000	77.83	3.864	47.4	0.000	87.58	3.726	42.0	0.030	0.000	1.246	47.3	0.048	0.000	0.989	58.6	0.091	0.001	0.646

Table 9. PALM parameter estimates on ImageNet-50 across different AL strategies and self-supervised embeddings. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β) for each method, capturing the efficiency and dynamics of learning across varying annotation budgets.

Embedding	**Random**				**Entropy**				**Uncertainty**				**DBAL**				**Margin**				**TypiClust**			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
MoCov2+	27.5	0.005	4.64	1.86	47.6	0.006	2.27	2.36	1.95	1.000	0.00	-1.00	66.3	0.080	13.25	-0.17	27.1	0.000	17.77	4.49	28.4	0.000	17.63	4.91
MoCov3	87.0	0.397	0.00	0.78	84.5	0.155	0.00	1.20	85.4	0.154	0.41	1.22	84.6	0.103	0.61	1.39	88.2	0.444	0.00	0.78	87.0	0.714	0.00	0.55
BYOL	8.48	0.314	0.00	0.67	42.0	0.071	0.00	0.30	12.0	0.170	0.00	0.61	23.1	0.129	0.00	0.31	8.08	0.250	0.43	0.81	2.96	0.130	6.56	1.05
SimCLR	66.5	0.131	0.22	0.98	45.9	0.001	3.81	2.39	50.0	0.009	1.11	1.96	46.9	0.001	4.42	2.71	62.7	0.138	0.00	1.15	59.5	0.363	1.13	0.67

Table 10. PALM parameter estimates on ImageNet-100 across different AL strategies and self-supervised embeddings. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β) for each method, providing insights into learning efficiency and model behavior across varying annotation budgets.

Embedding	**Random**				**Entropy**				**Uncertainty**				**DBAL**				**Margin**				**TypiClust**			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
MoCov2+	21.2	0.001	6.51	2.26	1.85	0.306	0.00	6.75	100	0.002	20.0	0.39	2.49	0.328	0.00	2.07	21.4	0.055	0.00	1.26	27.60	0.001	20.0	2.26
MoCov3	81.6	0.333	0.00	0.87	76.7	0.082	0.00	1.47	76.6	0.112	0.00	1.37	76.4	0.083	0.00	1.48	81.1	0.316	0.00	1.00	80.7	0.690	0.00	0.56
BYOL	8.17	0.188	0.00	0.47	8.52	0.125	0.34	0.60	22.7	0.070	0.00	0.36	8.64	0.090	1.71	0.72	4.44	0.302	0.00	0.72	2.46	0.000	20.0	4.41
SimCLR	44.0	0.226	0.00	0.73	27.7	0.087	0.00	1.08	36.2	0.105	0.00	0.86	27.3	0.089	0.00	1.09	44.0	0.171	0.30	0.80	36.9	0.090	3.36	1.20

Table 11. PALM parameter estimates on ImageNet-200 across different AL strategies and self-supervised embeddings. The table reports the maximum achievable accuracy (A_{max}), coverage efficiency (δ), early-stage performance offset (α), and scalability (β) for each method, highlighting differences in learning dynamics and sample efficiency across strategies.

Embedding	**Random**				**Entropy**				**Uncertainty**				**DBAL**				**Margin**				**TypiClust**			
	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β	A_{max}	δ	α	β
MoCov2+	18.9	0.001	4.65	2.33	1.12	0.411	0.00	1.05	3.90	0.000	18.9	4.63	1.40	0.000	5.44	5.17	32.0	0.020	1.41	1.62	13.1	0.060	3.57	1.85
MoCov3	76.9	0.330	0.00	0.80	47.4	0.092	0.00	1.38	70.3	0.129	0.00	1.28	69.2	0.087	0.00	1.43	76.0	0.348	0.00	0.83	70.4	0.694	0.00	0.57
BYOL	27.0	0.034	0.01	0.40	14.1	0.068	0.00	0.39	3.29	0.206	0.00	0.71	7.20	0.136	0.00	0.42	7.98	0.095	0.00	0.50	28.5	0.000	20.8	4.08
SimCLR	36.6	0.027	1.59	1.48	19.8	0.005	1.56	1.85	30.1	0.011	0.70	1.54	18.3	0.002	2.76	2.18	37.9	0.057	0.36	1.26	42.0	0.378	0.00	0.41