Project 16: Active Learning

02456 – Deep Learning

DTU Compute
Department of Applied Mathematics and Computer Science

Mads Fugl Petersen (s202795)

Evaluating Active Learning Strategies and CLIP-Predicted Label Initialization for Label-Efficient CIFAR-10 Classification with EfficientNet-B0

Introduction

Motivation & Goals: I investigate how active learning and automated labeling can reduce the data labeling bottleneck in deep learning while examining computational trade-offs. I evaluate three active learning strategies against stratified sampling on CIFAR-10, assessing efficiency in label requirements, training costs, and handling of class imbalance. Additionally, I explore CLIP for automated initial labeling, leveraging its zero-shot classification to bootstrap the active learning process with high-confidence predictions.

Key Questions:
- How do active learning strategies perform versus stratified sampling across varying sizes of balanced and imbalanced CIFAR-10 subsets?
- How do computational costs compare across strategies and configurations?
- Can CLIP's zero-shot classification provide reliable initialization labels while maintaining performance?

Active Learning 101

Active Learning Overview: Active learning iteratively selects informative samples for labeling, aiming to maximize model performance while minimizing labeling costs. The model identifies valuable unlabeled samples based on selection criteria, gets their labels, and retrains with the expanded dataset.

Strategies Evaluated:
- **Uncertainty Sampling:** Selects samples where model confidence (max class probability) is lowest, targeting ambiguous cases.
- **Margin-Based:** Focuses on decision boundary cases by selecting samples with smallest difference between top two predicted probabilities.
- **TypiClust:** Combines uncertainty (entropy) and typicality (cluster distance) with equal weights to select both uncertain and atypical samples.

Experiment Setup

Training Loop:
- EfficientNet-B0 (ImageNet pre-trained) fine-tuned on CIFAR-10.
- Identical hyperparameters across all experiments.
- 5 runs with fixed seeds for robustness.
- Data augmentation (training only) and 15% stratified validation.
- Early stopping with 10 epochs patience.

Key Parameters:
- **Training:** Adam optimizer, batch size 32, LR Scheduler.
- **Active Learning:** 10% initial labels (min. 150), 5% acquisition batch.

CLIP (Contrastive Language-Image Pre-training) Setup:
- Model: CLIP-ViT-Base-Patch32 from OpenAI.
- Zero-shot classification on CIFAR-10 classes; Batch size 32 for predictions.

Dataset Number	Percentage of Full Dataset	Total Samples	Distribution
Test Set	100%	10000	Balanced
1	100%	50000	Balanced
2	75%	37500	Balanced
3	50%	25000	Balanced
4	25%	12500	Balanced
5	10%	5000	Balanced
6	5%	2500	Balanced
7	3%	1500	Balanced
8	1%	500	Balanced
9	2.6%	1300	Imbalanced (8C@3%, 2C@1%)
10	8.2%	4100	Imbalanced (8C@10%, 2C@1%)
11	40.2%	20100	Imbalanced (8C@50%, 2C@1%)

Experiment 1: Performance Across Balanced Dataset Subsets

Balanced Dataset Evaluation:
- Compared three active learning strategies to a baseline that uses stratified sampling.
- Equal total label budgets across all sampling strategies.
- Active learning strategies have access to unrestricted label selection within the budget.

Key Findings:
- Similar effectiveness across all active learning strategies.
- Active learning achieves ~73% accuracy vs. baseline's 67.1% at 10% data.
- TypiClust shows 1.5x longer training time due to clustering.
- Performance advantage remains across dataset sizes until 1% data.
- F1 scores align with accuracy, indicating balanced performance.

Figure 1: Test Accuracy by Sample Size for Balanced Dataset Configurations

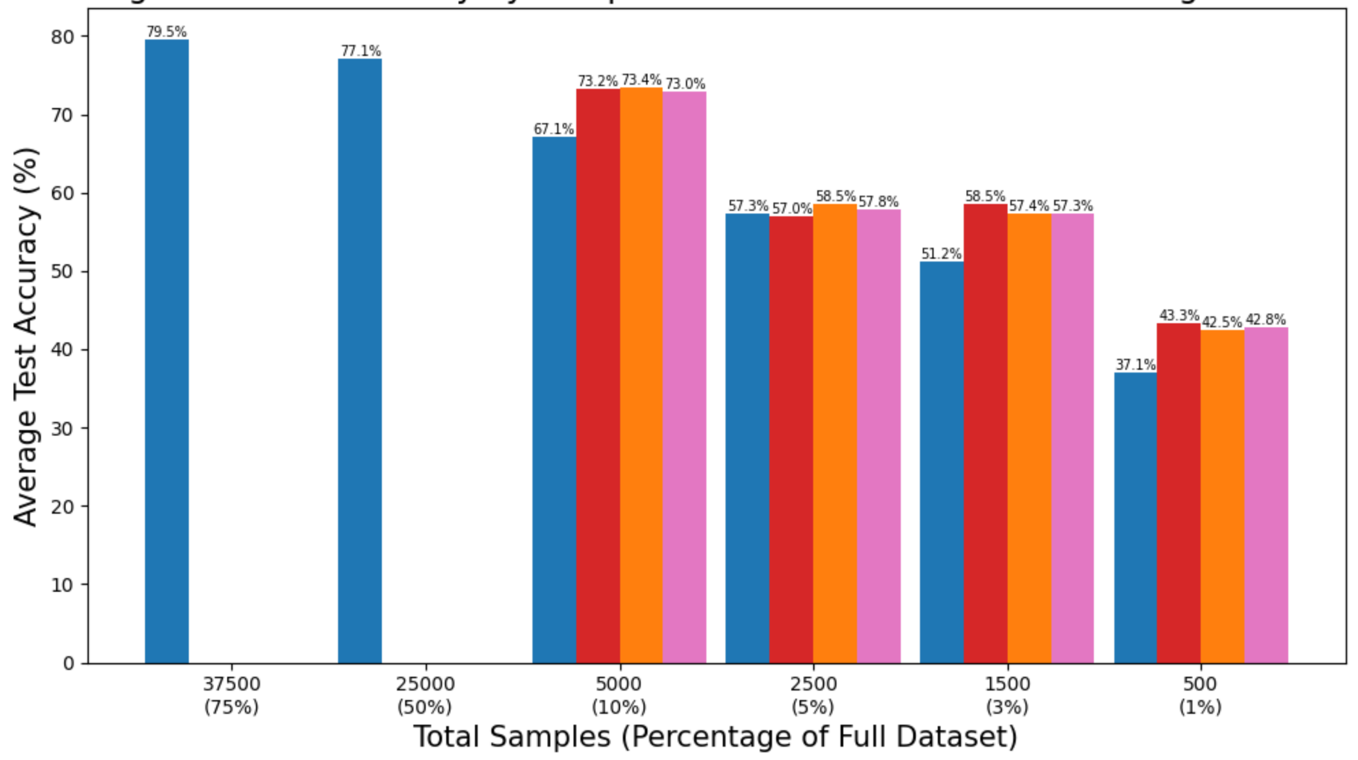
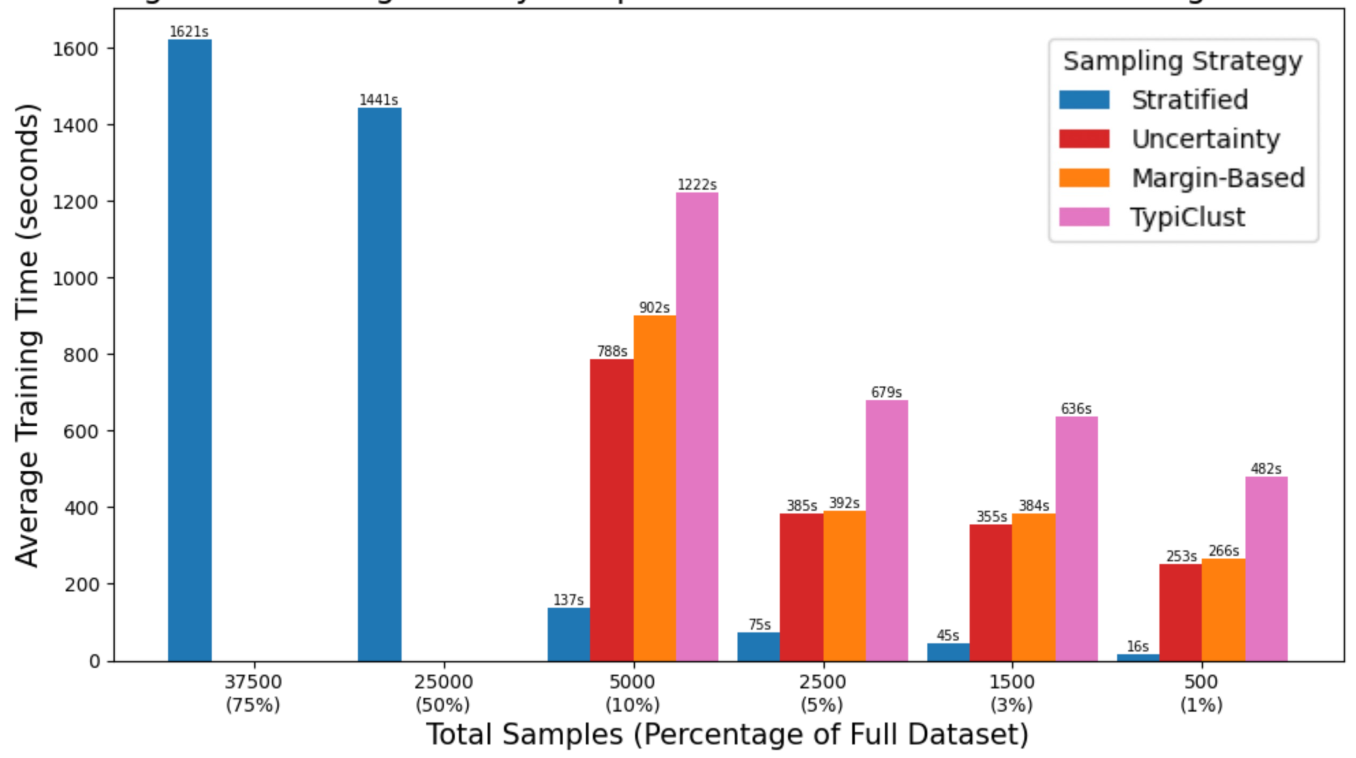


Figure 2: Training Time by Sample Size for Balanced Dataset Configurations

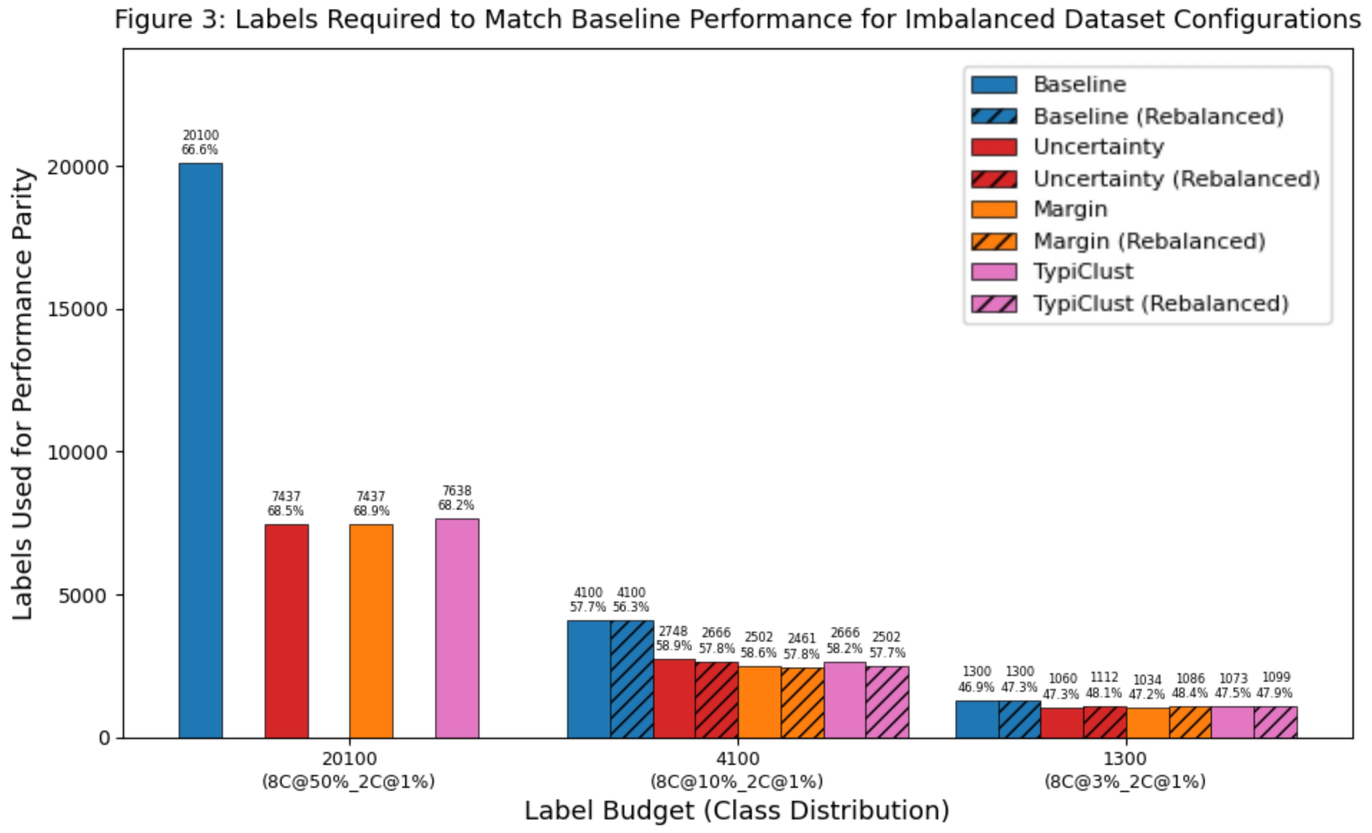


Experiment 2: Performance Under Class Imbalance

Imbalanced Dataset Evaluation:
- Tested on three imbalance scenarios:
  - 8 classes at 50%, 10%, or 3%.
  - 2 classes at 1%.
- Active learning models trained until exceeding baseline's performance (higher accuracy and lower loss on test set).
- Compared standard and class-weighted training.

Key Findings:
- Active learning strategies achieved baseline performance with dramatic label reductions:
  - 64% reduction at 50%/1% (~7.5K vs 20.1K labels).
  - 39% reduction at 10%/1% (~2.5K vs 4.1K labels).
  - 20% reduction at 3%/1% (~1.0K vs 1.3K labels).
- Class rebalancing slightly decreased performance in most cases.
- Similar efficiency across all active learning methods.
- Higher imbalance ratios enable greater label reduction.

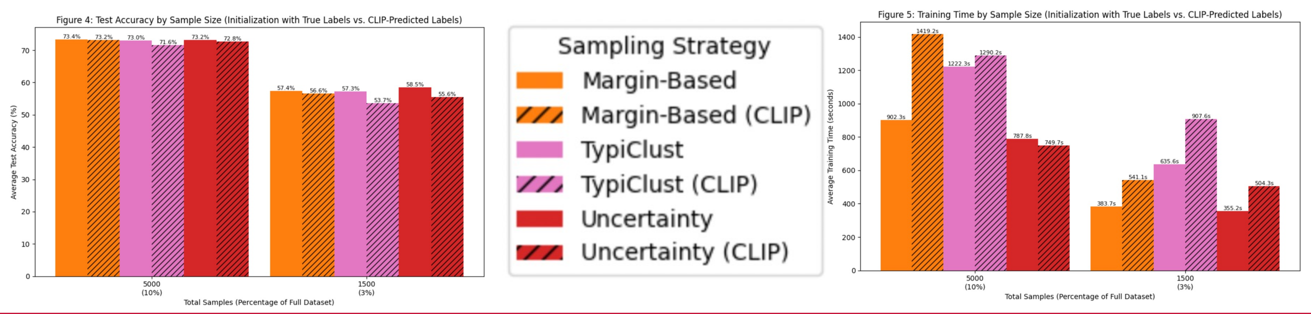
Figure 3: Labels Required to Match Baseline Performance for Imbalanced Dataset Configurations



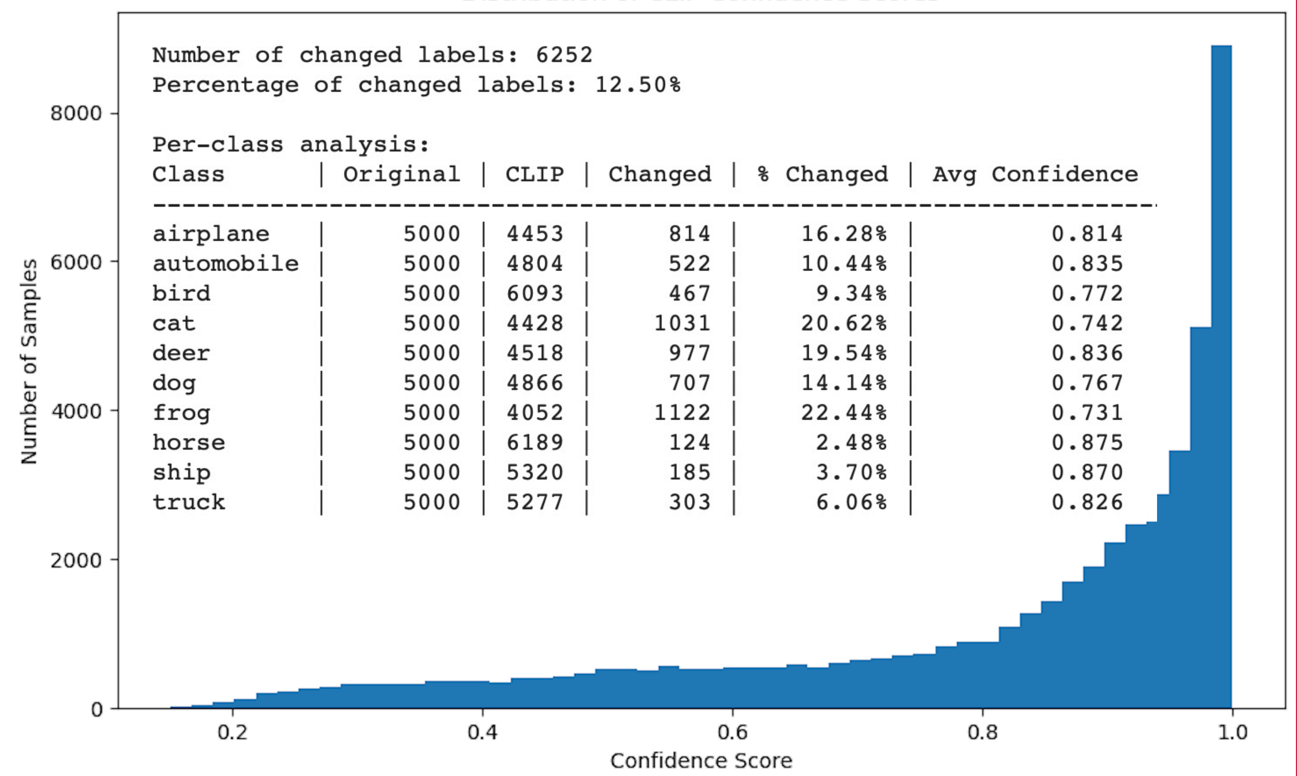
Experiment 3: Performance Using CLIP-Predicted Labels

CLIP Bootstrapping Evaluation:
- Used balanced 3% and 10% subsets of CIFAR-10.
- CLIP selected 10% of total samples, with predicted labels of highest confidence per class for the initial label pool.

Key Findings:
- At 3% dataset size:
  - Strategies using true labels slightly outperform CLIP initialization.
  - Consistent 40-42% increase in training time across all strategies.
- At 10% dataset size:
  - All strategies achieve comparable accuracy (71.5-73.5%) with both true and CLIP predicted labels.
  - Uncertainty Sampling and TypiClust show minimal changes (-5% and +5%).
  - Margin-Based Sampling shows 57% increase.



Distribution of CLIP Confidence Scores



Discussion

- Active learning strategies' similar effectiveness suggests fundamental robustness in sampling principles, regardless of approach complexity.
- Weight rebalancing's ineffectiveness challenges assumptions about class imbalance handling, potentially due to inherent robustness from transfer learning or guidance from stratified validation sets.
- Computational overhead patterns reveal important trade-offs between sophisticated sampling (TypiClust) and simpler approaches.
- CLIP initialization's varying impact across strategies suggests complex interactions between prediction confidence and sampling criteria.
- The synergy between Uncertainty Sampling and CLIP hints at complementary strengths in their approaches to selection.

Future Research Directions

- What characteristics truly determine sample informativeness across different sampling strategies?
- Do sample selection patterns reveal method-specific strengths in handling different class relationships?
- How can stratified sampling's speed be combined with active learning's efficiency?
- How does transfer learning from ImageNet weights contribute to class imbalance robustness?
- How can TypiClust's clustering mechanism be optimized dynamically for better efficiency?
- Can batch diversity criteria improve active learning sample selection?
- What factors affect CLIP's confidence patterns and initialization impact across classes and strategies?
- Why do CLIP initialization and Margin-Based sampling show unique computational patterns?"

Conclusion

- Active learning strategies consistently outperform stratified sampling baselines across all tested dataset sizes.
- Active learning significantly reduces labeling requirements while maintaining model performance.
- Advantages are most pronounced under high class imbalance, achieving 64% label reduction in 50%/1% scenarios.
- CLIP initialization achieves comparable accuracy but reveals dataset size-dependent training impacts.
- Study reveals intriguing patterns that raise rich questions for future research.

References

- EfficientNet: huggingface.co/docs/transformers/model_doc/efficientnet
- CIFAR-10: cs.toronto.edu/~kriz/cifar.html
- ImageNet: image-net.org
- Active Learning on a Budget: arxiv.org/pdf/2202.02794
- CLIP: huggingface.co/openai/clip-vit-base-patch32