A Unified Batch Selection Policy for Active Metric Learning

IIT BOMBAY

Priyadarshini K, Siddhartha Chaudhuri, Vivek Borkar, and Subhasis Chaudhuri

IIT Bombay and Adobe Research

What makes objects similar?

Clustering

Product recommendation

Wildlife search

Preference learning

Requires capturing humans' notion of perceptual similarity

Google Image

Human supervision for metric learning

A

B

C

Does A taste more similar to B or C?

{"yes", "no", "can't say"}

- Less subjective
- Less inconsistent
- Easy for human

- Annotation complexity is huge - $O(n^3)$

Active metric learning (AML) – label smarter

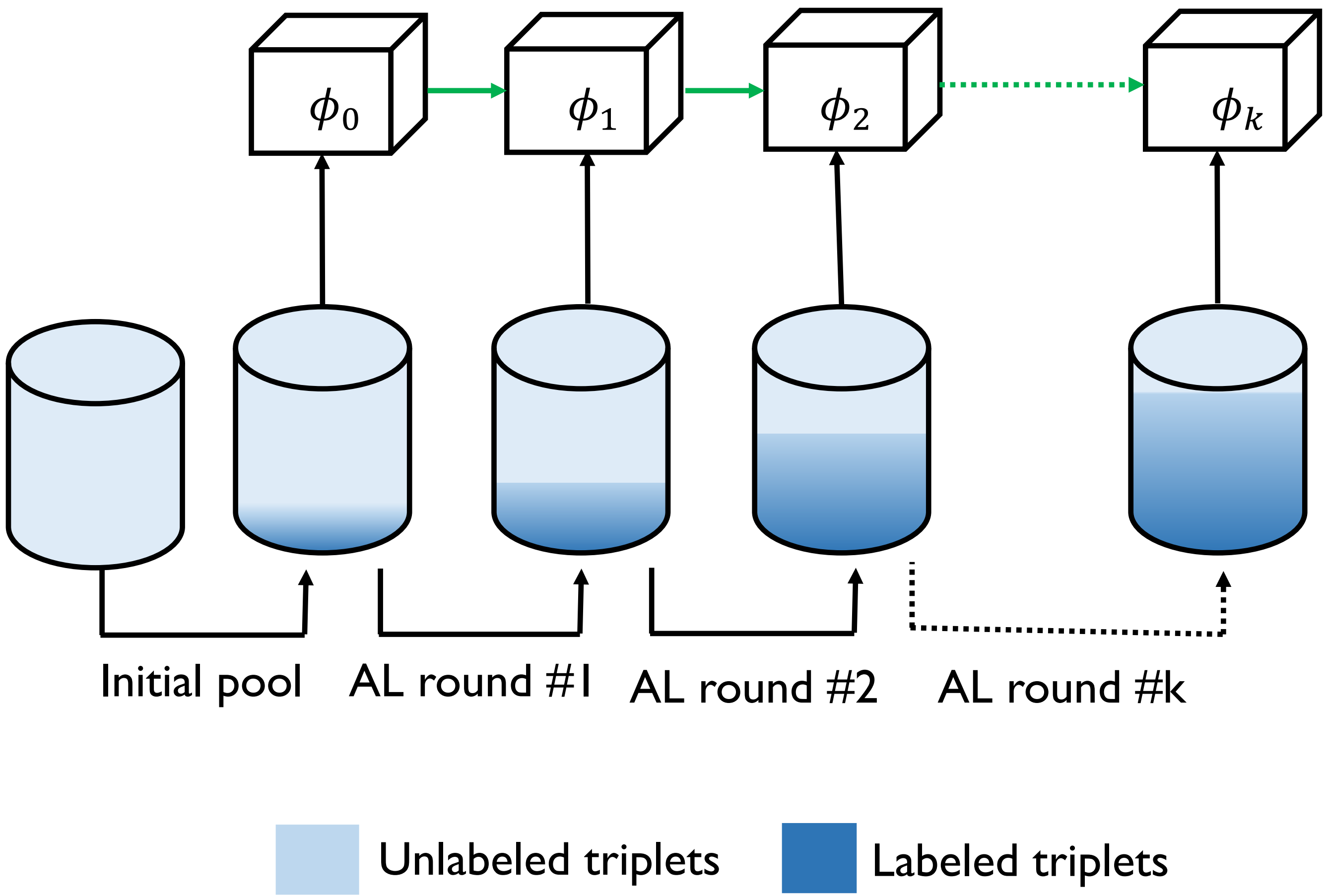
- Goal –
  - To learn an effective continuous perceptual metric using the minimum possible annotated triplets

- Key insights –
  - All triplets are not equally informative for the model
  - A good model can be trained on much fewer high-utility triplets

Two stages of active metric learning

(1) Triplet selection - Choose a subset of most informative triplets to annotate

(2) Model update - Train the model on the updated training set



Which triplets are informative?

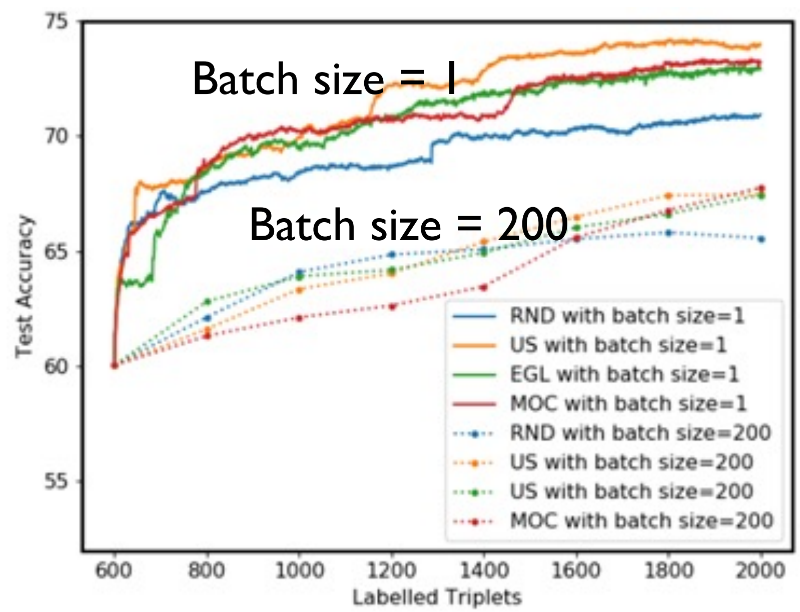
Existing work: use single-instance **uncertainty-based** informativeness measure

$$S^* = \operatorname*{argmax}_{\{S \subset U\}} H(S) = \sum_{t \in S} H(t)$$

Overestimates collective informativeness

$$= \sum_{t \in S} \underbrace{-p_{ijk} log p_{ijk} - p_{ikj} \log p_{ikj}}_{\text{Entropic measure}}$$

In batch mode, correlation b/w triplets is a huge problem



Batch AML – diversity is necessary

Our previous work: decoupled measures for informativeness and diversity

[Kumari, Chaudhuri, and Chaudhuri IJCAI2020]

Entropy

Decorrelation metrics

Separate measures do not ensure optimal tradeoff b/w both criteria

We proposed joint entropy as a unified measure to jointly balance both informativeness and diversity

$$H(S) = -\int p(x) log p(x) dx$$

How to define $p(x)$ for a batch of triplet ?

Defining probability distribution

Priors

We characterized the probability distribution by 2nd order moments estimated in distance margin space $\xi_t = d_\phi^2(x_i, x_k) - d_\phi^2(x_i, x_j)$ using dropout in neural network

Maximum entropy principle – Least biased estimate of probability distribution which best represents the prior state of knowledge is the one with the maximum entropy

$$maximize_{p(x)} - \int p(x) log p(x) dx$$

$$s.t \int p(x) r_i dx = m_i$$

Joint probability density function, $p(x)$ of a batch of triplets that satisfies 2nd order moment constraints and also maximizes the entropy is multivariate Gaussian

Optimum batch selection

Maximum informative batch $S^* = \underset{\substack{S \subset T_U \\ |S|=B}}{argmax} \;\; = \frac{1}{2} log((2\pi e)^B \det(G_S))$

Covariance matrix

Optimization is NP-hard

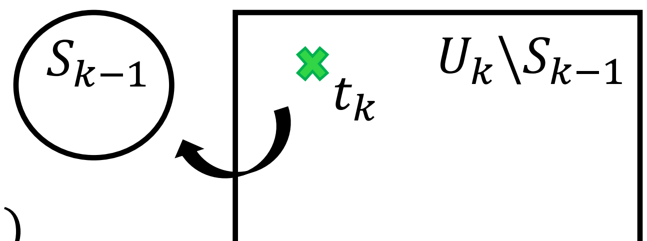
Monotone submodular optimization using greedy policy

$$S_0 = \emptyset$$

$$for\ k = 0, \ldots, B-1$$

$$t_k^* = argmax_{t_k \subset U \setminus S_{k-1}} \; H(\{t_k\} | S_{k-1})$$

$$S_k = S_{k-1} \cup \{t_k\}$$



How to efficiently compute conditional entropy?

[Nemhauser et al 78]

Recursive computation of conditional entropy

Maximize conditional entropy

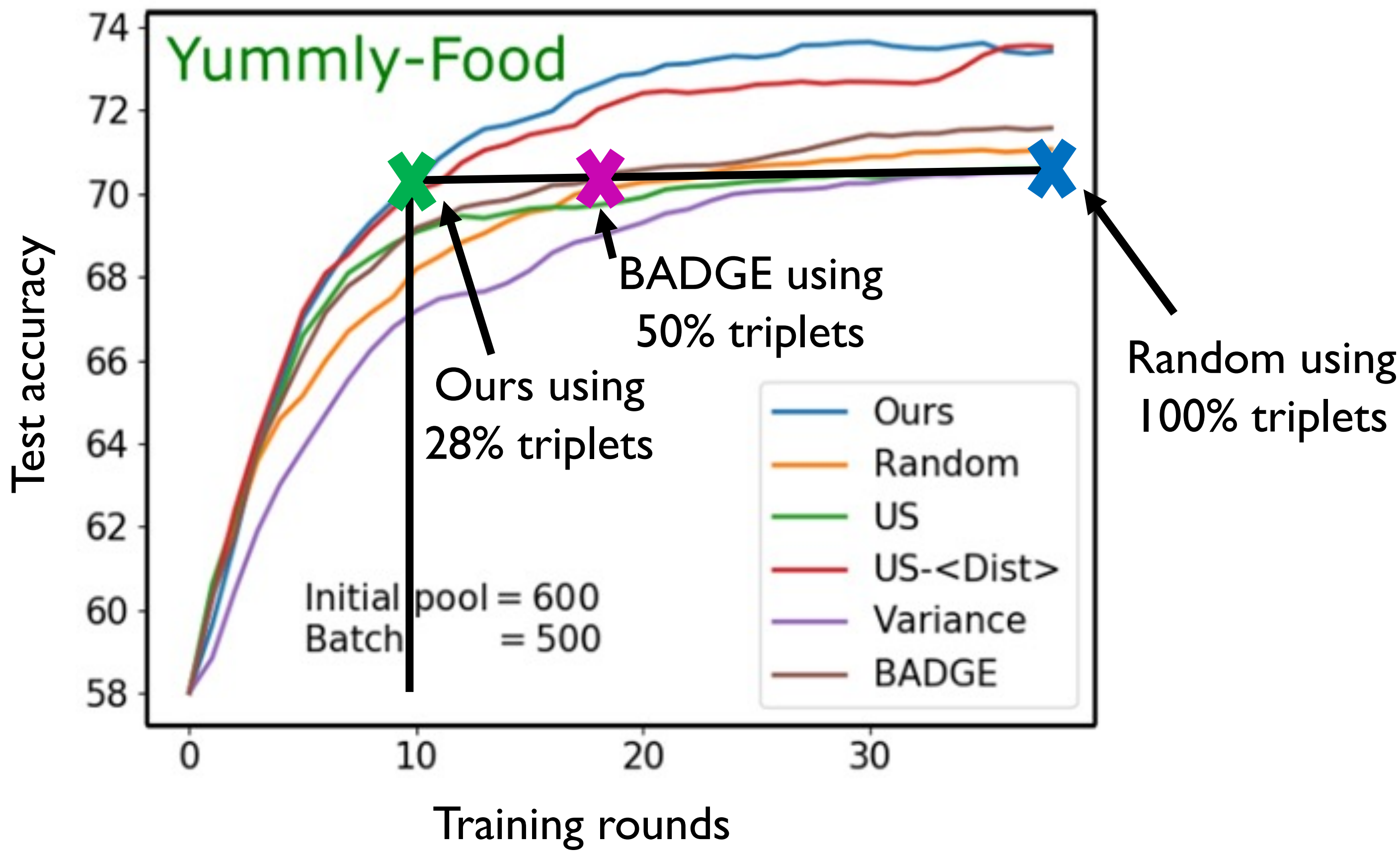
$$t_k^* = argmax_{t_k \subset U \setminus S_{k-1}} \; H(\{t_k\} | S_{k-1}) \rightarrow \log\left(\underbrace{\frac{\det\left(G_{S_{k-1} \cup \{t\}}\right)}{\det\left(G_{S_{k-1}}\right)}}_{\|\widetilde{u_t}\|^2}\right)$$

$\|\widetilde{u_t}\|^2$ is the orthogonal projection onto span of triplet set S_{k-1}

At each step the triplet that is least correlated with the already chosen triplets

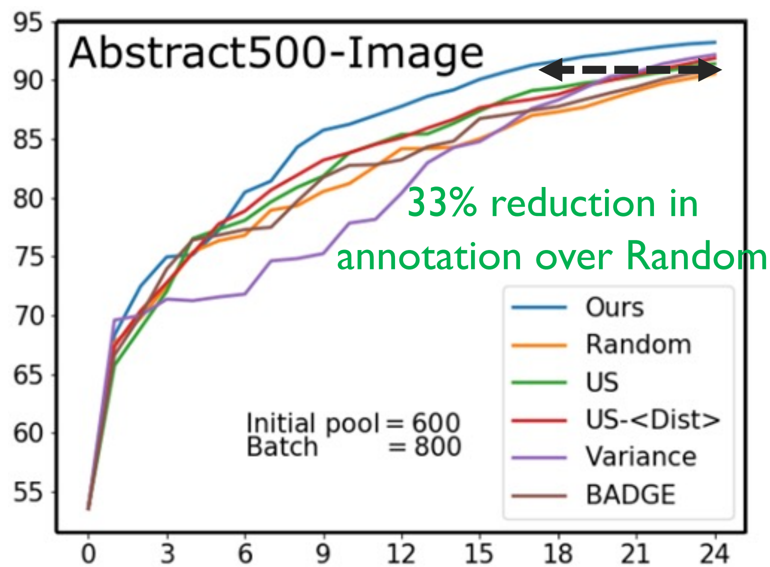
Results

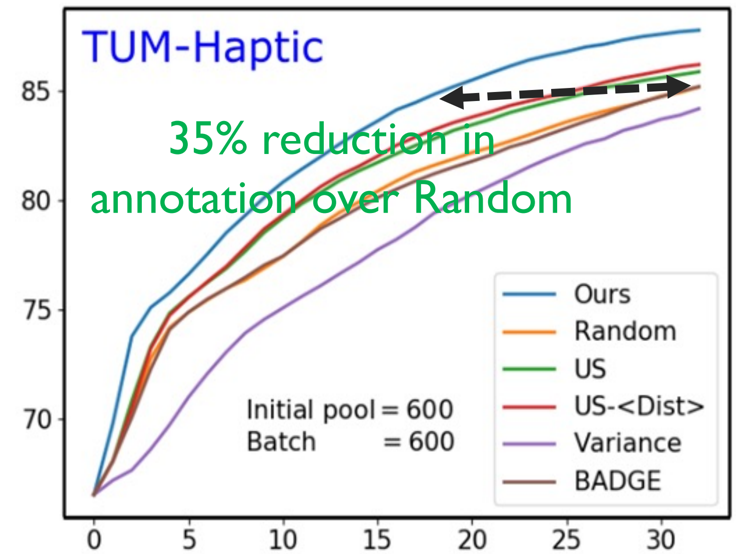
72% reduction in annotation over Random

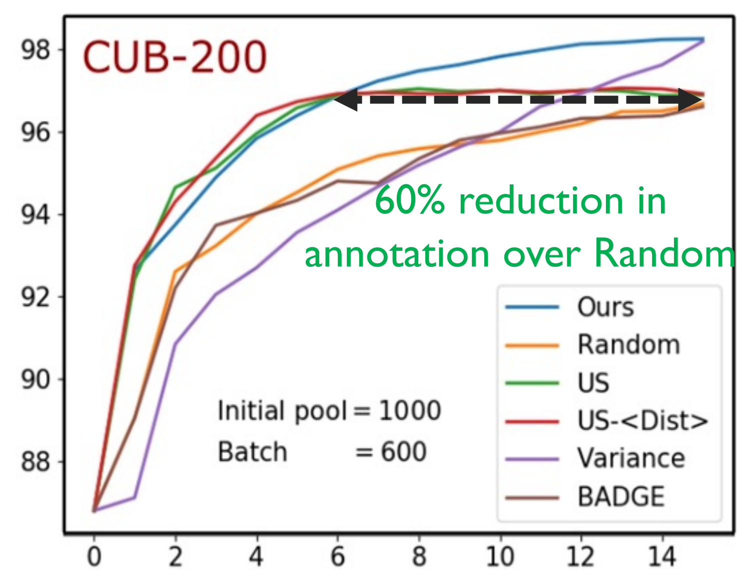


All variants of decorrelated active metric learning perform better than the Random and SoTA (BADGE)

Results







- Wide applicability across different modalities and dataset sizes
- Performance gain increases with increasing batch size

Results

Retrieved images in the order of increasing perceptual distance

Our method gives better perceptual matches with query than randomly-selected triplets at the same annotation cost

Takeaways

- Perceptual metrics can be effectively trained on far fewer examples if unlabeled samples (triplets, in our case) are chosen intelligently for annotation
- Unified measure for informativeness and diversity is important for optimum batch selection

Future directions

- Annotation effort can further be reduced if we learn the data selection policy dynamically
- Choose not just informative samples but also informative input modality