Efficient Techniques for Deep Fine-grained Feature Learning

Reduction in Training Time and Accuracy Improvements

Amar Saini & Wesam Sakla

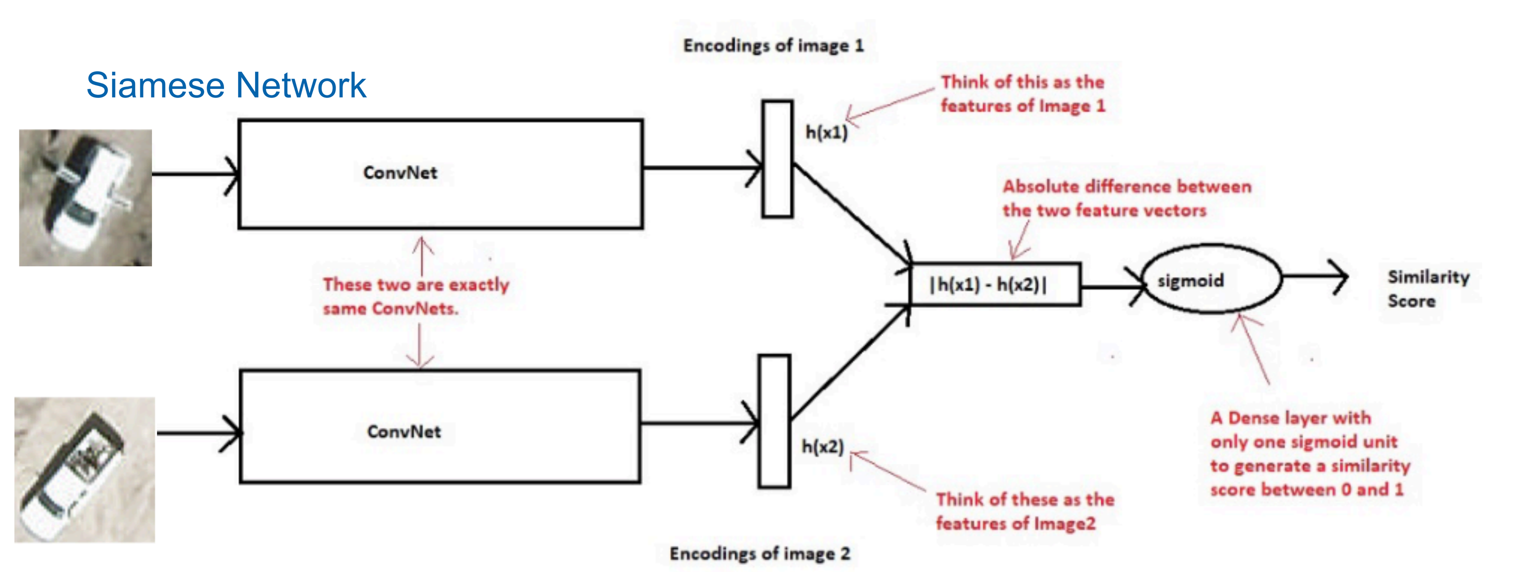
DSSI DATA SCIENCE SUMMER INSTITUTE

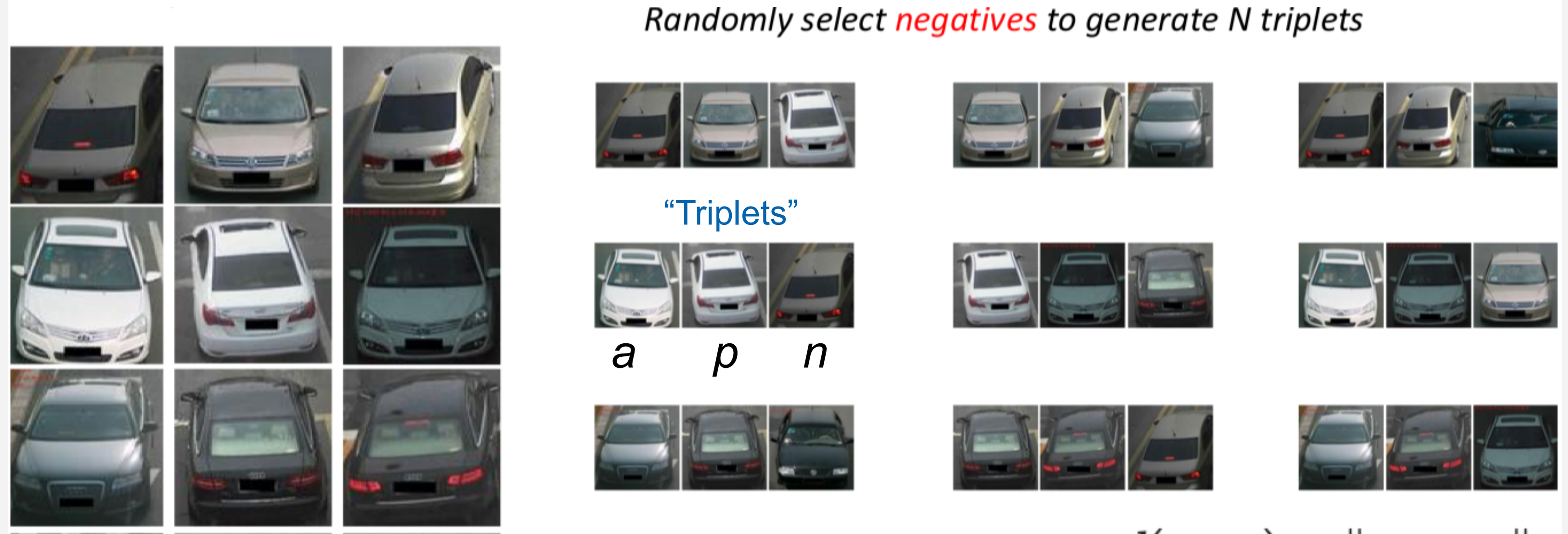
Lawrence Livermore National Laboratory

For fine-grained feature learning from overhead imagery, we are investigating alternate techniques for training deep CNNs that yield results comparable to the famous Triplet Loss, but with a reduction in training time. The Triplet Loss is popular for the image retrieval task, but suffers from extensive sampling during training. To improve training time and possibly retrieval accuracy, we propose to use an alternate loss function, *Weighted Approximate-Rank Pairwise Loss* (WARP Loss), with some minor optimizations to improve runtime. We are also experimenting with cosine distance layers to better learn features, and improve image retrieval accuracy.

Image Retrieval

- For the image retrieval task, the goal is to provide a query image and return the most similar images.
- Two main types of models that learn features for the image retrieval task: Siamese Convolutional Neural Networks (CNNs) and CNNs trained with a Triplet Loss objective.





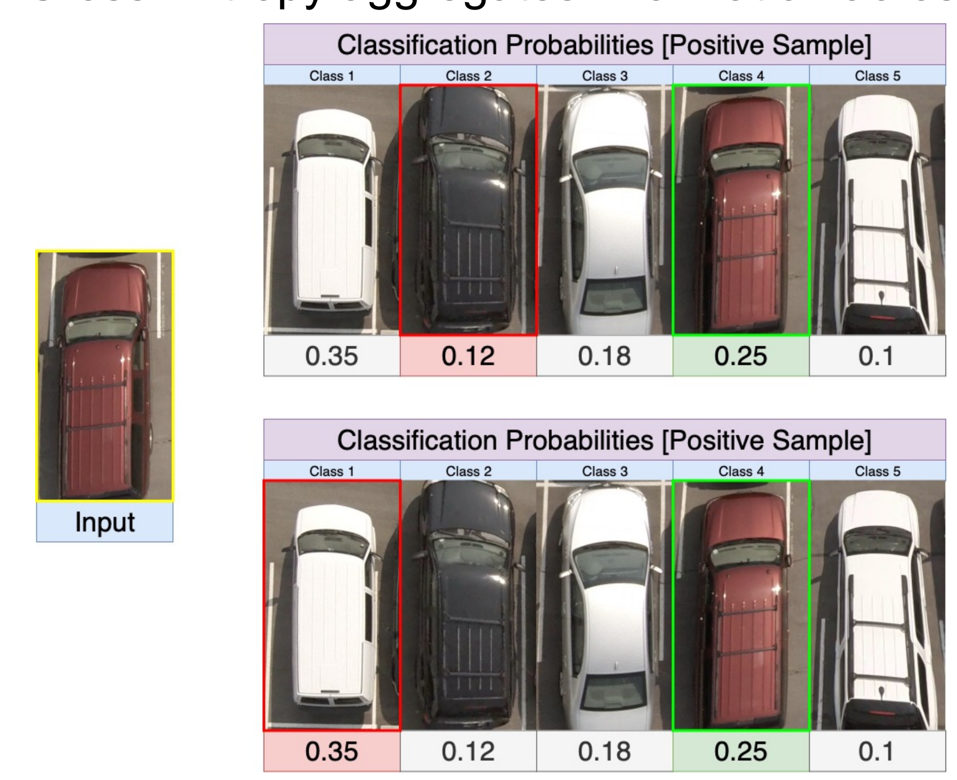
Triplet Loss Function

$$d(x_i, y_i) = \|\mathbf{x}_i - \mathbf{y}_i\|_p$$

$$L(a,p,n) = \max\{d(a_i,p_i) - d(a_i,n_i) + \text{margin}, 0\}$$

WARP Loss

- Focuses on learning using a pairwise ranking objective.
- Cross Entropy aggregates information across all classes.



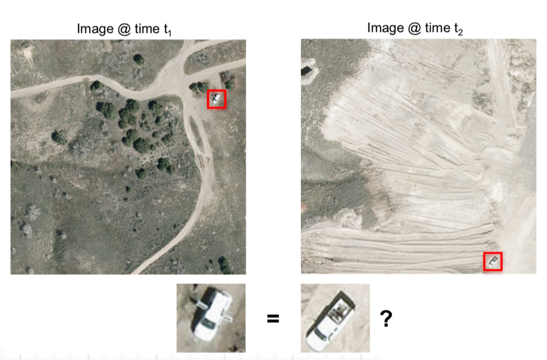
$$\text{Loss} = ln\left(\frac{X-1}{N}\right)(x_2^- - x_3^+)$$

WARP Loss Equation

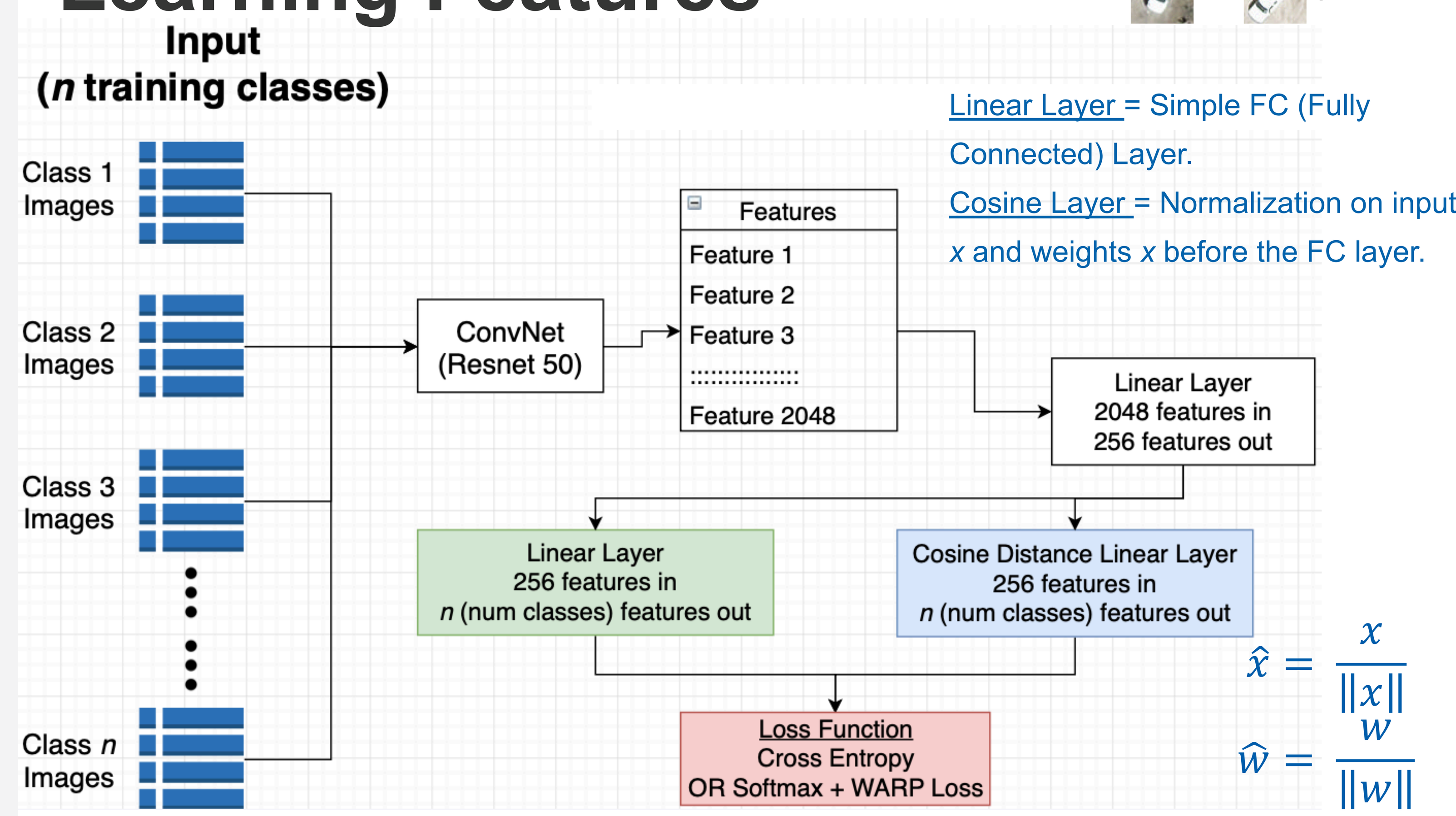
X = Number of samples

x^- = Prob. Of neg. sample

x^+ = Prob. Of pos. sample



Learning Features

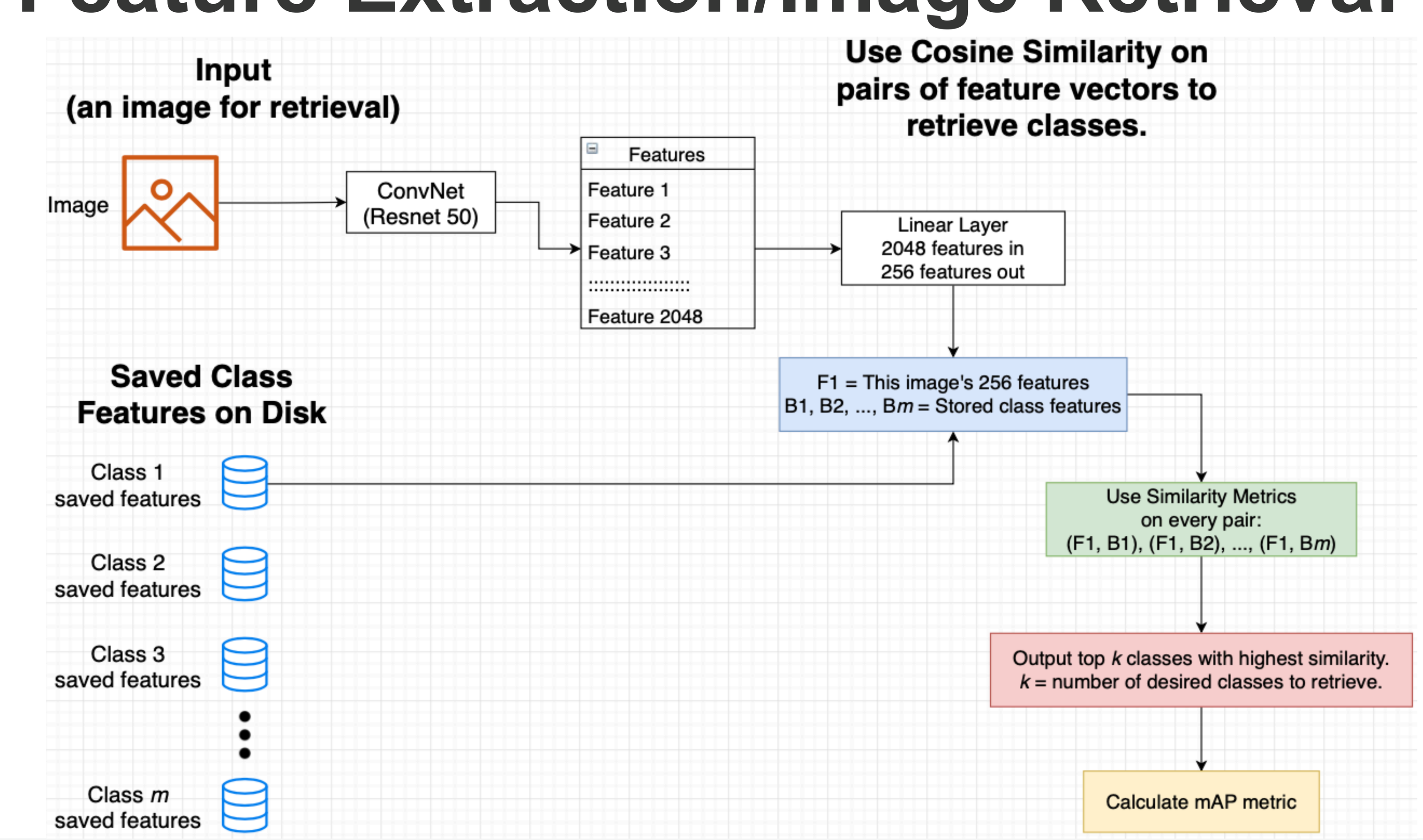


Linear Layer = Simple FC (Fully Connected) Layer.

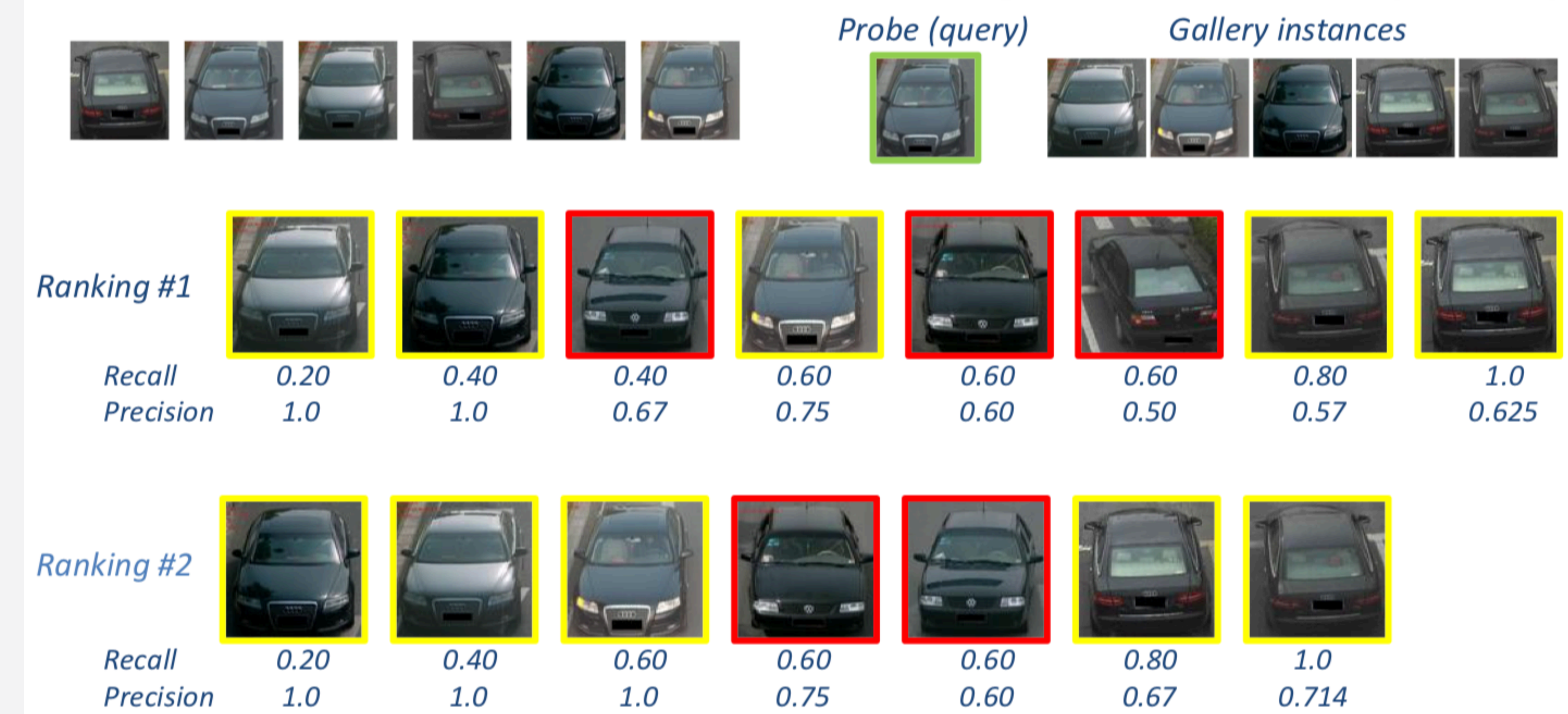
Cosine Layer = Normalization on input x and weights x before the FC layer.

$$\hat{x} = \frac{x}{\|x\|} \qquad \hat{w} = \frac{w}{\|w\|}$$

Feature Extraction/Image Retrieval



Evaluation Metric: mean Average Precision (mAP)



AP for Ranking #1: (1.0 + 1.0 + 0.75 + 0.57 + 0.625)/5 = 0.789

AP for Ranking #2: (1.0 + 1.0 + 1.0 + 0.67 + 0.714)/5 = 0.8768

Experimental Details & Results

- 60% classes: learning features via training (classification/triplet loss)
- 40% unseen classes: image retrieval (mAP calculation).
- 32x32 Grayscale for BYU-Cars dataset (8.3k images).
- 101x101 Grayscale for Satellite-Cars dataset (62.8k images).
- ResNet152 architectures, 256 embeddings/features.
- WARP Loss fine-tunes the "CrossEntropy No Cosine" model.

BYU-Cars Dataset	Training Time	Epochs	Classification Accuracy	mAP
Triplet Loss	24 min	35 (unfrozen)	N/A	76.41%
CrossEntropy No Cosine	22 min	50 (frozen)	83.8%	62.28%
WARP Loss	10 min	10 (fine-tuned)	84.10%	62.03%
CrossEntropy with Cosine	22 min	50 (frozen)	90.46%	67.82%
Satellite-Cars Dataset	**Training Time**	**Epochs**	**Classification Accuracy**	**mAP**
Triplet Loss	1 hr 39 min	35 (unfrozen)	N/A	17.66%
CrossEntropy No Cosine	1 hr 7 min	50 (frozen)	31.47%	16.76%
WARP Loss	12 min	6 (fine-tuned)	31.53%	16.10%
CrossEntropy with Cosine	1 hr 7 min	50 (frozen)	34.17%	17.45%

Conclusions

- *Triplet Loss* suffers in training time due to the online sampling process.
- Using image classification with *CrossEntropy* or *WARP Loss* is an alternate method for learning features for image retrieval, that can maintain relatively good mAP scores.
- Cosine distance-based classification leads to better classification accuracy and mAP (retrieval) scores.

References

- *A Closer Look at Few-shot Classification*, ICLR 2019.
- *WSABIE: Scaling Up To Large Vocabulary Image Annotation*, 2011.
- *FaceNet: A Unified Embedding for Face Recognition and Clustering*, IEEE CVPR 2015.

WARP Loss + Cosine Distance Layer improves Fine-grained Image Retrieval Accuracy and Reduces Training Time.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.