SCALABLE LOGO RECOGNITION USING PROXIES

Istvan Fehervari
istvanfe@amazon.com

Srikar Appalaraju
srikara@amazon.com

Problem

Detecting logos in the Amazon catalog enables better brand infringement detection which helps earning/keeping customer trust.

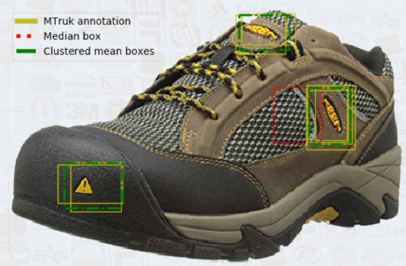
However, logo recognition is a challenging problem:

- no clear definition of a logo
- huge variations of logos, brands
- logo can appear in any context
- re-training to cover every variation is impractical

Dataset – PL2K

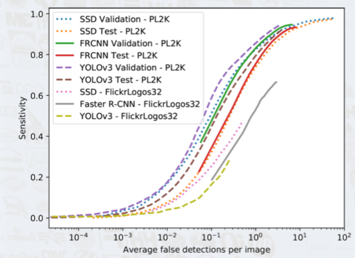
Sampled 1M product images from the Amazon Catalog biased towards high-visibility brands. Bounding boxes around each logo were added by MTurk. Data was clean up using density-based clustering.

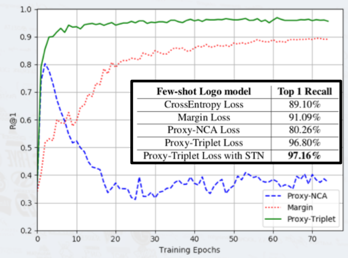
Image set	No. of images	No. of brands
Training	185247	206
Validation	46312	
Testing	57970	1528
Negatives	10000	2000



Results

All detectors reach high AP, SSD and YOLO3 produce a lot less FPs. SSD has the highest recall.
The proxy-triplet loss with spatial transformers outperform all other loss functions by a large margin.

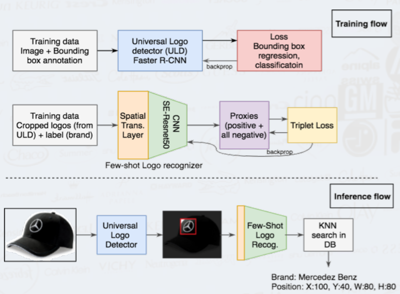




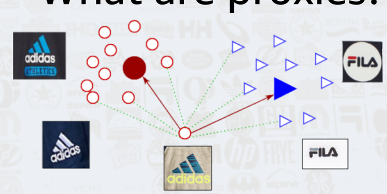
Left: FROC curve of the detector. Right: Test accuracy of the few-shot classifier with different loss functions.

Approach

We propose a deep learning-based two-step approach, where first a semantic logo detector identifies rectangular regions of an image where a logo might be located and a second model identifies its class/brand using only features in the identified region.
The class-agnostic detector is based on modern deep object detector neural network architecture. The classifier model is based on distance metric learning principles and is trained using proxies.



What are proxies?

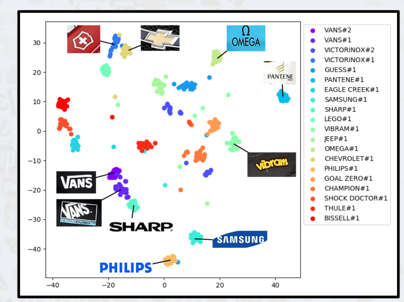


Proxies are randomly initialized “imaginary embeddings” that are co-learned with the embedding function using the following triplet-based loss function:

$$L_{triplet}(x, y, Z) = [d(x,p(y)) + M - d(x,p(Z))]_+$$

Proxies do not require any sampling strategy or large batch sizes. Performance and convergence rate is superior compared state of the art methods.

Analysis



Left: Retrieval results of the few-shot model. Left column contains query regions. Model successfully maps unseen logos even when there is noise, color swaps and partial occlusions.

Right: T-SNE plot of a random subset of (unseen) test classes. Logos of the same class are mapped close to each other, even with high inter- and intra-class variations.