A Survey of Techniques for Fine-Grained Pill Image Matching

Adam Allevato Andrew Sharp

The University of Texas at Austin

US NIH Pill Image Recognition

- **Goal:** Produce a system to identify cellphone-camera pictures of consumer medication
- **Input:** M consumer-quality (CQ) images and N high-quality reference (Ref) images
- **Output:** M x N matrix of similarity ranks

Dataset [http://pir.nlm.nih.gov/challenge]

2 reference images & 5 CQ images: taken by Cellphone cameras

1000 pills, 2000 ref images, 5000 CQ images

Challenges

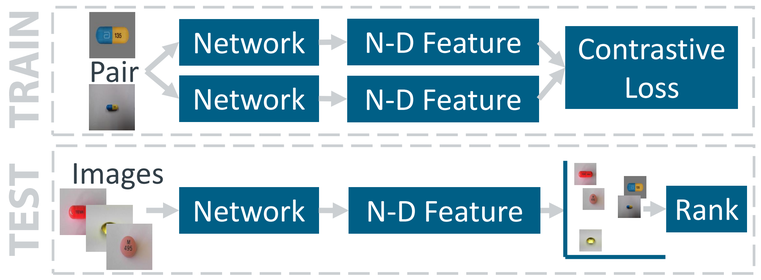
- **Matching**: Classes are unknown at training time
- **Fine-Grained:** Differences between pills are small
- **Small Dataset**: Near zero-shot matching task
- **Many Classes**: 1000-way matching
- **Incomplete Data:** Need to infer back of pill from front

Observation: even state-of-the-art CNN approaches struggle with this task

Key Idea: Siamese Networks

- Learn a discriminative mapping into N-D feature space
- Minimize *contrastive loss* (Y=1 if similar pair, else Y=0):

$$E = \frac{1}{2N}\sum_{n=1}^{N}(y)d^2 + (1-y)\max(margin - d, 0)^2$$ [Chopra et al. 06]



Experimental Setup

Preprocessing

- **Selective Search** [J. R. R. Uijlings et al. 13] was used to crop consumer-quality images
- Networks were tested on selective search images (SS) and raw images (Full), as well as both together (Comb)
- SS alone generally improved mAP by about 0.001 (7%)

Successful SS Crops

Failure Cases

LeNet

8 layers
Trained from scratch

Siamese configuration
2-dimensional mapping

CaffeNet

21 layers
Fine-tuned last layer

Siamese configuration
20-dimensional mapping

Data Setup

- 30% of images used as test (N=600)
- Hard Negative Mining: Test networks on training set, generate new training pairs from errors, train again
- Front/Back: Use contextual knowledge (filenames) to correlate pill fronts with pill backs

Results: mAP (N=600)

-Based on **top 2** ranks (pill front and back)

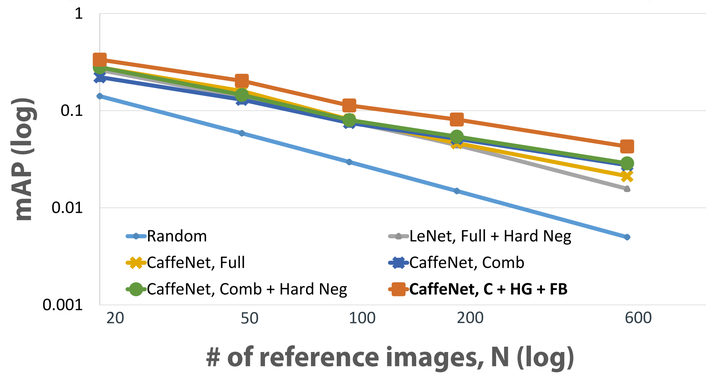
$$MAP = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{2}\sum_{j=1}^{2}\frac{j}{MT(i,j)}\right)$$

Network	mAP	+ Hard Neg	+ Front/Back
Random	0.0050	-	-
LeNet, Full	0.0148	0.0157	0.0280
LeNet, SS	0.0161	0.0175	0.0326
CaffeNet, Full	0.0278	-	0.0323
CaffeNet, Comb	0.0275	**0.0287**	**0.0428**

Observation: Combination of data mining techniques results in best performance for this challenging dataset

Analysis

Dependence on number of reference pills



T-SNE Visualizations

CaffeNet Comb+Hard Neg

LeNet + Hard Neg

Best Matches (N=600)