Paper reading seminar

Yan Wang

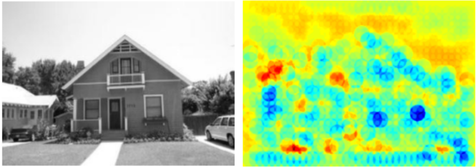
September 6, 2012

What Has My Classifier Learned? Visualizing the Classification Rules of Bag-of-Feature Model by Support Region Detection

- Interesting
  - Motivation
  - Applications
- Motivation
  - Critical parts of the image for classification
  - Sparse Coding + Max Polling

Background

- Related work: heat map (Li Fei-Fei, CVPR 11)
  - SVM + random forest + dense sampling ⇒ image classification
  - Image region selection ⇒ heatmap



(a) Original Image (b) Heat Map Image

- Assumption
  - SC + MP ⇒ no linearity: $C(A \cup B) \neq \mu C(A) + (1 - \mu) C(B)$
  - Interested in regions without which the image will be misclassified

Approach

- SC + MP
  - Sparse coding: $u_j \in \mathbb{R}^V$. u_{jk}
  - Pooling: $\max_j\{u_{jk}\}$
  - Classification: $\hat{y} = sgn\left(w_k \max_j\{u_{jk}\} + b\right)$
- Support Region R_s
  - Only interested in positive samples

$$\sum_k \max_{\{j|P_j \notin R_s\}} \{u_{jk}\} + b < 0$$

  - Restriction on area

Approach

- An efficient (DP-like) way to detect such regions

$$J(R_p, R_q) = \sum_k w_k \left(\max_{\{j|P_j \in R_p\}} \{u_{jk}\} - \max_{\{j|P_j \in R_p - R_q\}} \{u_{jk}\} \right)$$

- Don't really know what it means... The larger, the better? $R_q \subset R_p$

$$K(I, R_s) \geq S_0, S_0 = \sum_k \max_j \{u_{jk}\} + b$$

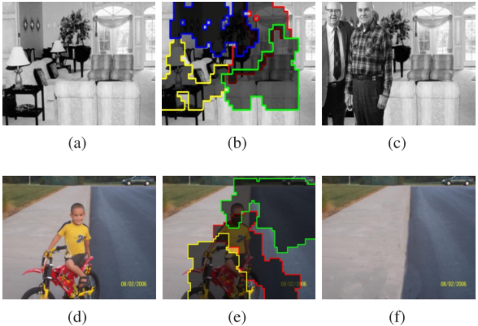
$$J(I, R_t) = J(I, R_{t-1}) + J(I - R_{t-1}, P_t)$$

- Iteration way. Start from P_0, $R_t = R_{t-1} \cup \hat{P}_t$

$$\hat{P}_t = \arg \max_{P_t \in \text{boundary}\{R_{t-1}\}} J(I - R_{t-1}, P_t)$$

Applications

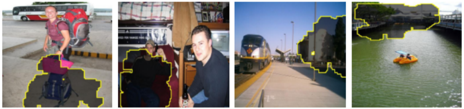
- Predict the failure mode of classifiers



Applications

- Understand the classification and discover database bias

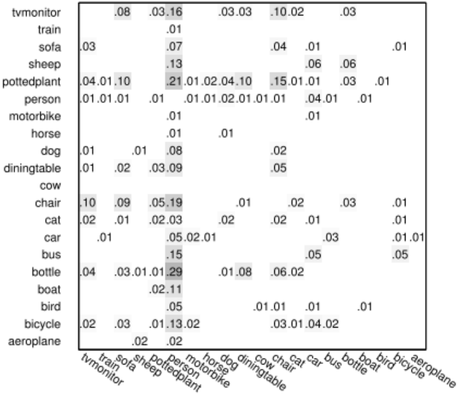
(a) chair (b) aeroplane (c) cow (d) sheep



(e) bus (f) TV/monitor (g) train (h) boat

Applications

- Understand the classification and discover database bias



	tvmonitor	train	sofa	sheep	pottedplant	person	motorbike	horse	dog	diningtable	cow	chair	cat	car	bus	bottle	boat	bird	bicycle	aeroplane
tvmonitor			.08		.03	.16			.03	.03		.10	.02			.03				
train						.01														
sofa	.03					.07						.04		.01					.01	
sheep						.13								.06		.06				
pottedplant	.04	.01	.10			.21	.01	.02	.04	.10		.15	.01	.01		.03		.01		
person	.01	.01	.01		.01		.01	.01	.02	.01	.01	.01		.04	.01		.01			
motorbike						.01								.01						
horse						.01			.01											
dog	.01			.01		.08						.02								
diningtable	.01		.02		.03	.09						.05								
cow																				
chair	.10		.09		.05	.19				.01			.02			.03			.01	
cat	.02		.01		.02	.03			.02			.02		.01					.01	
car		.01				.05	.02	.01							.03				.01	.01
bus						.15								.05					.05	
bottle	.04		.03	.01	.01	.29			.01	.08		.06	.02							
boat					.02	.11														
bird						.05					.01	.01		.01			.01			
bicycle	.02		.03		.01	.13	.02					.03	.01	.04	.02					
aeroplane				.02		.02														

Applications

- Build better classifiers
  - User annotates bad support regions $\Rightarrow$ remove the regions, put remaining image as a positive sample
  - Cross-dataset validation: context won't help, must focus on the objects themselves

Applications

- Build better classifiers
  - User annotates bad support regions ⇒ remove the regions, put remaining image as a positive sample
  - Cross-dataset validation: context won't help, must focus on the objects themselves

Test Setting	Bike	Car	Person
Without new samples Test on Graz	93.3	79.2	86.4
With new samples Test on Graz	93.8	79.7	**88.3**
Without new samples Test on Pascal	73.2	77.5	67.7
With new samples Test on Pascal	73.4	**80.9**	**68.4**

Discussions

- Is the new classifier simply getting rid of the contexts?
  - That depends on how to define the "bad support regions"
  - Define "good support regions"
    - On the object itself
    - On "reasonable" contexts (like boat river)
  - $\Rightarrow$ include less-bias contexts in the new classifier
- Is it possible to get rid of the human label?
  - Without the "good context", yes. Directly use the ground-truth objects
  - Otherwise need a way to determine whether a support region is "good"

Ensemble of Exemplar-SVMs for Object Detection and Beyond

- Motivation
  - "Semantic classifier" hard because things are not visually similar
  - Build "exemplar classifier" to do instance classification and then fuse the scores
  - SVM-style extension of KNN-based classification
  - (Another extreme from "Objectness")
- New problem
  - Detection dataset $\Rightarrow$ matching instance $\Rightarrow$ associated metadata (class, label, 3D model, segmentation, ...)

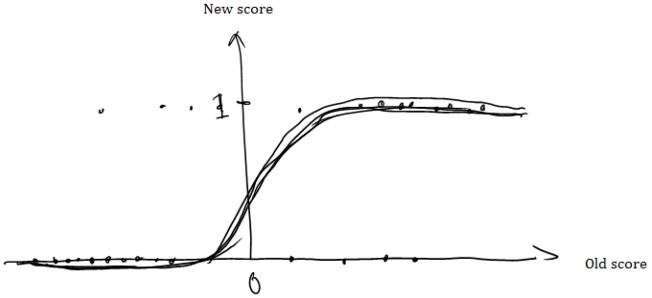
Approach

- Exemplar-SVM
  - Corresponds to every (detection) instance
  - Positive training sample: detection instance
  - Negative training sample: all the other instances
  - HoG + Linear SVM
  - Use non-maximum suppression in testing phrase

Approach

- Calibration
  - The individually trained SVMs have different scales
  - Cannot be integrated for final result directly
  - Calibration to normalize their score
  - For every Exemplar-SVM

$$f(x|w_E, \alpha_E, \beta_E) = \frac{1}{1 + e^{-\alpha_E(W_E^T x - \beta_E)}}$$



Approach

- Calibration
  - Sample collection
    - Agreed samples (w/in one class?) $\Rightarrow$ positive samples
    - Other samples
- Test
  - Sliding window + HoG
  - Individual test score
  - Calibrated test score
  - Sum? Max?

Data-driven Visual Similarity for Cross-domain Image Matching

- Problem: cross-domain image matching
- Motivation
  - Visual similarity metric is hard to design
  - Directly use the (unlabeled) data
  - Training a discriminative classifier to discover important parts of an image $\Rightarrow$ learn a similarity metric from data

Approach

- Train an SVM for every image
  - Pos: one image with shifting/rotating/scaling
  - Neg: (millions of) sliding windows from (10K) randomly sampled Flickr images
  - Feature: grid-like features with high enough dimensions, like HoG or Dense SIFT
- Technical details
  - Efficiency: SVM only keeps support vectors in the negative part $\Rightarrow$ efficient in testing phrase
  - "Sample expansion": explore hard negative samples with current trained w
  - Ranking: didn't find in the paper

Experiments

- Perform well in saliency detection
- Image-Image matching
  - Dataset: INRIA Holiday Dataset to ensure the exact query instance exists in the dataset
  - Baselines: GIST, Tiny images, spatial pyramid
  - Measurement: true positive rate in top 100 results

Experiments

- Sketch-Image matching
  - Dataset: 25 car sketch + 25 bike sketch against PASCAL VOC 2007 and SBIR Dataset
  - Baselines: proposed algo. with different dataset size
  - Measurement: mAP of top K
- Paint-Image matching (qualitative)
  - Dataset: self-collected, 50 outdoor paintings agains 5K GPS-enabled images within 50 miles and 5K random images

Applications

- Scene completion
- Internet re-photography
- Painting2GPS
- Visual Scene Exploration

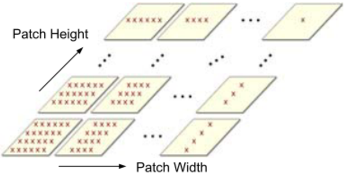
Backup slides

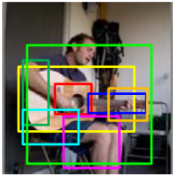
Combining Randomization and Discrimination for Fine-Grained Image Categorization

- Motivation
  - Fine-grained image categorization
  - Bird species
  - Human activity classification
- Intuition
  - Dense sampling $\Rightarrow$ patches
  - Correlation among patches
  - Random forest $+$ SVM

Approach

- Dense sampling

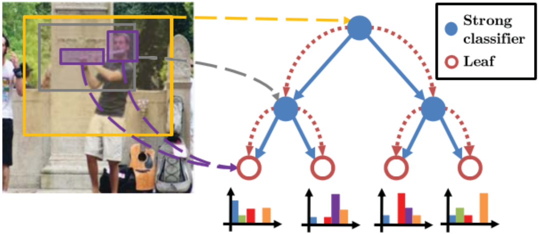




- Feature
  - Single patch: BoW
  - Patch pair: concatenation/intersection/absolute of difference of BoW histogram

Approach

- Random forest + SVM



(b) Discriminative decision tree.

  - Randomly select patches (or patch pairs) + SVM
  - Train random forest with information gain
  - Use “ancestor” features
  - Q: invariance?