TRAINING FREE, ONE SHOT DETECTION

Team: Marvin (Siddhant Katyan)
2018801005, IIIT HYDERABAD

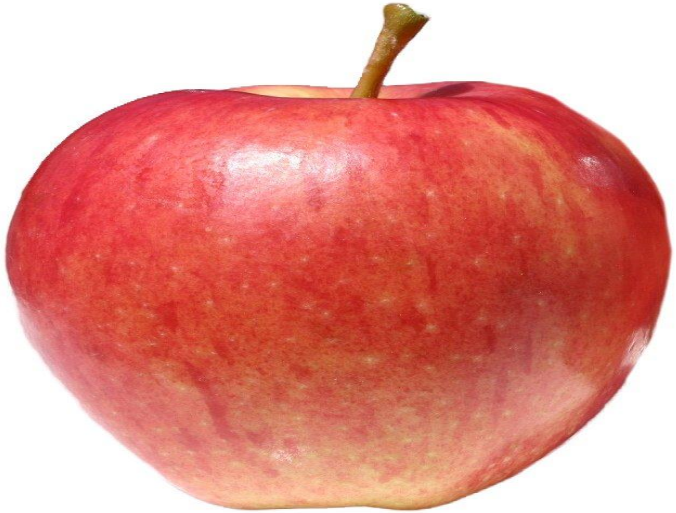
Take a look at this:

See it here?

How about here?

Single Example, No Training!

(Most) people can find the apple from one look.



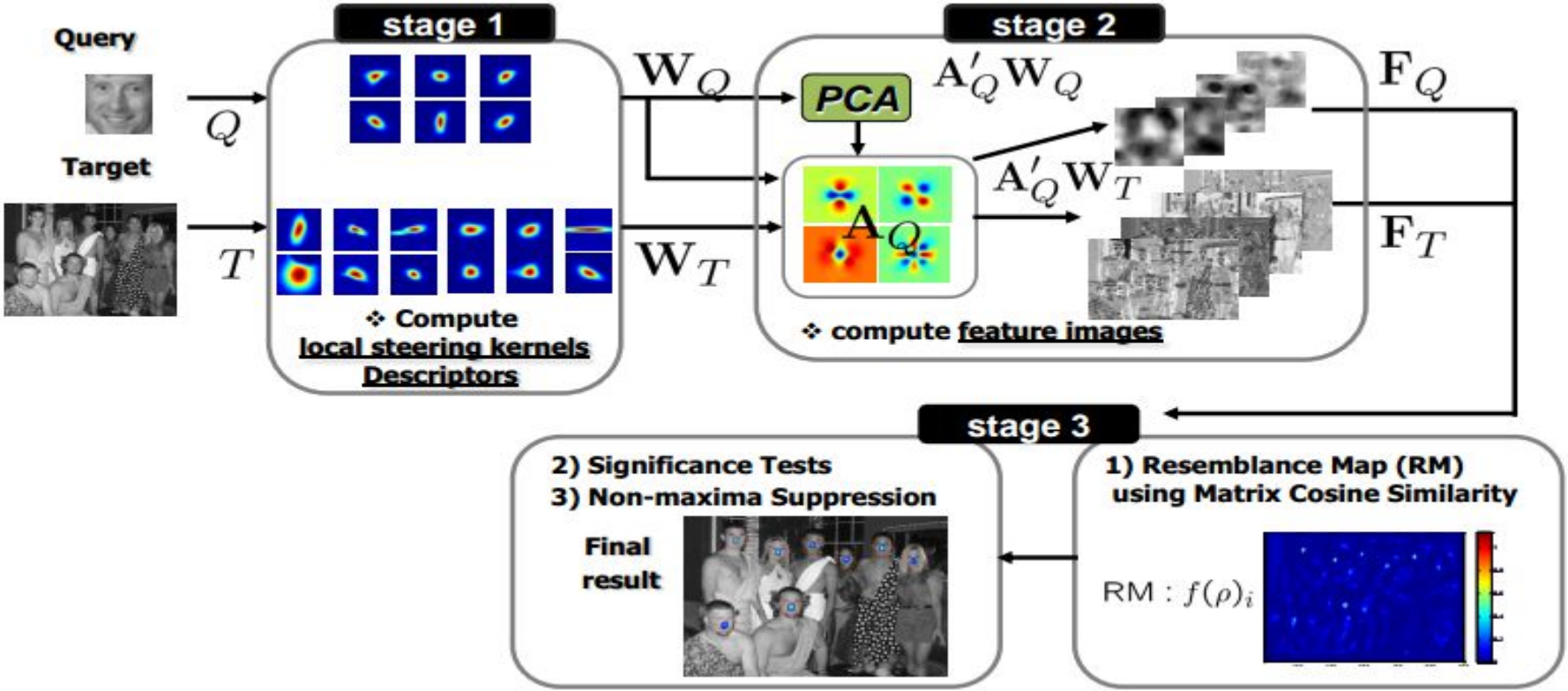
Even if they’ve never seen it before.

Overview of One Shot Detection Scheme

Query

Output

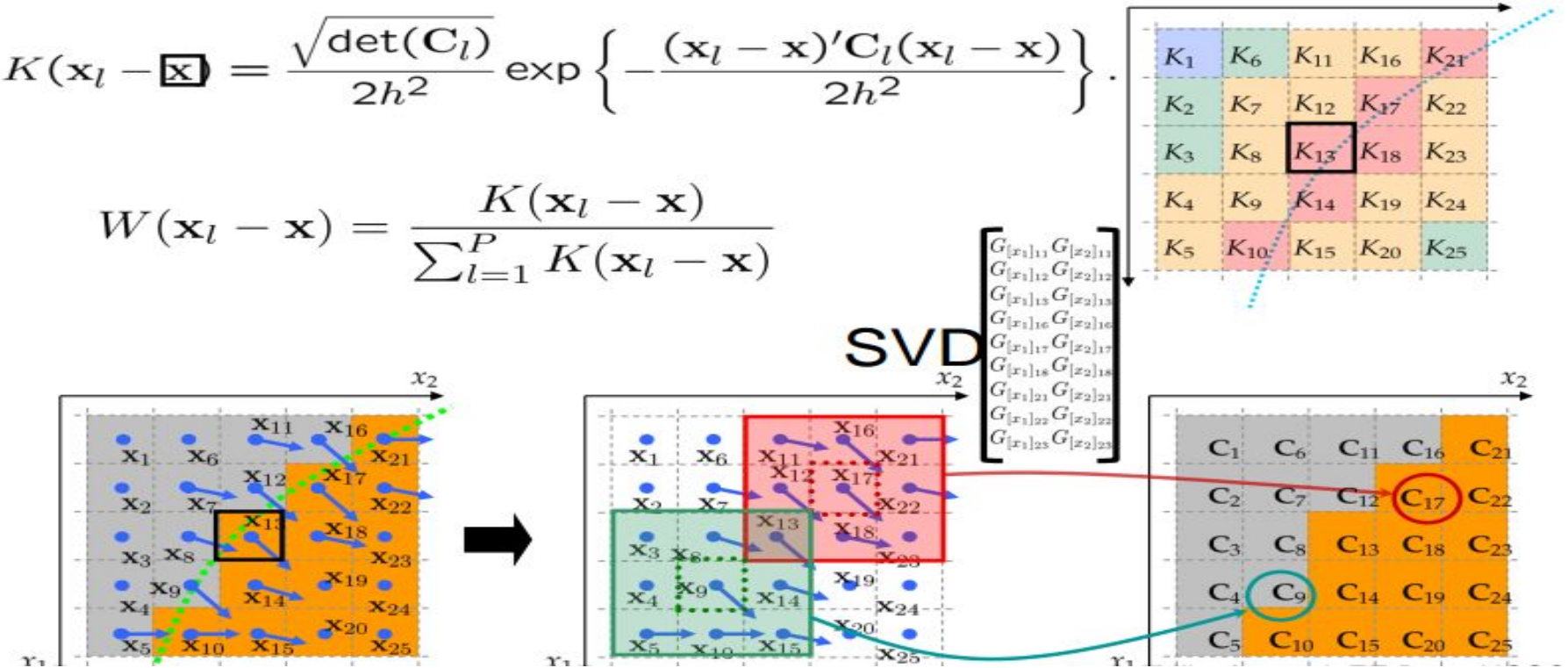
One Shot Detection System Pipeline



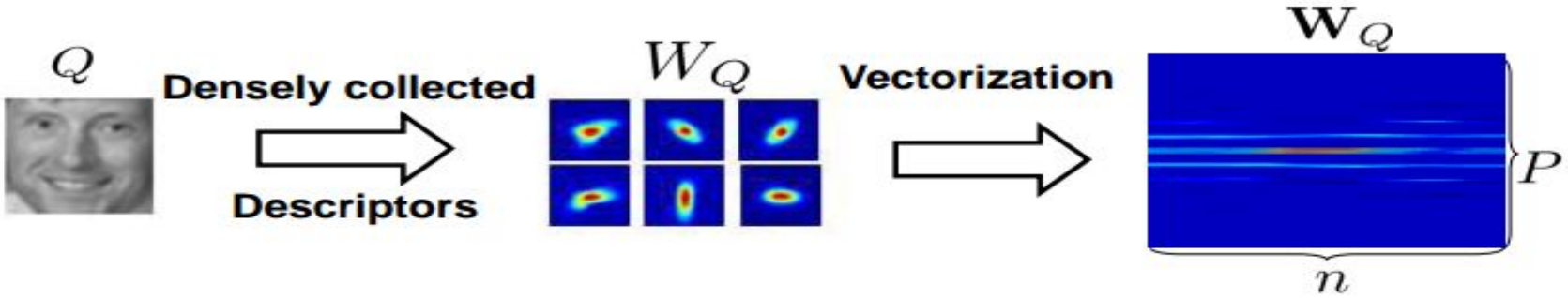
Stage 1: Calculation of Local Descriptors

$$K(\mathbf{x}_l - \mathbf{x}) = \frac{\sqrt{\det(\mathbf{C}_l)}}{2h^2} \exp\left\{-\frac{(\mathbf{x}_l - \mathbf{x})'\mathbf{C}_l(\mathbf{x}_l - \mathbf{x})}{2h^2}\right\}.$$

$$W(\mathbf{x}_l - \mathbf{x}) = \frac{K(\mathbf{x}_l - \mathbf{x})}{\sum_{l=1}^{P} K(\mathbf{x}_l - \mathbf{x})}$$



Stage 2: Feature Extraction from Descriptors



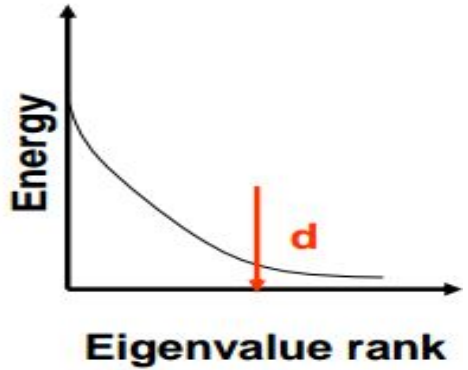
Apply PCA to $\mathbf{W}_Q$ for dimensionality reduction

→ Retain the d largest principal components $\mathbf{A}_Q \in \mathbb{R}^{P \times d}$

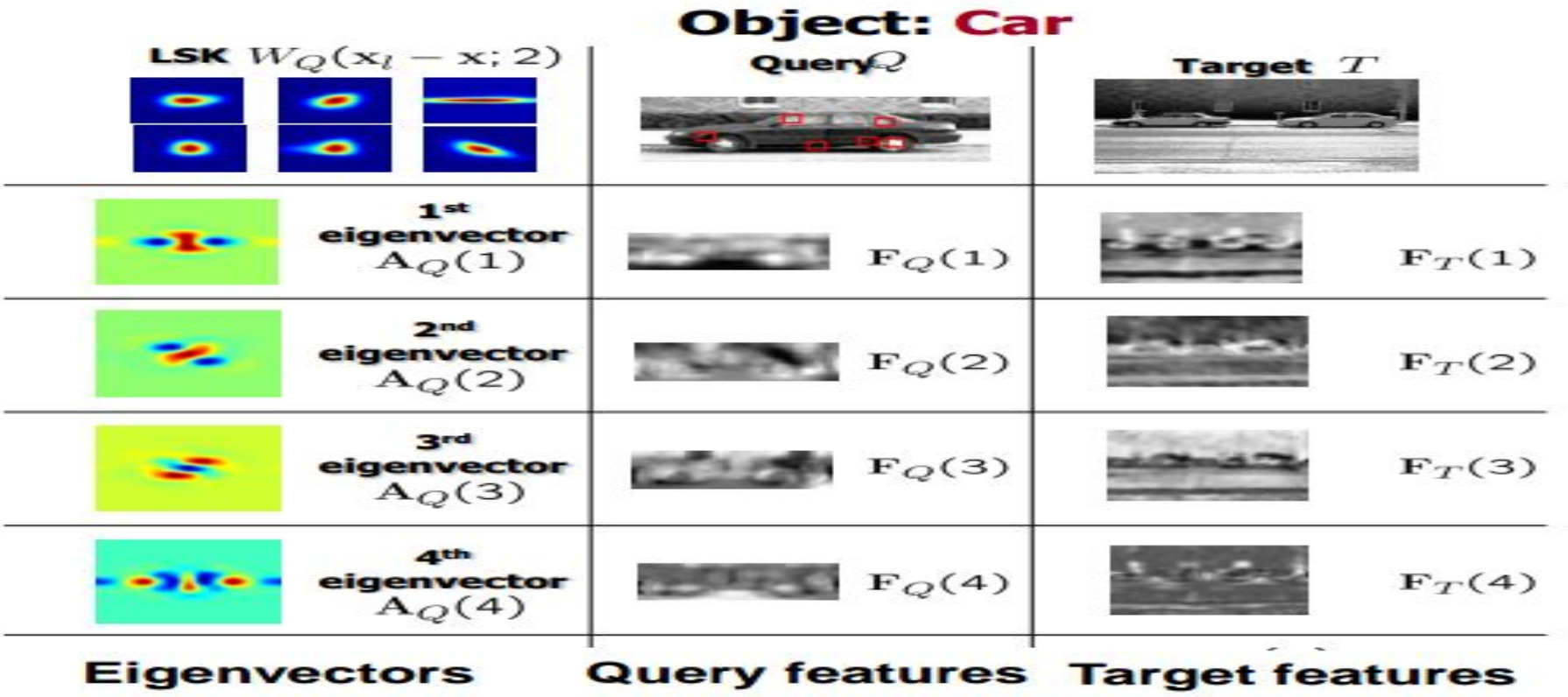
→ Project $\mathbf{W}_Q$ and $\mathbf{W}_T$ onto $\mathbf{A}_Q$

$$\mathbf{F}_Q = [\underline{\mathbf{f}}_Q^1, \cdots, \underline{\mathbf{f}}_Q^n] = \mathbf{A}_Q^{'} \mathbf{W}_Q$$

$$\mathbf{F}_T = [\underline{\mathbf{f}}_T^1, \cdots, \underline{\mathbf{f}}_T^{n_T}] = \mathbf{A}_Q^{'} \mathbf{W}_T$$

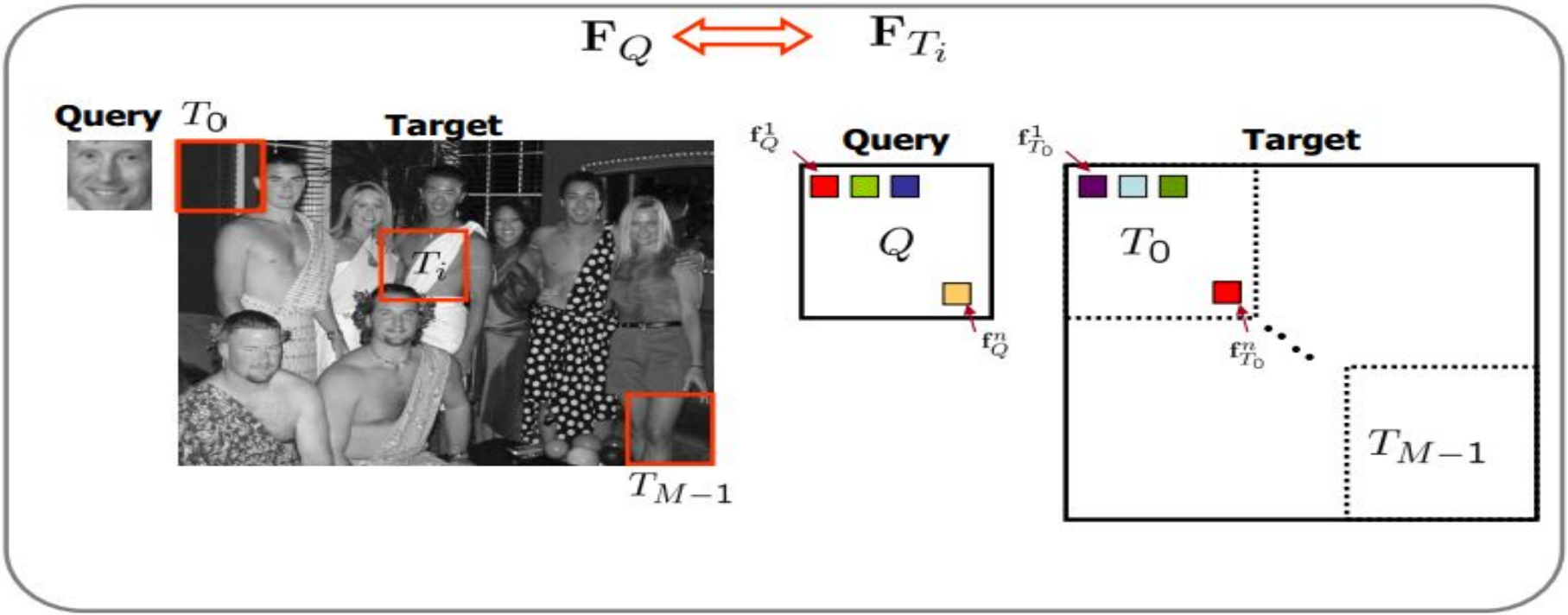


Stage 2: Salient features after PCA



Stage 3: Finding similarity between features

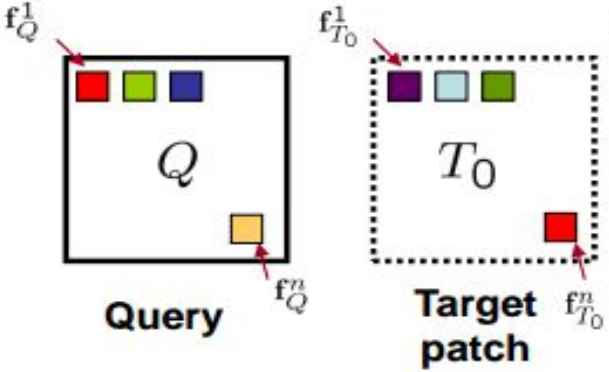
Target image is divided into a set of overlapping patches



Stage 3: Matrix Cosine Similarity

What about a set of vectors? **Matrix Cosine Similarity**

→ Frobenius Inner product between normalized matrices



$$\rho(\mathbf{F}_Q, \mathbf{F}_{T_i}) = <\overline{\mathbf{F}}_Q, \overline{\mathbf{F}}_{T_i}>_F = \text{trace}\left(\frac{\mathbf{F}_Q' \mathbf{F}_{T_i}}{\|\mathbf{F}_Q\|_F \|\mathbf{F}_{T_i}\|_F}\right) \in [-1, 1],$$

$$= \sum_{\ell=1}^{n} \frac{\mathbf{f}_Q^{\ell\,\prime} \mathbf{f}_{T_i}^{\ell}}{\|\mathbf{F}_Q\|_F \|\mathbf{F}_{T_i}\|_F},$$

$$= \sum_{\ell=1}^{n} \rho(\mathbf{f}_Q^{\ell}, \mathbf{f}_{T_i}^{\ell}) \frac{\|\mathbf{f}_Q^{\ell}\| \|\mathbf{f}_{T_i}^{\ell}\|}{\|\mathbf{F}_Q\|_F \|\mathbf{F}_{T_i}\|_F}.$$

A weighted sum of the column-wise vector cosine similarities

$$= \rho(\text{colstack}(\mathbf{F}_Q), \text{colstack}(\mathbf{F}_{T_i}))$$

Stage 3: Generate Resemblance Map

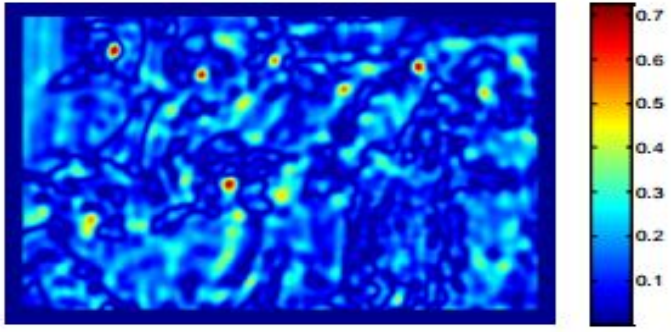
Resemblance Map (RM)

$$\text{RM} : f(\rho_i) = \frac{\rho_i^2}{1 - \rho_i^2}$$

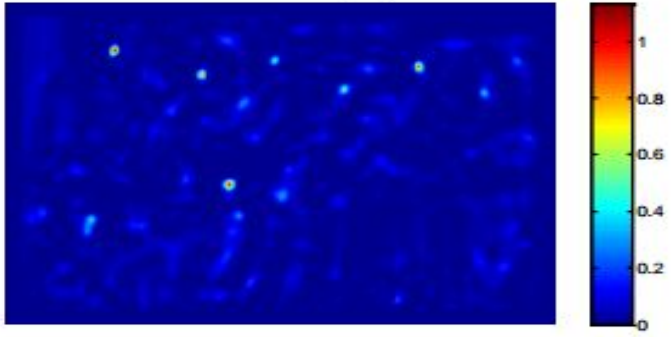
Describes the proportion of variance in common between two features

Lawley-Hotelling Trace statistic

$\text{RM} : |\rho_i|$



$\text{RM} : \frac{\rho_i^2}{1 - \rho_i^2}$



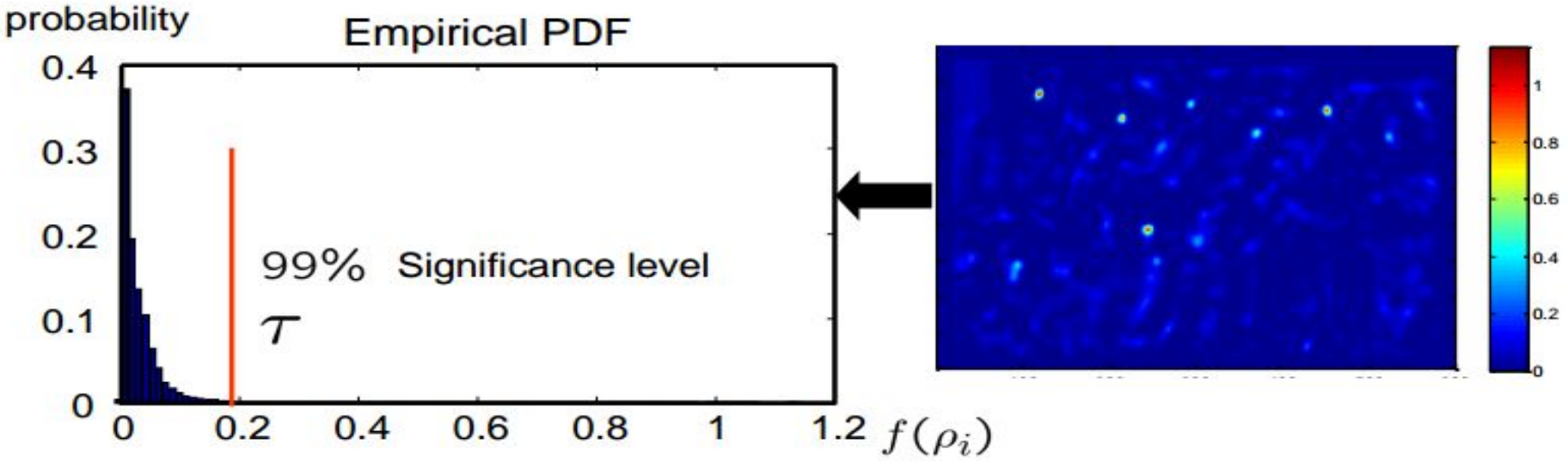
Stage 3: Significance Tests & Non Maxima Suppression

1. Is any sufficiently similar object present?

$$\max f(\rho_i) > \tau_0$$

i.e., $\tau_o = 0.96$ so that ~ 50 % of variance in common

2. How many objects of interest are present?



Experimental Results

query

target

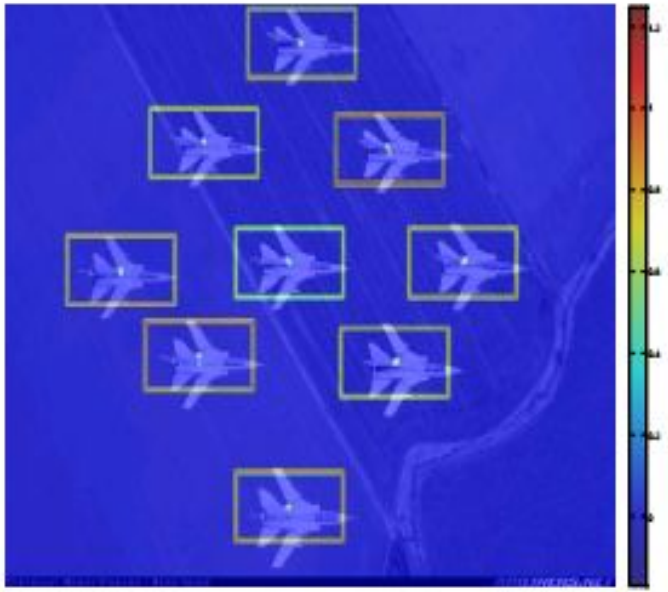
output

Experimental Results

query

target

output



Limitations & Future Work

1. **Make algorithm scalable for image and (video) retrieval**
2. **Increase accuracy by incorporating “context”**
3. **Detect /recognize objects of interest in general degraded data without explicit restoration**
4. **Extension to large-scale data set requires a significant improvement of the computational complexity of the proposed method.**

THANK YOU!