Sparse Kernel Learning for Image Annotation

School of informatics

Sean Moran†, Victor Lavrenko
† sean.moran@ed.ac.uk

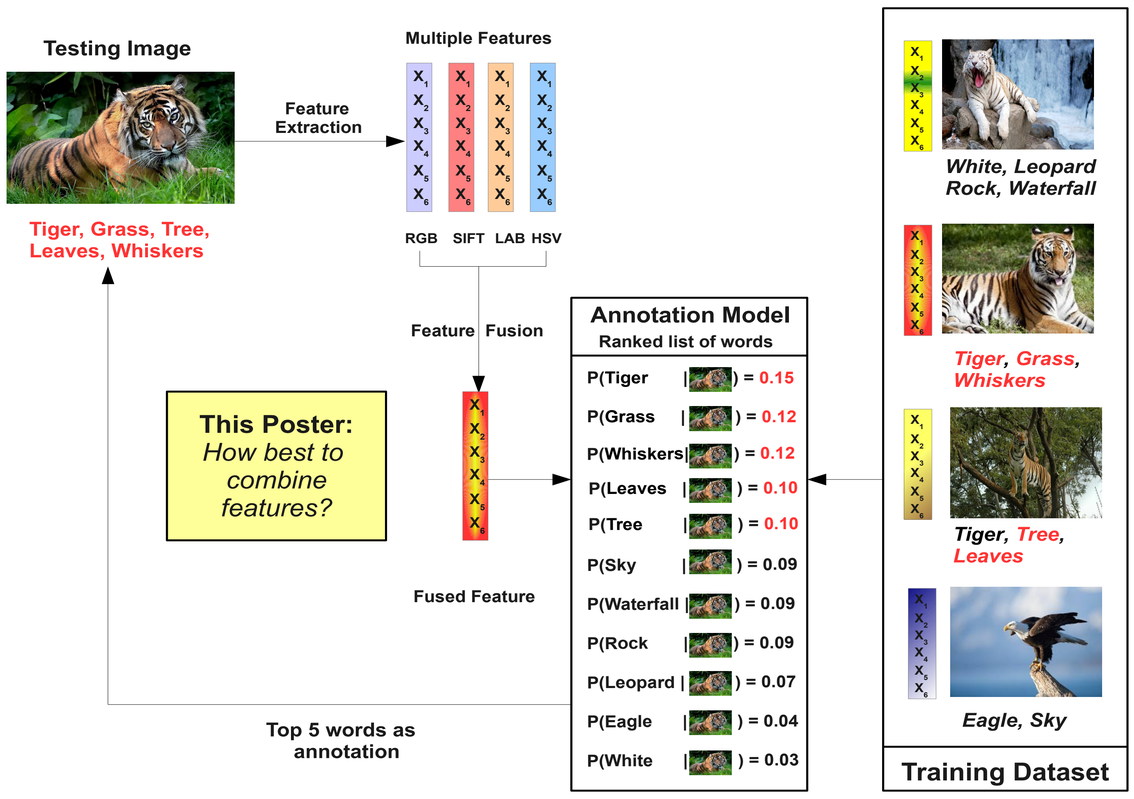
The University of Edinburgh

Research Question

- How do we exploit multiple features for image annotation?

Introduction

- **Problem:** Assigning one or more keywords to describe an image.
- **Probabilistic generative modelling approach:**
  - Compute the conditional probability of word given image $P(w|I)$.
  - Take the 5 words with highest $P(w|I)$ as the image annotation.



- **Advantages:**
  - Permits image search based on natural language keywords.

Continuous Relevance Model (CRM): Lavrenko et al. '03

- $P(\mathbf{w}, \mathbf{f})$: joint expectation of words $\mathbf{w}$ and image features $\mathbf{f}$ defined by images J in the training set T:

$$P(\mathbf{w}, \mathbf{f}) = \sum_{J \in T} P(J) \prod_{i=1}^{K} P(w_i|J) \prod_{i=1}^{M} P(\vec{f}_i|J) \quad (1)$$

- $P(w_i|J)$ is modelled using a Dirichlet prior:

$$P(w_i|J) = \frac{\mu p_v + N_{v,J}}{\mu + \sum_{v'} N_{v',J}} \quad (2)$$

- $N_{v,J}$: number of times the word v appears in annotation of training image J, p_v: relative frequency of word v, μ: smoothing parameter.

- $P(\vec{f}_j|J)$ is modelled with a kernel-based density estimator:

$$P(\vec{f}_i|J) = \frac{1}{R} \sum_{j=1}^{R} P(\vec{f}_i|\vec{f}_j) \quad (3)$$

- Each region $j = 1 \ldots R$ instantiates a Gaussian kernel, bandwidth β:

$$P(\vec{f}_i|\vec{f}_j) = \frac{1}{\sqrt{2^d \pi^d \beta}} exp\left\{\frac{-||\vec{f}_i - \vec{f}_j||^2}{\beta}\right\} \quad (4)$$

Sparse Kernel Learning CRM (SKL-CRM)

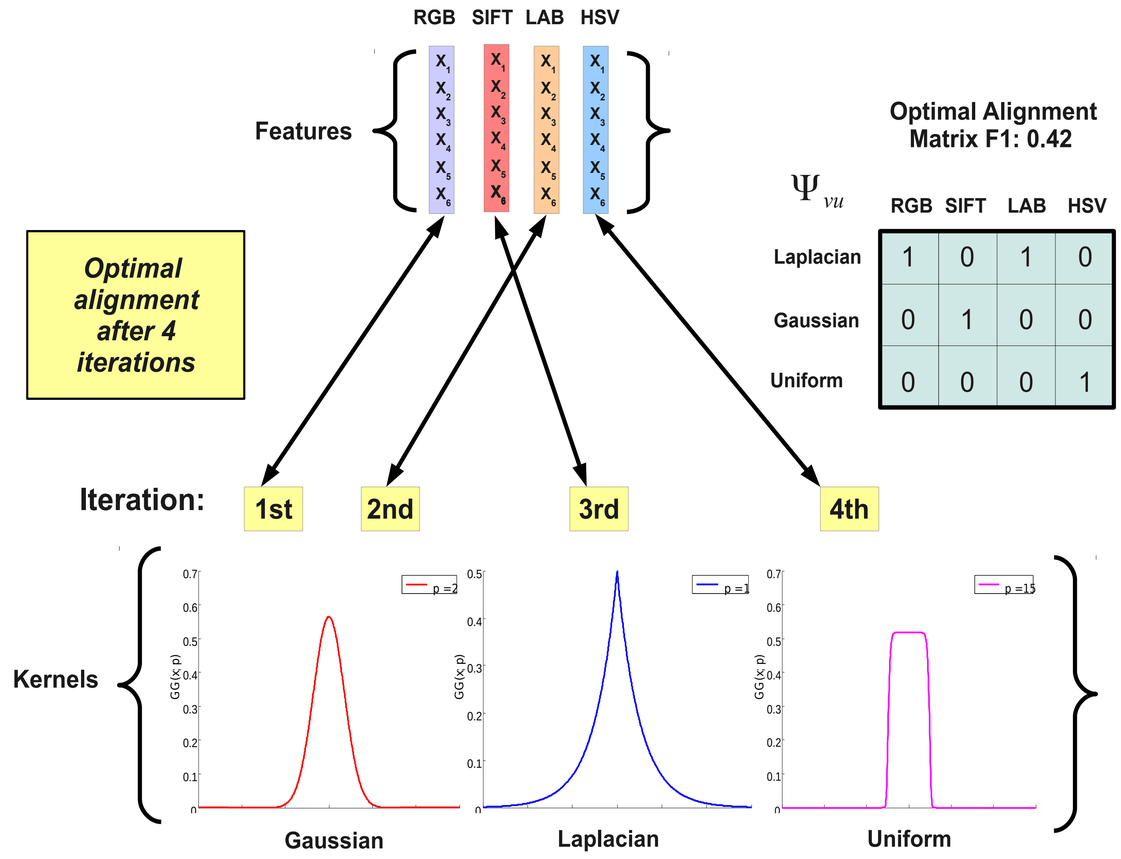
- Extend the CRM to M feature types (e.g. SIFT, HSV, RGB ...):

$$P(I|J) = \prod_{i=1}^{M} \sum_{j=1}^{R} exp\left\{-\frac{1}{\beta} \sum_{u,v} \mathbf{\Psi}_{u,v} k^v(\vec{f}_i^u, \vec{f}_j^u)\right\} \quad (5)$$

- SKL-CRM learns $\mathbf{\Psi}_{u,v}$: an alignment matrix mapping kernel k^v (e.g. Gaussian) to a feature type u (e.g. SIFT)

Greedy Kernel-Feature Alignment Algorithm

- Greedily solve for the kernel-feature alignment matrix $\mathbf{\Psi}_{u,v}$
- At each iteration add the kernel-feature that maximises F_1



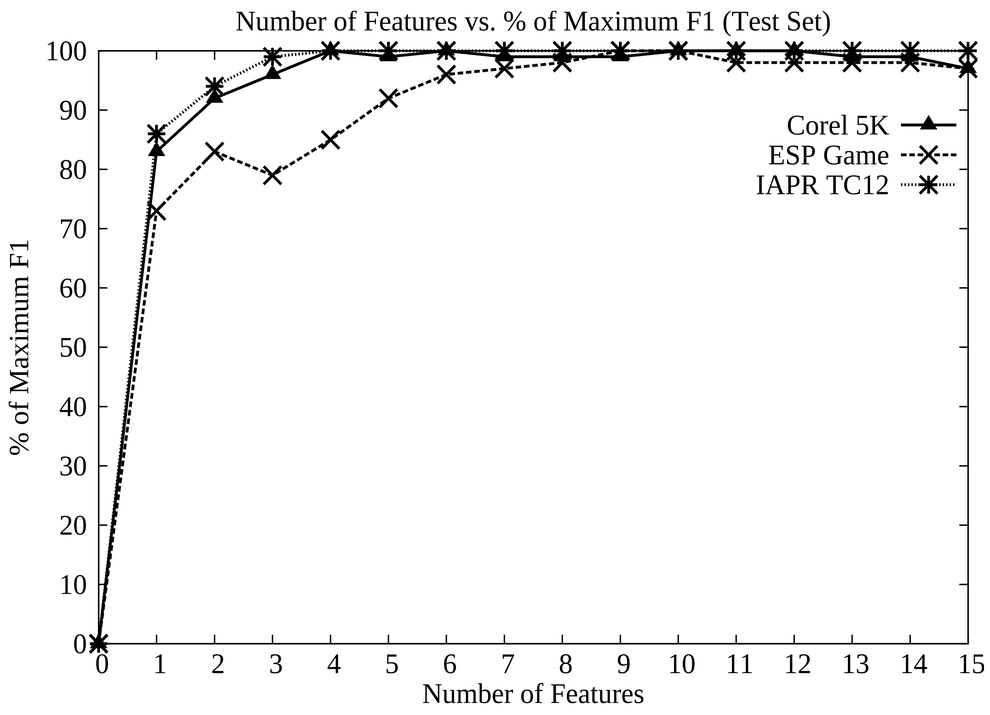
- SKL-CRM aligns to the generalised Gaussian, χ^2, Hellinger and Multinomial kernels

Quantitative Results

- Mean per word precision (P), recall (R), Number of words > 0 (N+) and F_1 measure:

Dataset	Corel 5K				IAPR TC12				ESP Game			
	R	P	F_1	N+	R	P	F_1	N+	P	R	F_1	N+
CRM	19	16	17	107	–	–	–	–	–	–	–	–
JEC	32	27	29	139	29	28	28	250	25	22	23	224
RF-opt	40	29	34	157	31	44	36	253	26	41	32	235
GS	42	33	37	160	29	32	30	252	–	–	–	–
KSVM-VT	42	32	36	179	29	47	36	268	32	33	33	259
Tagprop	42	33	37	160	35	46	40	266	27	39	32	239
SKL-CRM	46	39	42	184	32	47	38	274	26	41	32	248

- Rapid convergence to a *sparse subset* of the available features:



Summary of Key Findings

- Better to choose kernels based on the data than opt for default assignment advocated in literature.
- Only a small number of carefully chosen features are required for the best annotation performance.
- **See our ICMR'14 paper for further information and results.**