Learning Clustered Sub-spaces for Sketch-based Image Retrieval

Koustav Ghosal, Ameya Prabhu, Riddhiman Dasgupta, Anoop Namboodiri

Center for Visual Information Technology

IIIT Hyderabad, India

Content Based Image Retrieval

Different Form Of Queries :

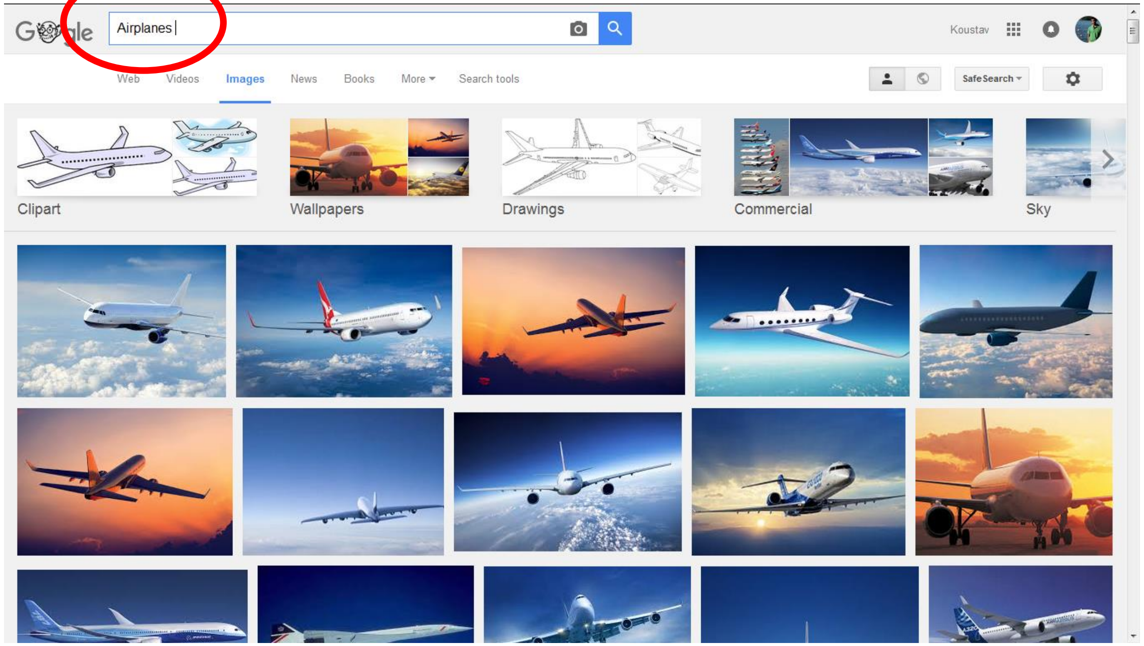
Text Example Sketch

" Classical Problem in Computer Vision : Divided into 3 categories as mentioned above. "

Content Based Image Retrieval

Different Form Of Queries :

Text Example Sketch

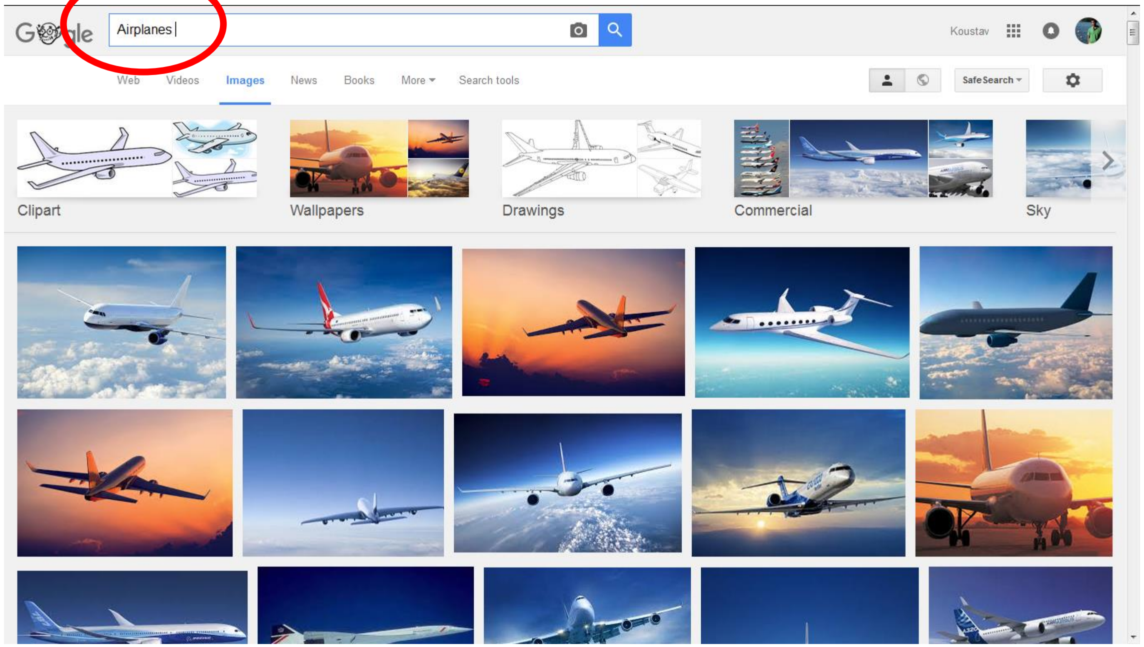


Most popular approach.

Content Based Image Retrieval

Different Form Of Queries :

Text Example Sketch



Search is based on metadata (hash-tags, comments) and NOT the actual content within the image

Content Based Image Retrieval

Different Form Of Queries :

Text **Example** Sketch

Works well when examples are available …

" images.google.com "

Content Based Image Retrieval

Different Form Of Queries :

Text **Example** Sketch

But examples are not always available.

Content Based Image Retrieval

Different Form Of Queries :

Text Example Sketch

What about these objects ?

Content Based Image Retrieval

Different Form Of Queries :

Text Example Sketch

"A table lamp with a blue base, blue shade and a black neck."

"A navy-blue t-shirt with USA written on it."

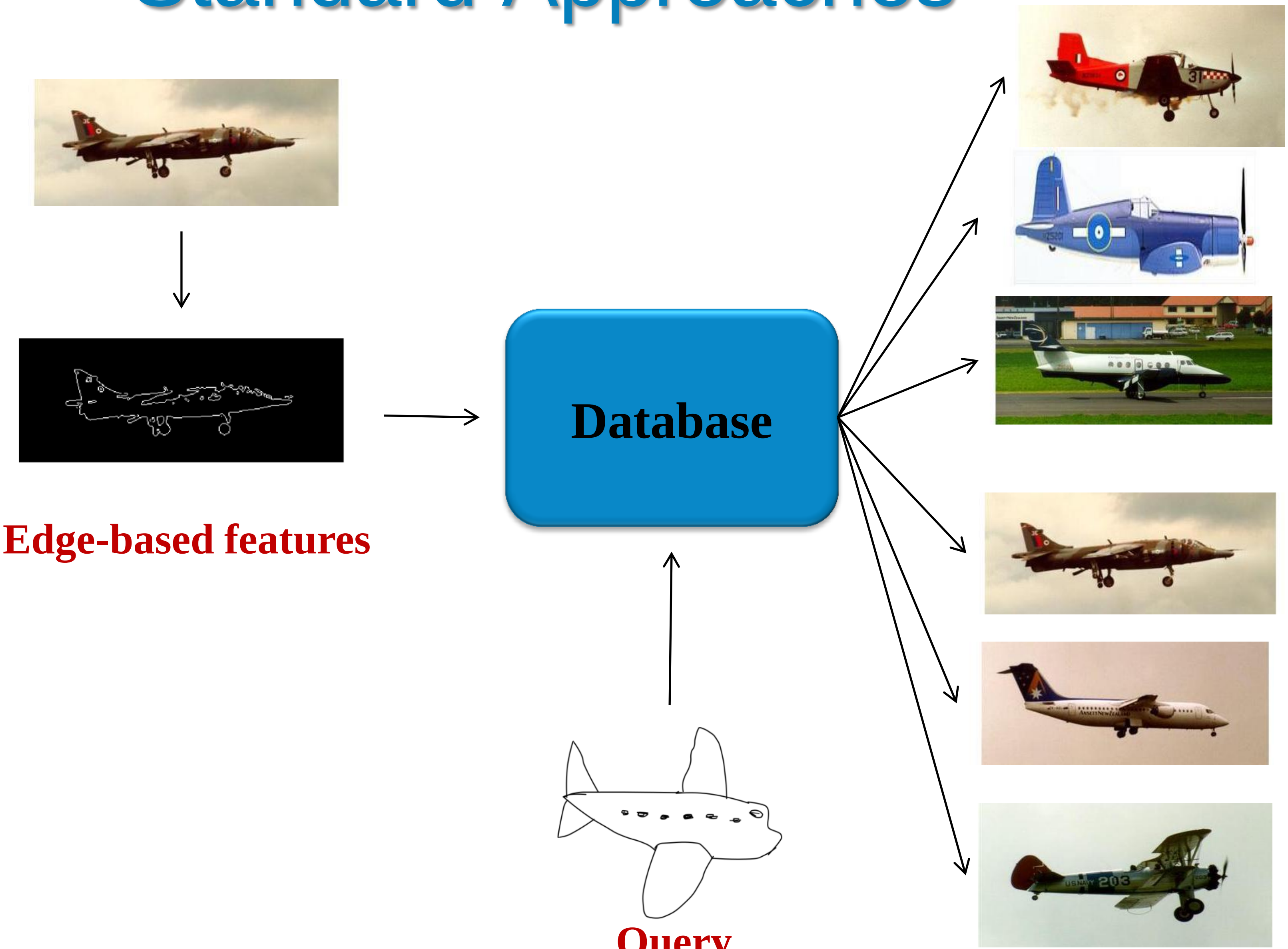
"A black travel purse with two base pockets and a main pocket."

Content Based Image Retrieval

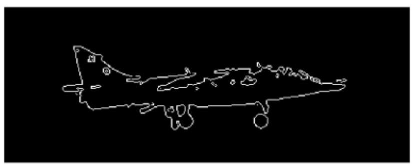
Different Form Of Queries :

Text Example **Sketch**

Standard Approaches

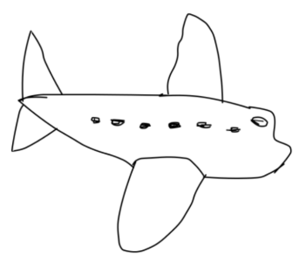


Standard Approaches

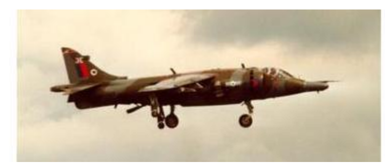


Edge-based features

Database



Query



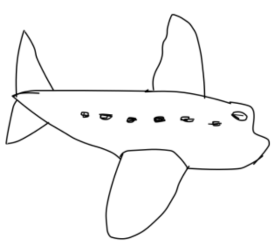
- **Information Loss**
- **Most of us are not faithful artists**

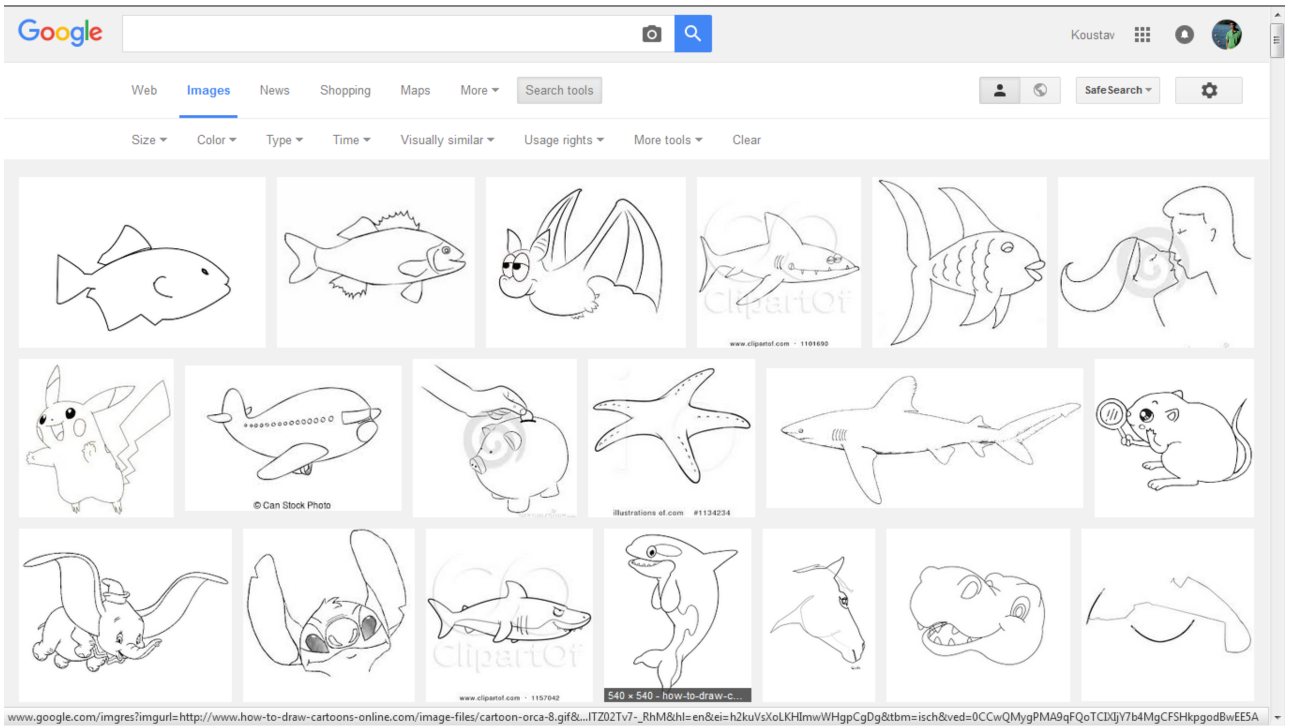
Sparsity : Information Loss

Yang *et al*, Zeiler *et al*

Results

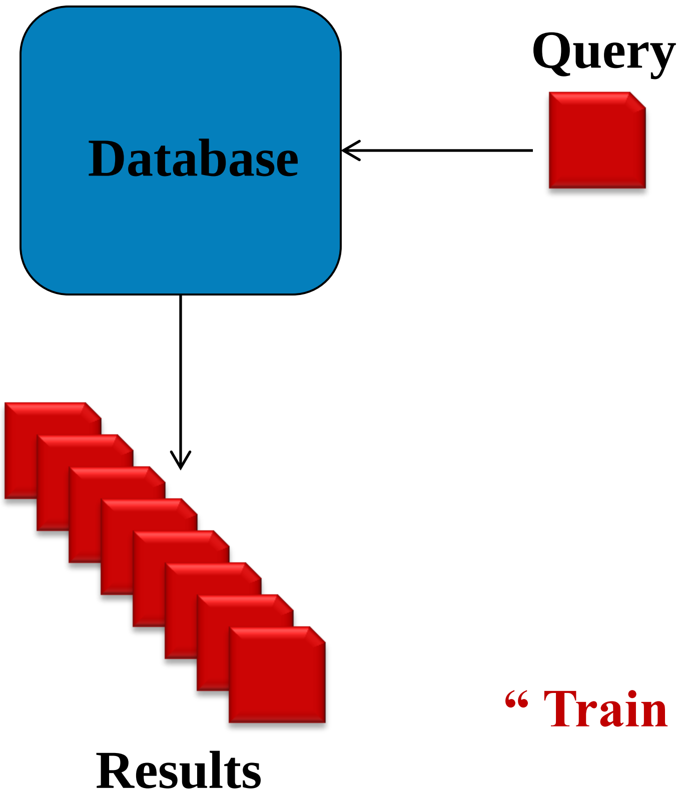
Query





Two different modalities shouldn't be compared directly

Cross-Modal Problem



Uni-Modal

Text Text

Image Image

Cross-Modal

Image Text

Image Sketch

Motivation

“ Train the system both with images and sketches ”

Our Model

N dimensions

car bird

bag ball

Images

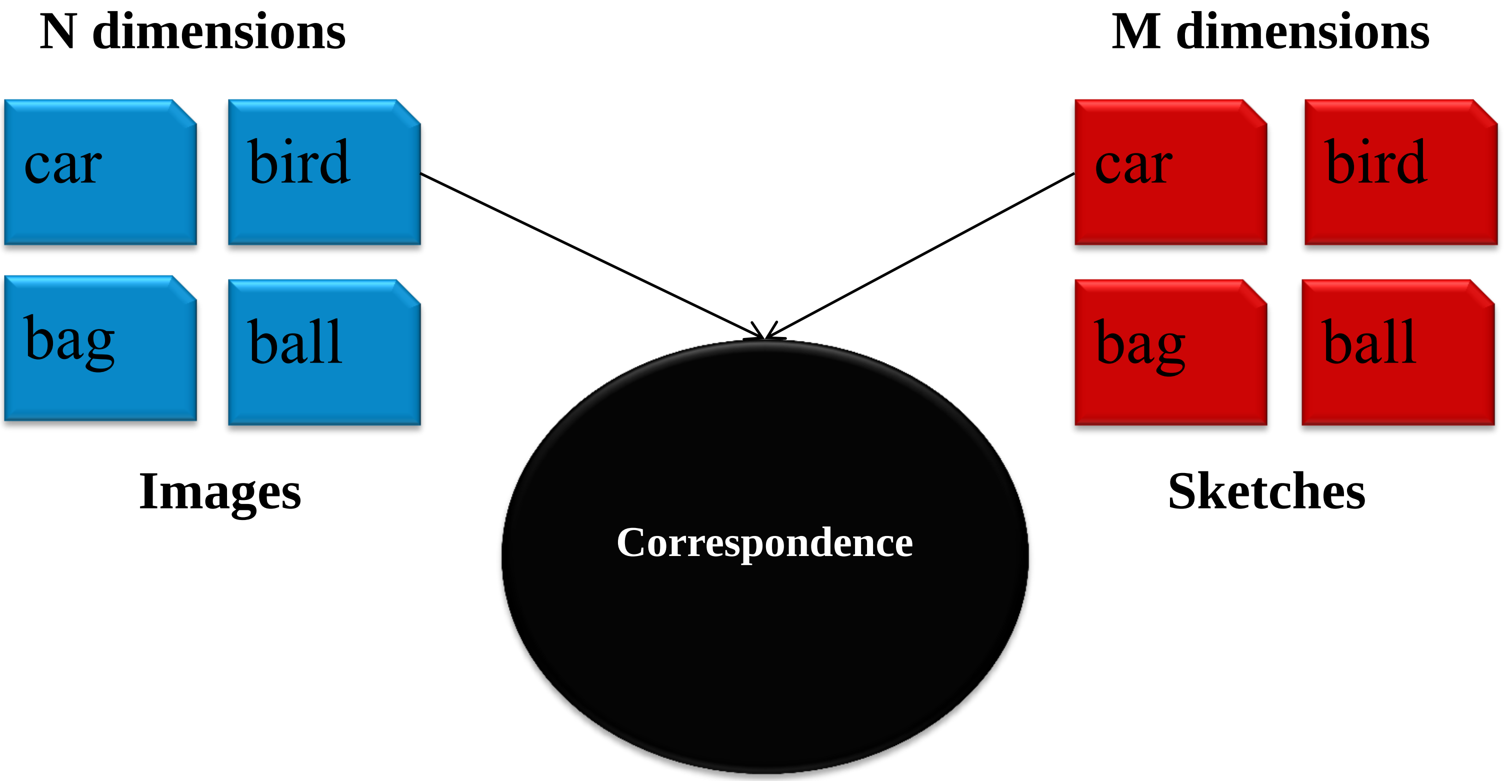
M dimensions

car bird

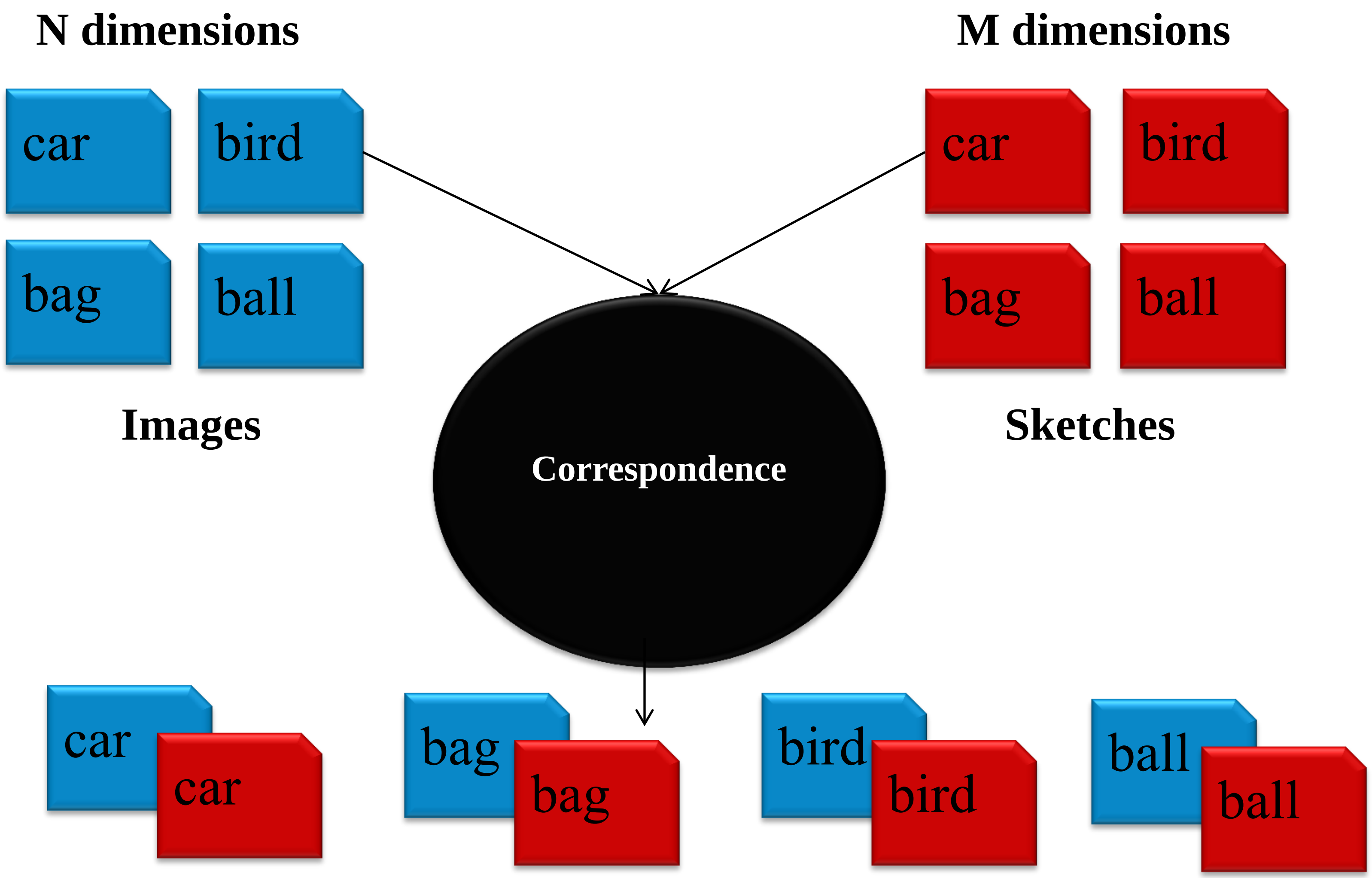
bag ball

Sketches

Our Model

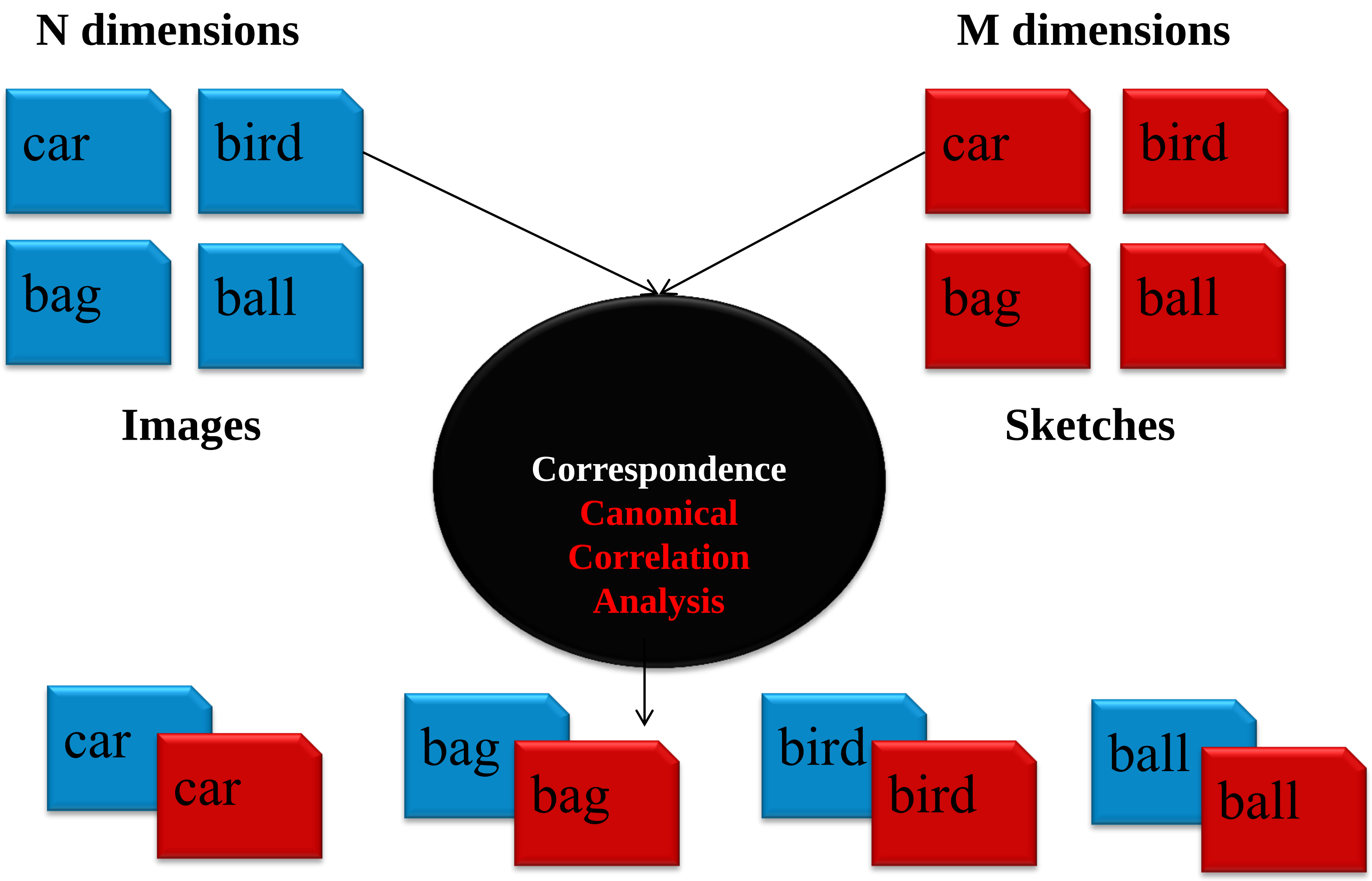

N dimensions
car
bird
bag
ball
Images
M dimensions
car
bird
bag
ball
Sketches
Correspondence

Our Model



N > K dimensions < M

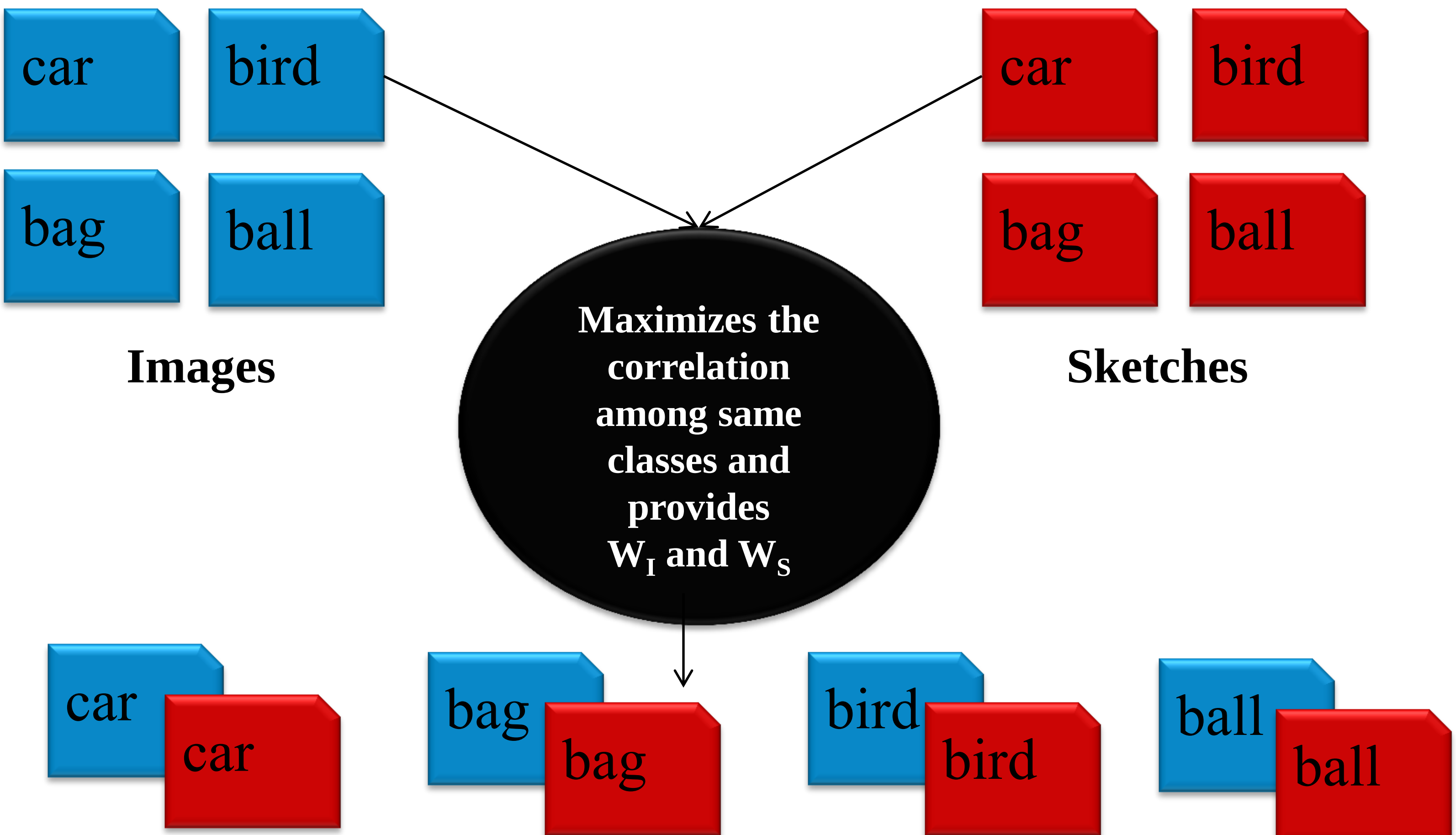
Our Model



N > K dimensions < M

CCA

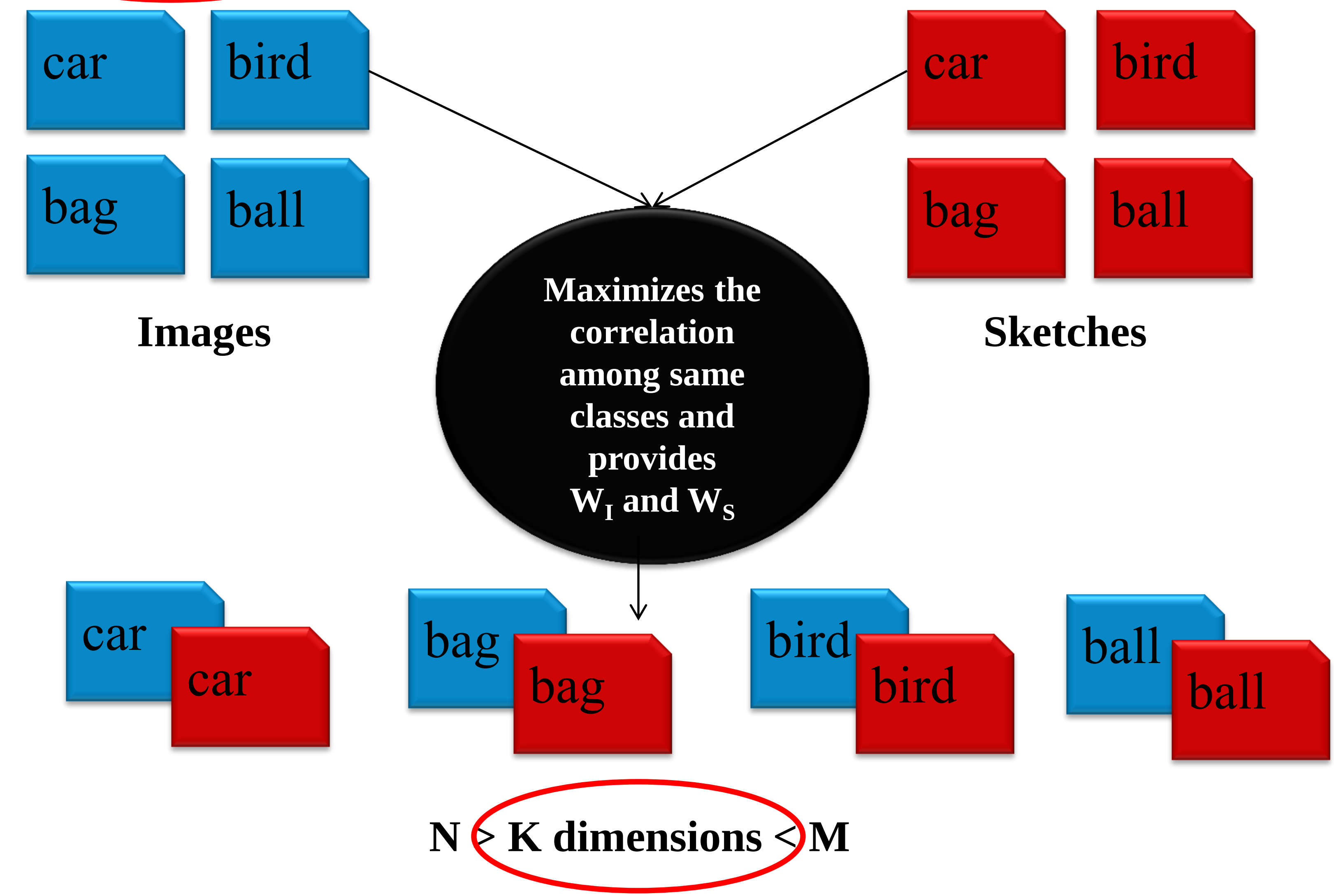
N dimensions Hardoon *et al* **M dimensions**



N > K dimensions < M

CCA

Hardoon *et al*



Canonical Correlation Analysis

Hardoon *et al*

$$X = (X_1, X_2, ...X_n)$$

$$Y = (Y_1, Y_2, ...Y_n)$$

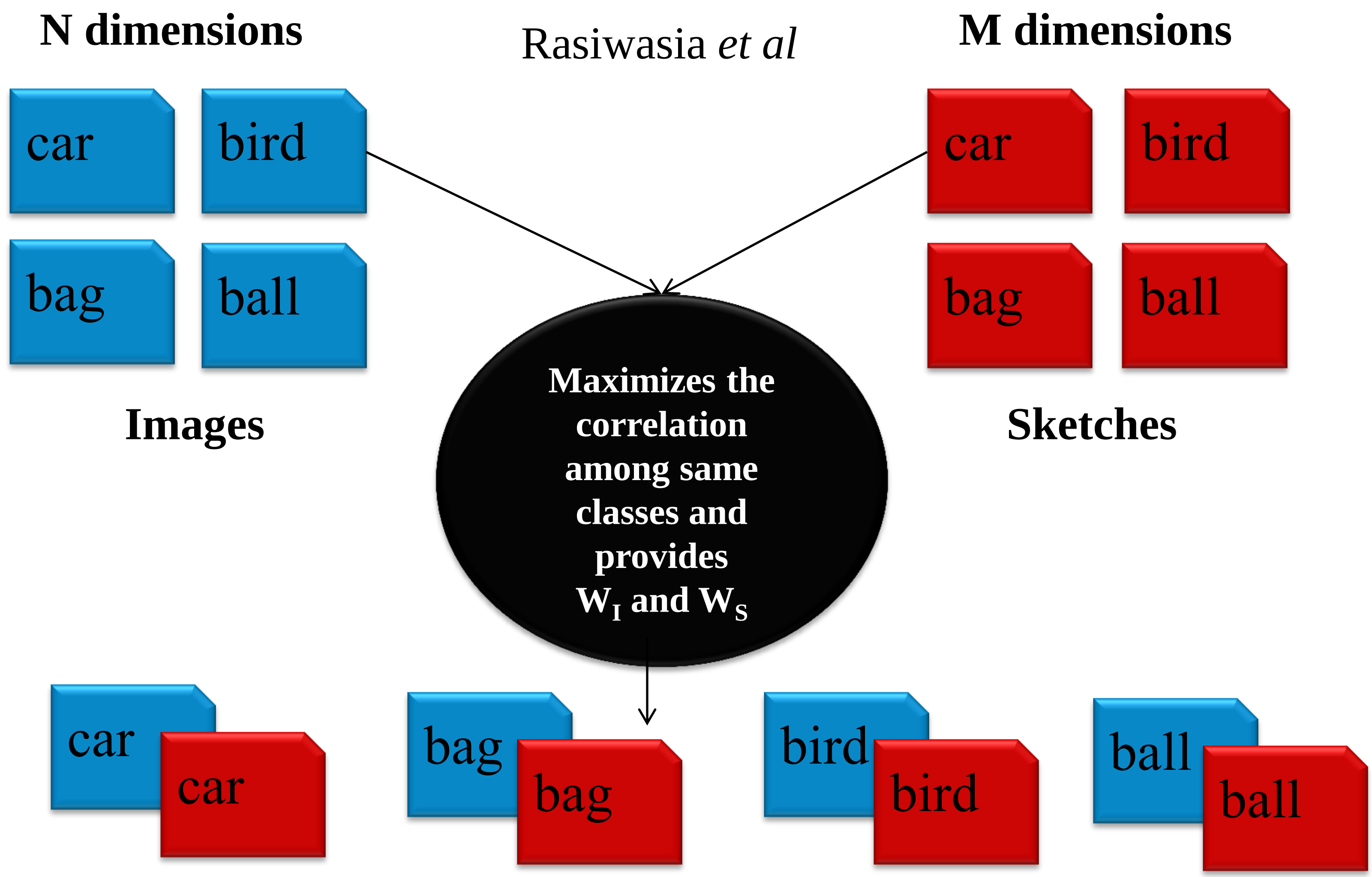
$$S_x = < W_x, X >, S_y = < W_y, Y >$$

$$\rho = \frac{W_x' \Sigma_{XY} W_y}{\sqrt{W_x' \Sigma_{XX} W_x} \sqrt{W_y' \Sigma_{YY} W_y}}.$$

Solved as an Eigen Value problem and we find

W_x and W_y which maximizes "ρ"

Cluster CCA



N > K dimensions < M

Cluster CCA

Rasiwasia *et al*

The covariance matrix is calculated for each class

$$\Sigma_{IS} = \frac{1}{M} \sum_{c=1}^{C} \sum_{j=1}^{|I^c|} \sum_{k=1}^{|S^c|} I_j^c S_k^{c'}$$

$$\Sigma_{II} = \frac{1}{M} \sum_{c=1}^{C} \sum_{j=1}^{|I^c|} |S^c| I_j^c I_j^{c'}$$

$$\Sigma_{SS} = \frac{1}{M} \sum_{c=1}^{C} \sum_{k=1}^{|S^c|} |I^c| S_k^c S_k^{c'}$$

Our Formulation

Standard CCA

$$\rho = \frac{W_x' \Sigma_{XY} W_y}{\sqrt{W_x' \Sigma_{XX} W_x} \sqrt{W_y' \Sigma_{YY} W_y}}.$$

Our formulation

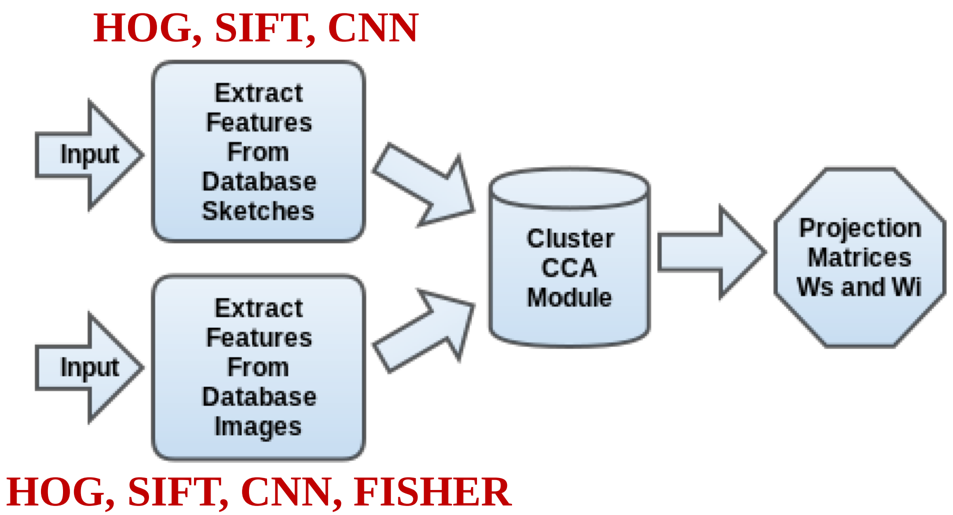
$$\rho = \frac{W_S' \Sigma_{IS} W_I}{\sqrt{W_S' \Sigma_{SS} W_S} \sqrt{W_I' \Sigma_{II} W_I}}.$$

CCA vs Cluster CCA

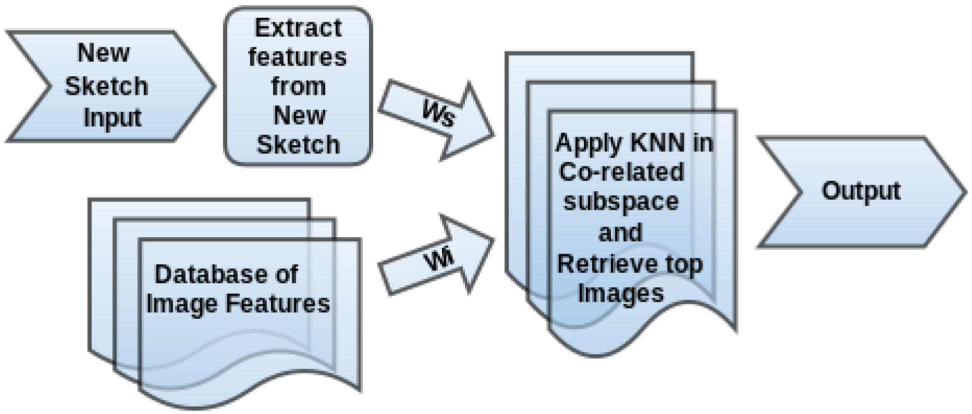
Element-wise correspondence	Class-wise correspondence
Covariance matrix computation $\mathcal{O}(n)$	Covariance matrix computation $C * \mathcal{O}(n^2)$

Pipeline

Training



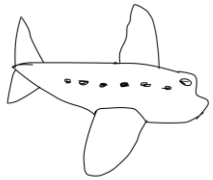
Testing



Datasets

Sketches

TU-Berlin Dataset (Eitz et al)



Images

Caltech 256, Pascal VOC 2007

105 common classes for Caltech 256

19 Common classes for Pascal VOC 2007

80 samples from each class.

Results

MAP values for Image-Sketch Feature Combinations

Dataset	SIFT-SIFT	SIFT-HOG	SIFT-Fisher	HOG-SIFT	HOG-HOG	HOG-Fisher	CNN-CNN
Caltech	0.06	0.03	**0.20**	0.14	0.02	0.01	**0.20**
Pascal	0.13	0.12	0.05	**0.18**	0.09	0.06	0.06

Results

Raw features vs cluster CCA

MAP values

Dataset	Features	Before CCA	After CCA
Caltech	SIFT-Fisher	0.01	**0.20**
Caltech	CNN-CNN	0.01	**0.20**
Pascal	HOG-SIFT	0.01	**0.18**
Pascal	SIFT-SIFT	0.06	**0.13**

Summary

- Content based retrieval
- Close to human perception.
- Efficiency and Usability
- More information.

Suggestions and Questions ?

Thank you

Appendix

CCA

$$S_{x,\mathbf{w}_x} = (\langle \mathbf{w}_x, \mathbf{x}_1 \rangle, \ldots, \langle \mathbf{w}_x, \mathbf{x}_n \rangle)$$

with the corresponding values of the new $\mathbf{y}$ co-ordinate being

$$S_{y,\mathbf{w}_y} = (\langle \mathbf{w}_y, \mathbf{y}_1 \rangle, \ldots, \langle \mathbf{w}_y, \mathbf{y}_n \rangle)$$

The first stage of canonical correlation is to choose $\mathbf{w}_x$ and $\mathbf{w}_y$ to maximise the correlation between the two vectors. In other words the function to be maximised is

$$\begin{aligned} \rho &= \max_{\mathbf{w}_x, \mathbf{w}_y} corr(S_x \mathbf{w}_x, S_y \mathbf{w}_y) \\ &= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\langle S_x \mathbf{w}_x, S_y \mathbf{w}_y \rangle}{\|S_x \mathbf{w}_x\| \|S_y \mathbf{w}_y\|} \end{aligned}$$

If we use $\hat{\mathbb{E}}[f(\mathbf{x}, \mathbf{y})]$ to denote the empirical expectation of the function $f(\mathbf{x}, \mathbf{y})$, were

$$\hat{\mathbb{E}}[f(\mathbf{x}, \mathbf{y})] = \frac{1}{m} \sum_{i=1}^{m} f(\mathbf{x}_i, \mathbf{y}_i)$$

Appendix

CCA

Algorithm 4

we can rewrite the correlation expression as

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\hat{\mathbb{E}}[\langle \mathbf{w}_x, \mathbf{x}\rangle \langle \mathbf{w}_y, \mathbf{y}\rangle]}{\sqrt{\hat{\mathbb{E}}[\langle \mathbf{w}_x, \mathbf{x}\rangle^2]\hat{\mathbb{E}}[\langle \mathbf{w}_x, \mathbf{x}\rangle^2]}}$$

$$= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\hat{\mathbb{E}}[\mathbf{w}_x' \mathbf{x} \mathbf{y}' \mathbf{w}_y]}{\sqrt{\hat{\mathbb{E}}[\mathbf{w}_x' \mathbf{x}\mathbf{x}' \mathbf{w}_x]\hat{\mathbb{E}}[\mathbf{w}_y' \mathbf{y}\mathbf{y}' \mathbf{w}_y]}}$$

follows that

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' \hat{\mathbb{E}}[\mathbf{x}\mathbf{y}'] \mathbf{w}_y}{\sqrt{\mathbf{w}_x' \hat{\mathbb{E}}[\mathbf{x}\mathbf{x}'] \mathbf{w}_x \mathbf{w}_y' \hat{\mathbb{E}}[\mathbf{y}\mathbf{y}'] \mathbf{w}_y}}.$$

Where we use A' to denote the transpose of a vector or matrix A.
Now observe that the covariance matrix of $(\mathbf{x}, \mathbf{y})$ is

$$C(\mathbf{x}, \mathbf{y}) = \hat{\mathbb{E}}\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}'\right] = \begin{bmatrix} C_{\mathbf{xx}} & C_{\mathbf{xy}} \\ C_{\mathbf{yx}} & C_{\mathbf{yy}} \end{bmatrix} = C. \tag{2.1}$$

The total covariance matrix C is a block matrix where the within-sets covariance matrices are $C_{\mathbf{xx}}$ and $C_{\mathbf{yy}}$ and the between-sets covariance matrices are $C_{\mathbf{xy}} = C_{\mathbf{yx}}'$

Hence, we can rewrite the function ρ as

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}} \tag{2.2}$$

the maximum canonical correlation is the maximum of ρ with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$.

Appendix

CCA

3.1 Canonical Correlation Analysis

Observe that the solution of equation (2.2) is not affected by re-scaling $\mathbf{w}_x$ or $\mathbf{w}_y$ either together or independently, so that for example replacing $\mathbf{w}_x$ by $\alpha\mathbf{w}_x$ gives the quotient

$$\frac{\alpha \mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\alpha^2 \mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}} = \frac{\mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}}.$$

Since the choice of re-scaling is therefore arbitrary, the CCA optimisation problem formulated in equation (2.2) is equivalent to maximising the numerator

Algorithm 5

subject to

$$\begin{aligned} \mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x &= 1 \\ \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y &= 1. \end{aligned}$$

Appendix

CCA

The corresponding Lagrangian is

$$L(\lambda, \mathbf{w}_x, \mathbf{w}_y) = \mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y - \frac{\lambda_x}{2}(\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x - 1) - \frac{\lambda_y}{2}(\mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y - 1)$$

Taking derivatives in respect to $\mathbf{w}_x$ and $\mathbf{w}_y$ we obtain

$$\frac{\partial f}{\partial \mathbf{w}_x} = C_{\mathbf{xy}} \mathbf{w}_y - \lambda_x C_{\mathbf{xx}} \mathbf{w}_x = \mathbf{0} \quad (3.1)$$

$$\frac{\partial f}{\partial \mathbf{w}_y} = C_{\mathbf{yx}} \mathbf{w}_x - \lambda_y C_{\mathbf{yy}} \mathbf{w}_y = \mathbf{0}. \quad (3.2)$$

Subtracting $\mathbf{w}_y'$ times the second equation from $\mathbf{w}_x'$ times the first we have

$$\begin{aligned} 0 &= \mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y - \mathbf{w}_x' \lambda_x C_{\mathbf{xx}} \mathbf{w}_x - \mathbf{w}_y' C_{\mathbf{yx}} \mathbf{w}_x + \mathbf{w}_y' \lambda_y C_{\mathbf{yy}} \mathbf{w}_y \\ &= \lambda_y \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y - \lambda_x \mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x, \end{aligned}$$

which together with the constraints implies that $\lambda_y - \lambda_x = 0$, let $\lambda = \lambda_x = \lambda_y$. Assuming $C_{\mathbf{yy}}$ is invertible we have

$$\mathbf{w}_y = \frac{C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} \mathbf{w}_x}{\lambda} \quad (3.3)$$

and so substituting in equation (3.1) gives

$$\frac{C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} \mathbf{w}_x}{\lambda} - \lambda C_{\mathbf{xx}} \mathbf{w}_x = 0$$

or

$$C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} \mathbf{w}_x = \lambda^2 C_{\mathbf{xx}} \mathbf{w}_x \quad (3.4)$$

We are left with a generalised eigenproblem of the form $A\mathbf{x} = \lambda B\mathbf{x}$. We can therefore find the co-ordinate system that optimises the correlation between corresponding co-ordinates by first solving for the generalised eigenvectors of equation (3.4) to obtain the sequence of $\mathbf{w}_x$'s and then using equation (3.3) to find the corresponding $\mathbf{w}_y$'s.

As the covariance matrices $C_{\mathbf{xx}}$ and $C_{\mathbf{yy}}$ are symmetric positive definite we are able to decompose them using a complete Cholesky decomposition (more details on Cholesky decomposition can be found in section 4.2)

$$C_{\mathbf{xx}} = R_{\mathbf{xx}} \cdot R_{\mathbf{xx}}'$$

Appendix

CCA

$$C_{\mathbf{xx}} = R_{\mathbf{xx}} \cdot R'_{\mathbf{xx}}$$

where R_{xx} is a lower triangular matrix. If we let $\mathbf{u}_x = R'_{\mathbf{xx}} \cdot \mathbf{w}_x$ we are able to rewrite equation (3.4) as follows

$$\begin{aligned} C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} R_{\mathbf{xx}}^{-1'} \mathbf{u}_x &= \lambda^2 R_{\mathbf{xx}} \mathbf{u}_x \\ R_{\mathbf{xx}}^{-1} C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{yx}} R_{\mathbf{xx}}^{-1'} \mathbf{u}_x &= \lambda^2 \mathbf{u}_x . \end{aligned}$$

We are therefore left with a symmetric eigenproblem of the form $A\mathbf{x} = \lambda \mathbf{x}$.