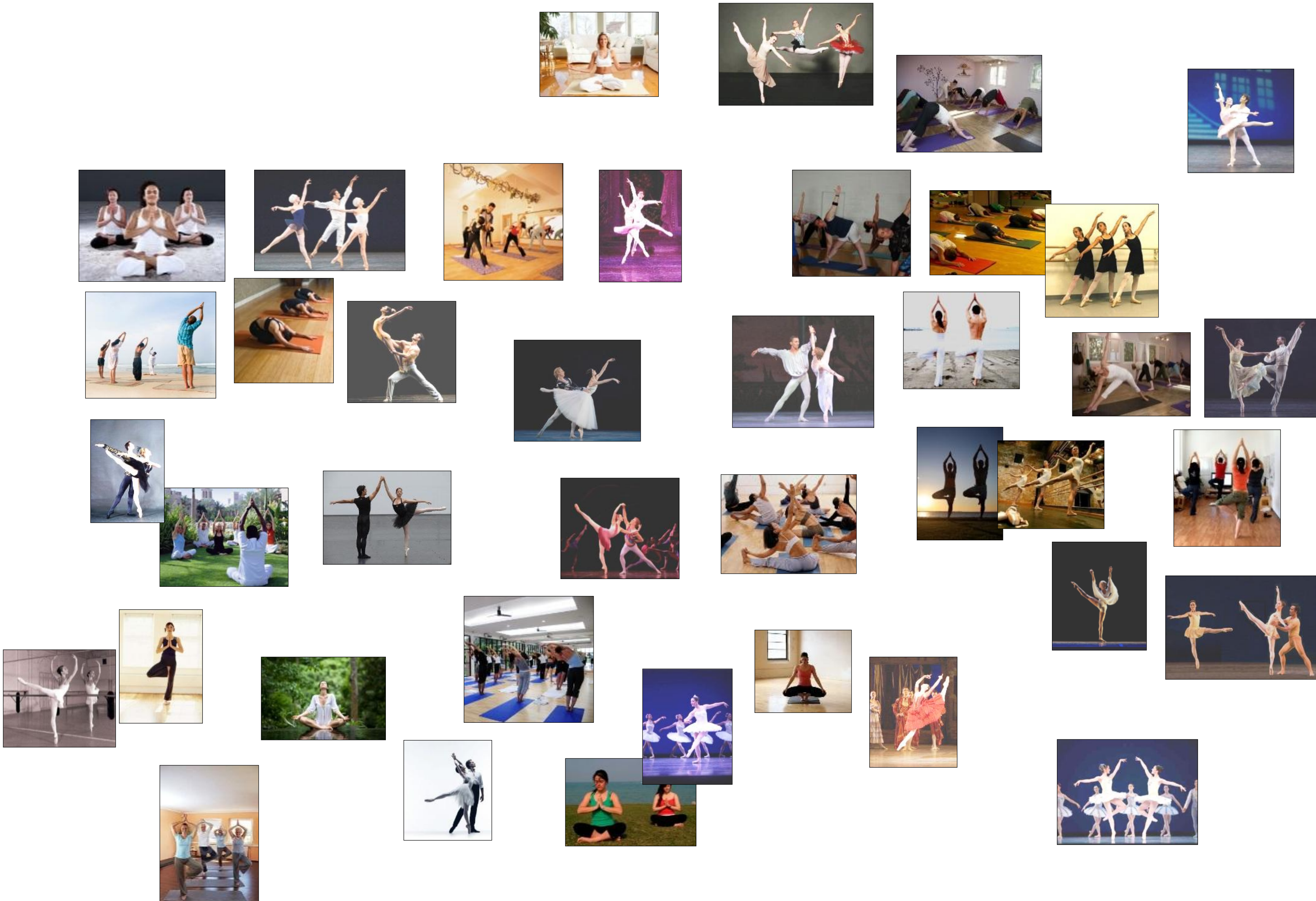
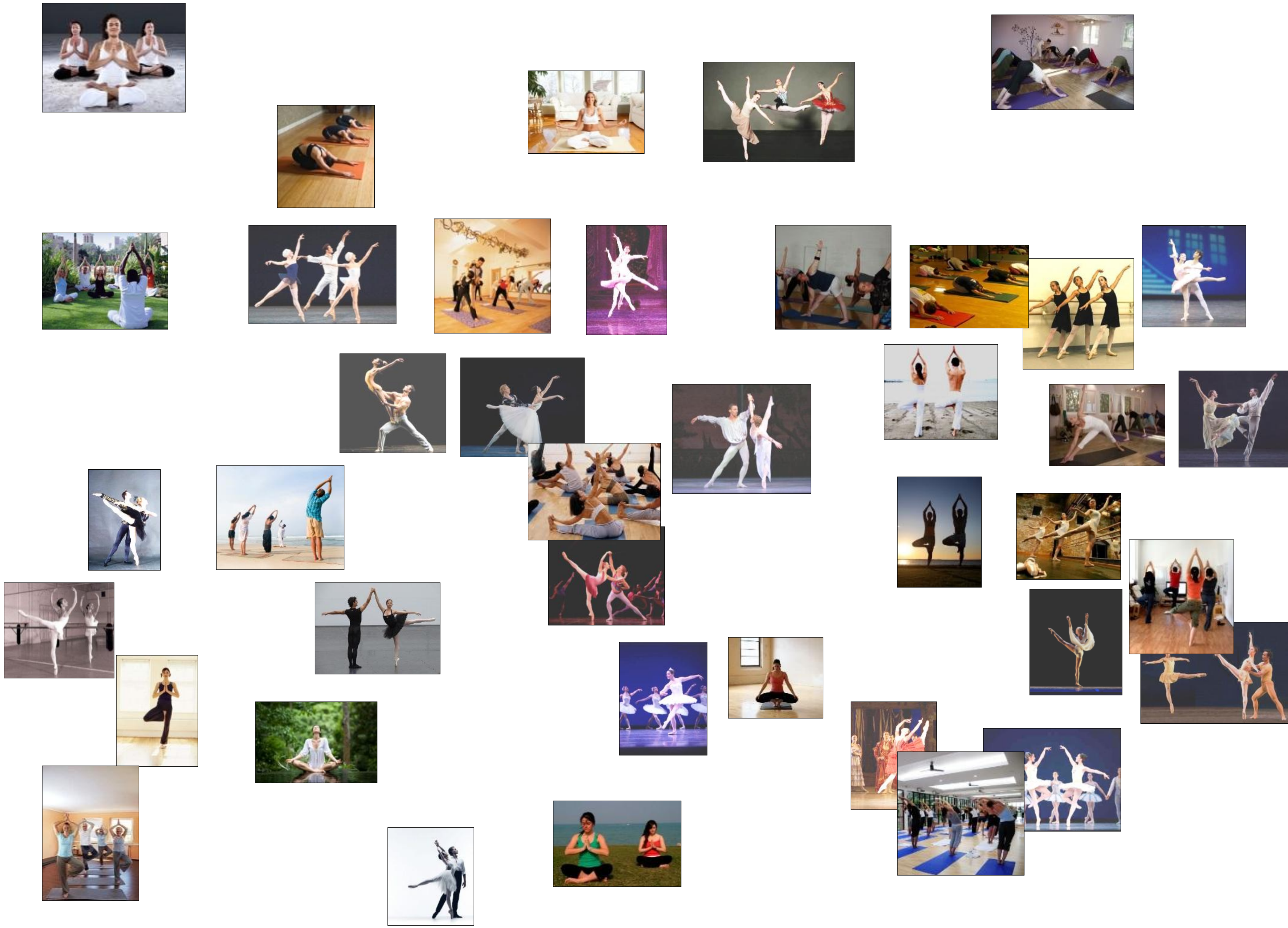
“Clustering by Composition”:
Unsupervised Discovery of Image Categories

Alon Faktor & Michal Irani
The Weizmann Institute of Science



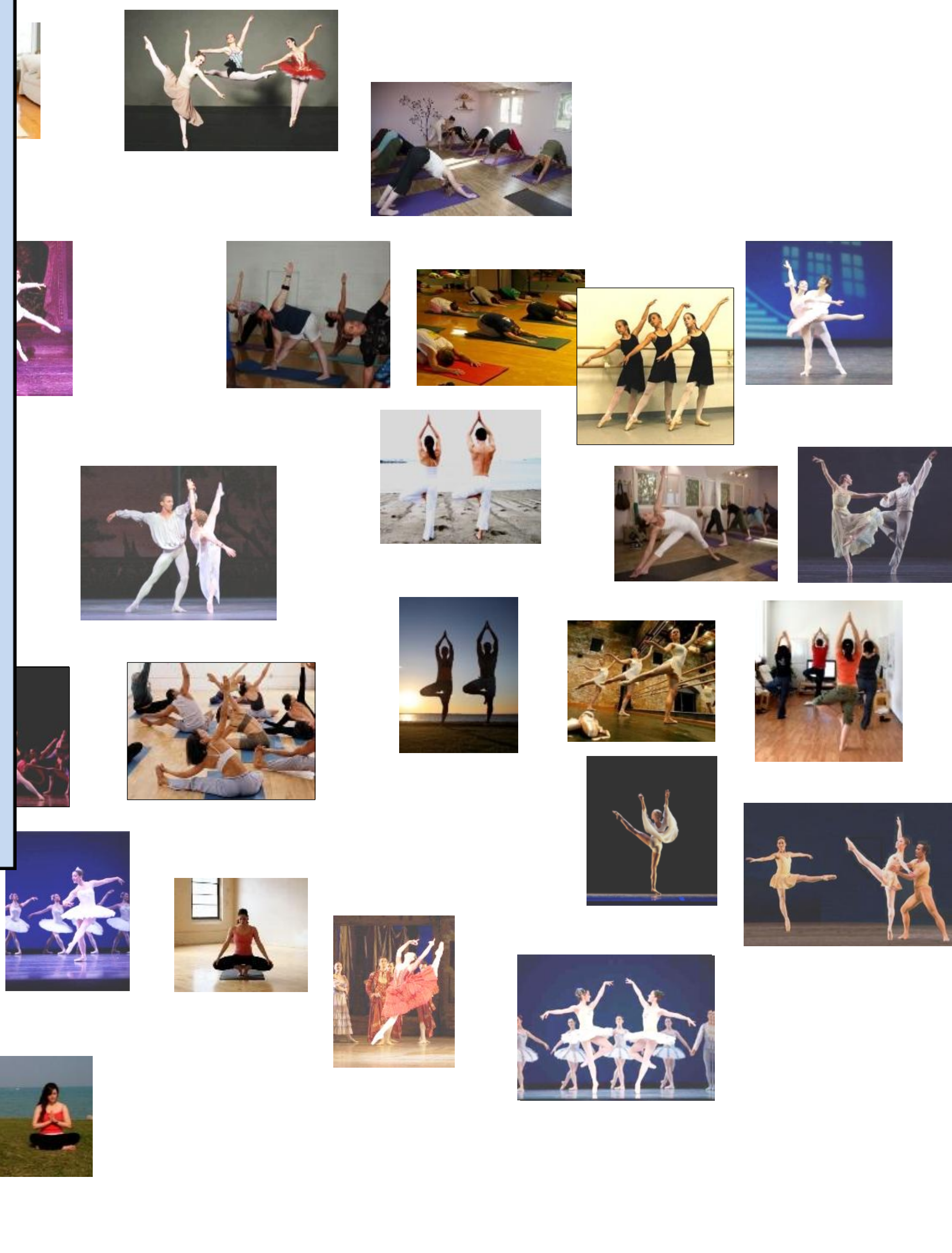


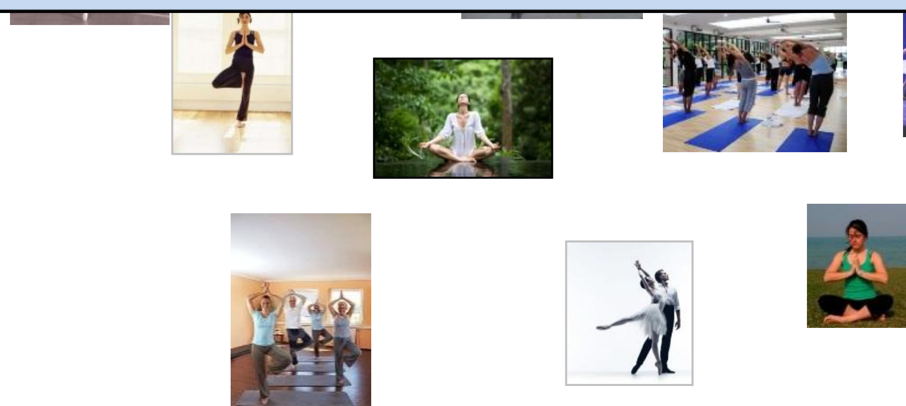
Goal: Separate into 2 clusters (Yoga & Ballet)

Previous Work:
Unsupervised Category Discovery

Simple pairwise affinities:
- Grauman & Darrell 2006
  (Pyramid match kernel)

Discover common "cluster model":
- Sivic et al 2005 (PLSA)
- Russell et al 2006 (segments)
- Kim et al 2008 (link analysis)
- Lee & Grauman 2009 (foreground)
- Payet & Todorovic 2010 (shape)
- …





<u>Goal:</u> Separate into 2 clusters (Yoga & Ballet)

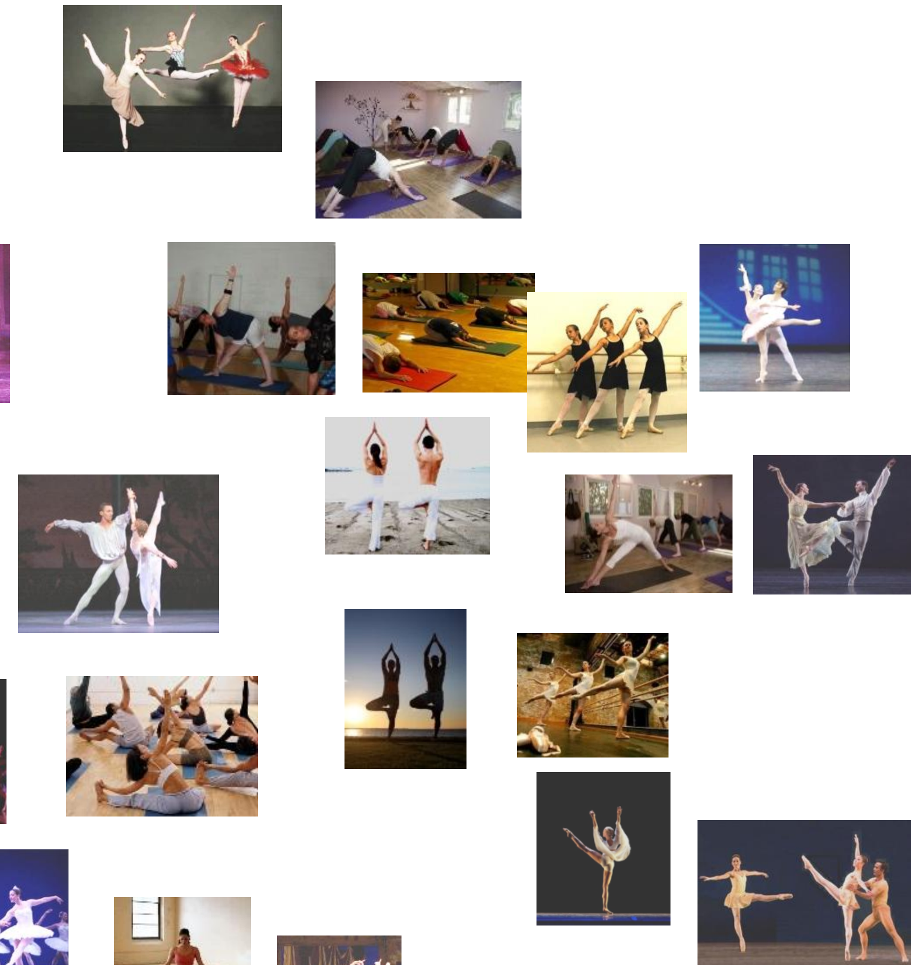
<u>Overview of Our Approach:</u>

- *"Affinity by composition"*

"Good region" ≠ "Good segment"

1) **"Rare"** *(low chance to occur at random).*
2) Shared by **2 images**.

<u>Goal:</u>* *Separate into 2 clusters (Yoga & Ballet)



<u>Overview of Our Approach:</u>

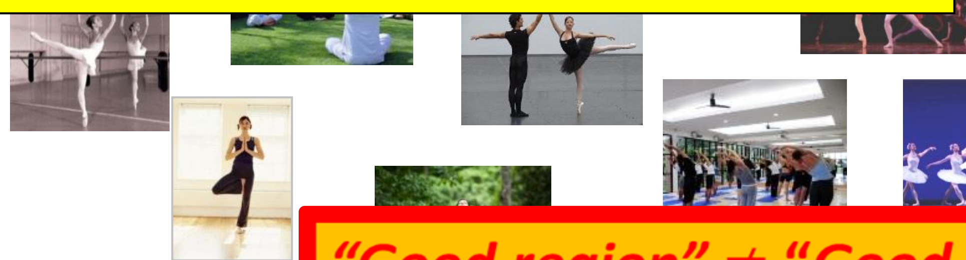
- ***"Affinity by composition"***

- **Random Search process**

  +

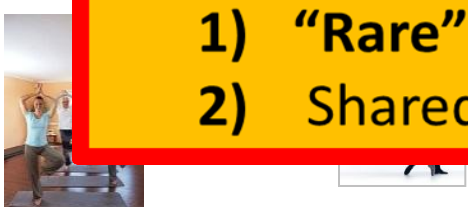
  Image collaboration

  Linear complexity



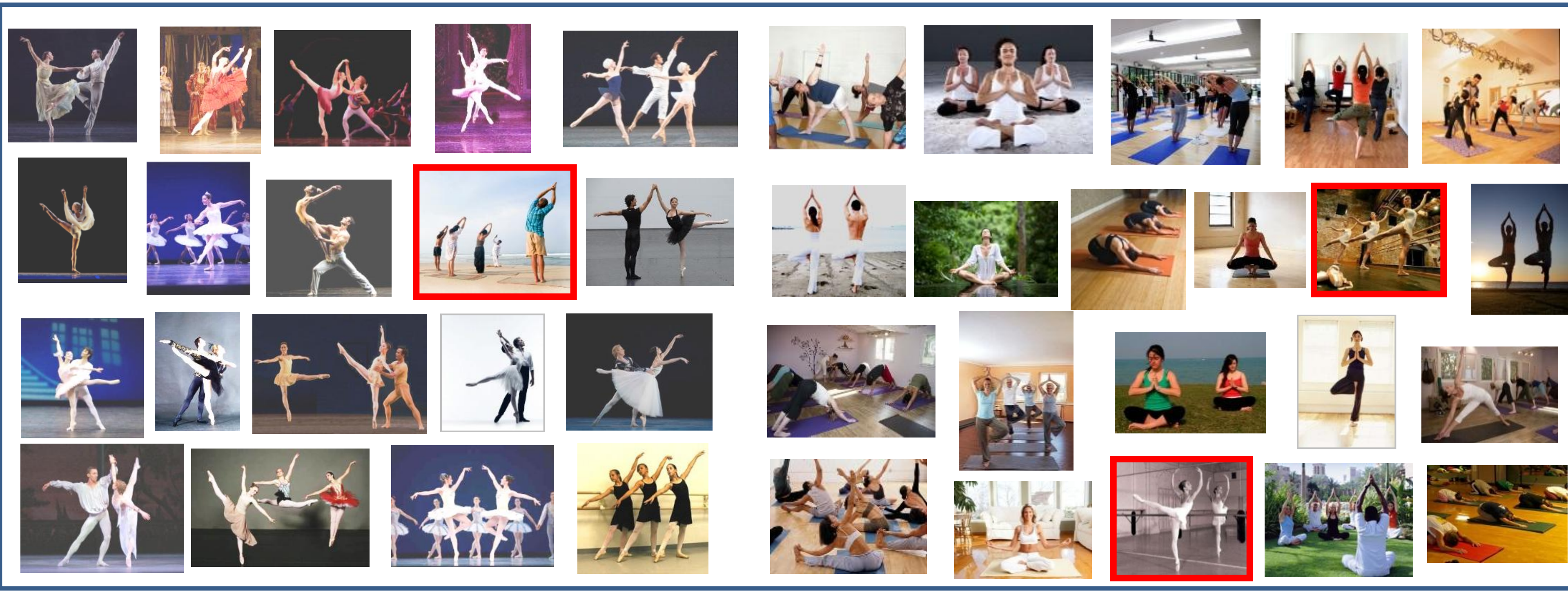
"Good region" ≠ "Good segment"

1) **"Rare"** *(low chance to occur at random).*
2) Shared by **2 images**.



Goal: Separate into 2 clusters (Yoga & Ballet)

Cluster#1 (mostly Ballet images) **Cluster#2 (mostly Yoga images)**

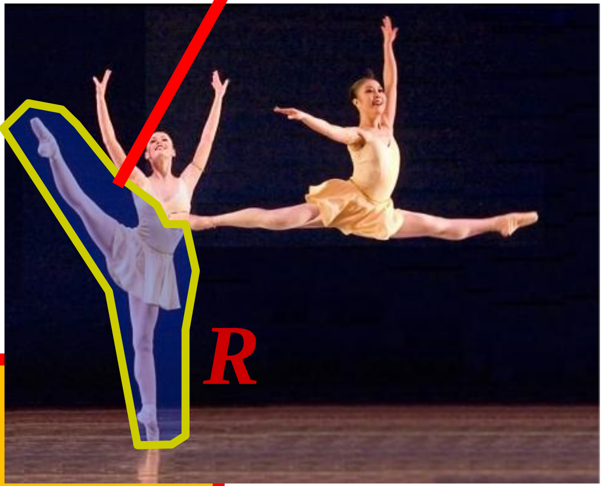


Purity = 18/19 = 95% **Purity = 19/21 = 90%**

"Affinity by Composition"

extending *[Boiman & Irani, NIPS'06]*

I_1



I_2

$$\frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

How good is the match

How rare is *R*

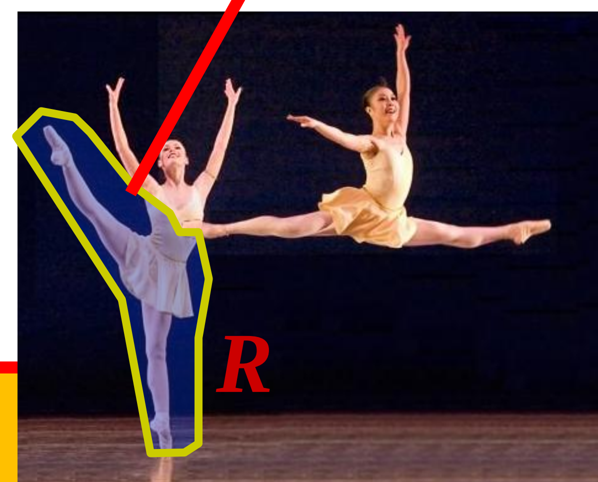
"Good region" R :

1) **Shared by 2 images.**
2) **"Rare" (low chance to occur at random).**

"Affinity by Composition"

extending *[Boiman & Irani, NIPS'06]*

I_1



I_2

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} \rightarrow$$ **contribution of *R* to affinity(*I_1*,*I_2*)**

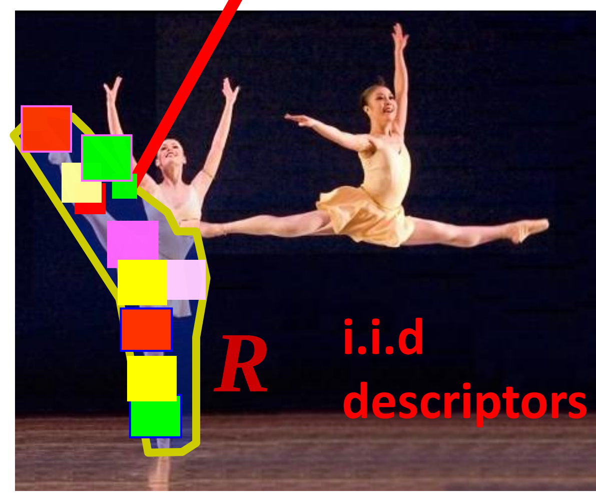
"Good region" R :

1) **Shared by 2 images.**
2) **"Rare" (low chance to occur at random).**

“Affinity by Composition”

extending *[Boiman & Irani, NIPS’06]*

I_1



I_2

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} = \sum_{d_i \in R} \Big[\underbrace{\log p(d_i \mid I_1, I_2)}_{\text{Quality of descriptor match}} - \log p(d_i \mid H_0) \Big]$$

"Affinity by Composition"

extending *[Boiman & Irani, NIPS'06]*

I_1

I_2

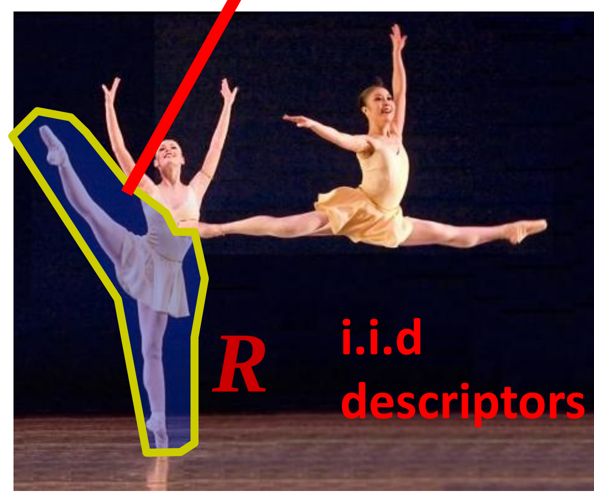
$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} =$$

$$\sum_{d_i \in R} \Big[\underbrace{\log p\big(d_i \mid I_1, I_2\big)}_{\text{Quality of descriptor match}} - \log p\big(d_i \mid H_0\big) \Big]$$

"Affinity by Composition"

extending *[Boiman & Irani, NIPS'06]*

I_1



I_2

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} =$$

$$\sum_{d_i \in R} \Big[\underbrace{\log p(d_i \mid I_1, I_2)}_{\boxed{-Err(d_i \mid I_1, I_2)}} - \underbrace{\log p(d_i \mid H_0)}_{\text{How rare is the descriptor?}} \Big]$$

Statistically Significant Descriptors

How rare is each descriptor:

$$-\log p(d_i \mid H_0) = ?$$




Generate a "codebook" (quantized descriptors):

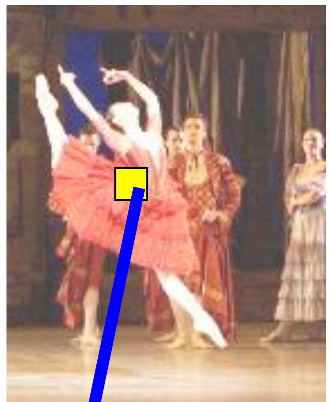
- *Frequent descriptors → Low error*
- *Rare descriptors → High error*

Codebook

Statistically Significant Descriptors

How rare is each descriptor:

$$-\log p(d_i \mid H_0) = ?$$



Codebook

Compute error to *Nearest Codeword*

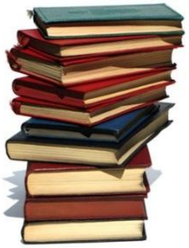
Statistically Significant Descriptors

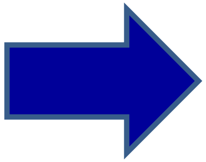
How rare is each descriptor: $$-\log p(d_i \mid H_0) = ?$$




$$\boxed{-\log p(d_i \mid H_0) \approx Err(d_i \mid Codebook)}$$

Codebook



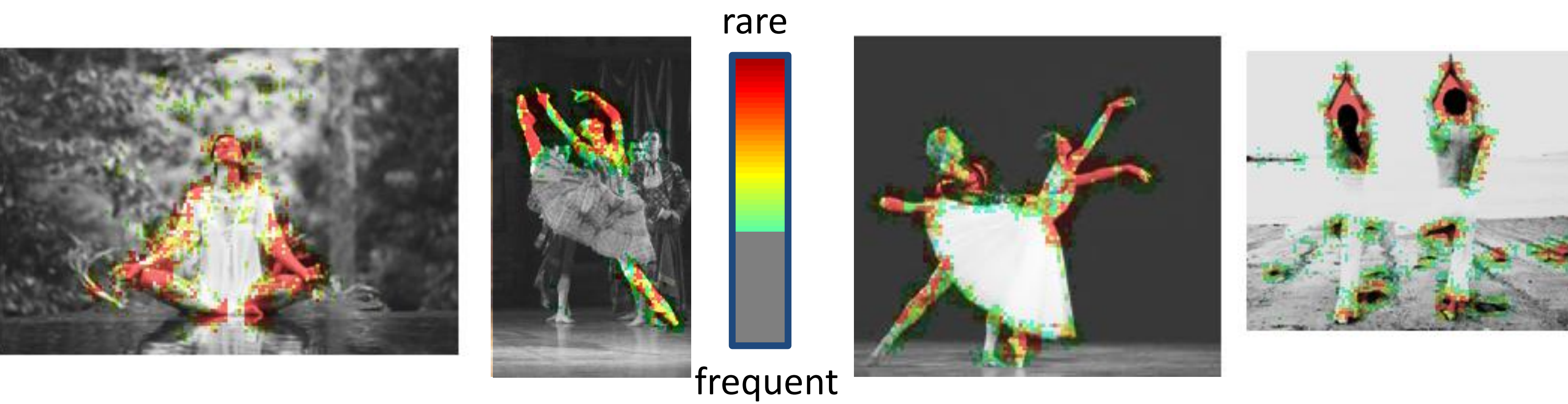


Compute error to *Nearest Codeword*

Statistically Significant Descriptors

How rare is each descriptor:

$$-\log p(d_i \mid H_0) = ?$$

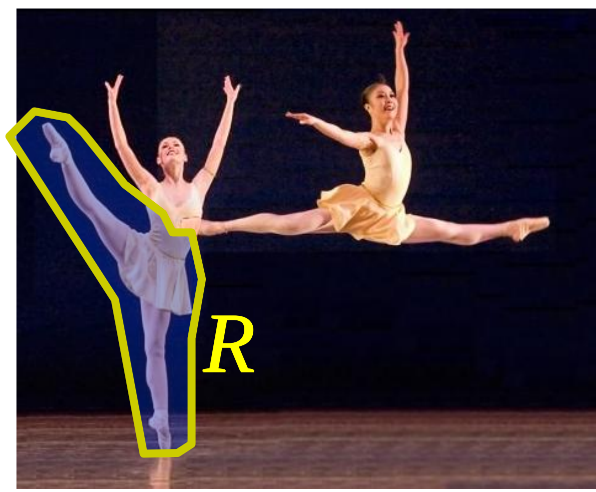


$$\boxed{-\log p(d_i \mid H_0) \approx Err(d_i \mid Codebook)}$$

The most informative descriptors are the rare ones !

“Affinity by Composition”

I_1



I_2

Contribution of *R* to affinity(I_1,I_2)

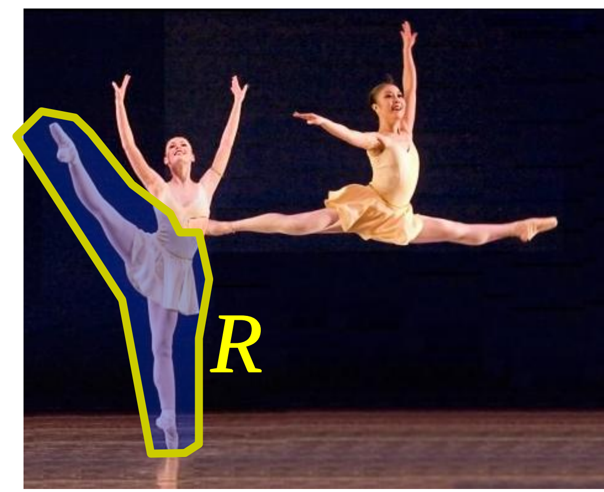
How good is the match

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

How rare is *R*

"Affinity by Composition"

I_1



I_2

Contribution of R to affinity(I_1,I_2)

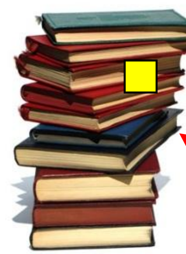
$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{d_i \in R} \left[\boxed{Err(d_i \mid codebook) - Err(d_i \mid I_1, I_2)} \right]$$

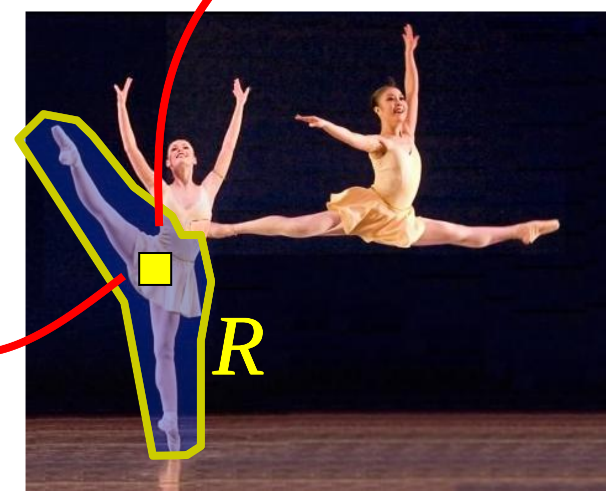
"Affinity by Composition"

***I*₁**

Codebook



Contribution of *R* to affinity(I_1,I_2)



***I*₂**

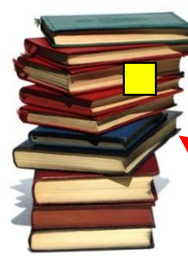
$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{d_i \in R} \left[Err\left(d_i \mid codebook\right) - Err\left(d_i \mid I_1, I_2\right) \right]$$

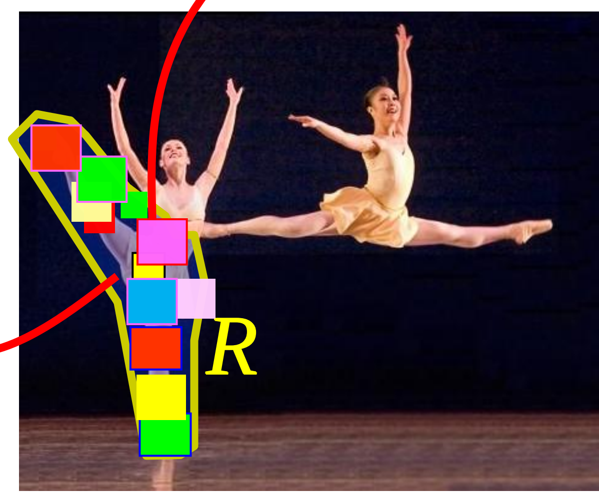
"Affinity by Composition"

I_1

Codebook



Contribution of R to affinity(I_1,I_2)



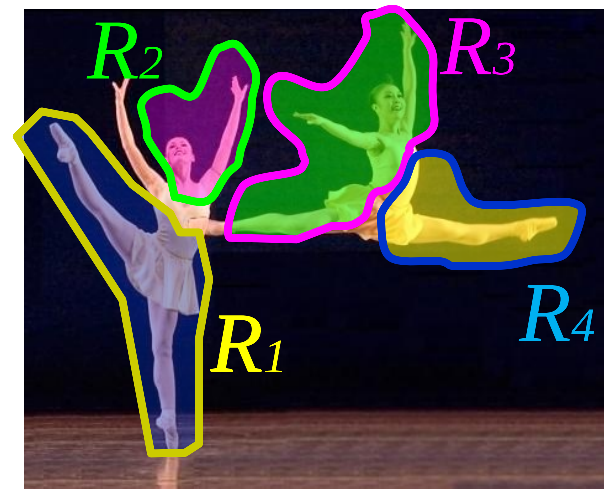
$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{\boxed{d_i \in R}} \left[Err\left(d_i \mid codebook\right) - Err\left(d_i \mid I_1, I_2\right) \right]$$

I_2

“Affinity by Composition”

I_1



I_2

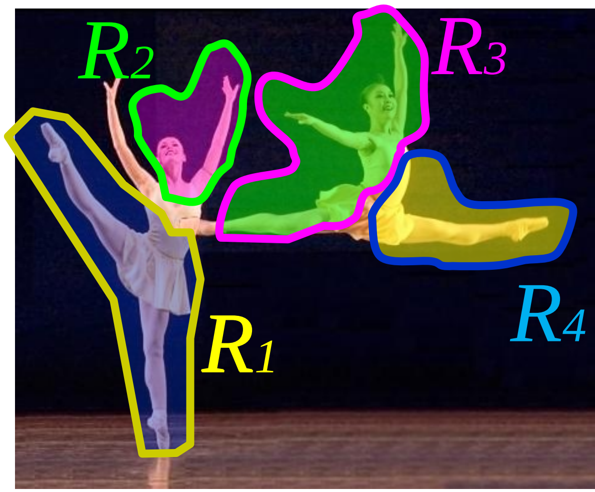
Contribution of *R* to affinity(I_1,I_2)

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{d_i \in R} \left[Err\left(d_i \mid codebook\right) - Err\left(d_i \mid I_1, I_2\right) \right]$$

“Affinity by Composition”

***I*1**



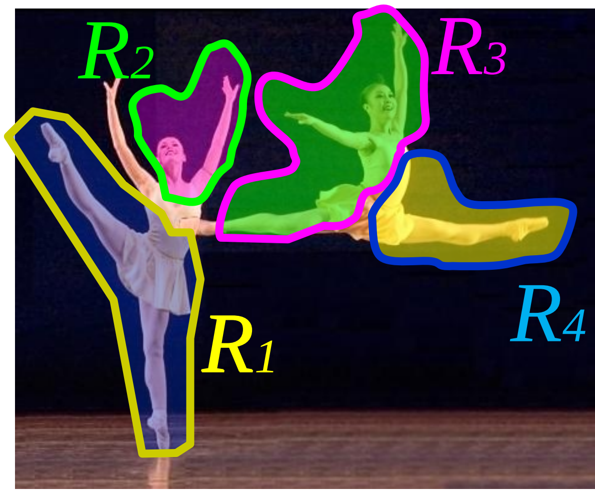
***I*2**

$$\sum_{R} \log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{R} \sum_{d_i \in R} \left[Err\left(d_i \mid codebook\right) - Err\left(d_i \mid I_1, I_2\right) \right]$$

"Affinity by Composition"

I_1



I_2

$$\textbf{Affinity}(I_1, I_2) = \sum_{R} \sum_{d_i \in R} \left[Err(d_i \mid codebook) - Err(d_i \mid I_1, I_2) \right]$$

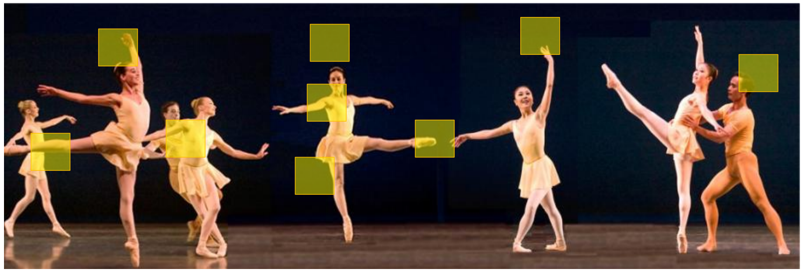
“Region Match”

Using Randomized Search

Extending “Patch Match”

[Barnes et al 2009]

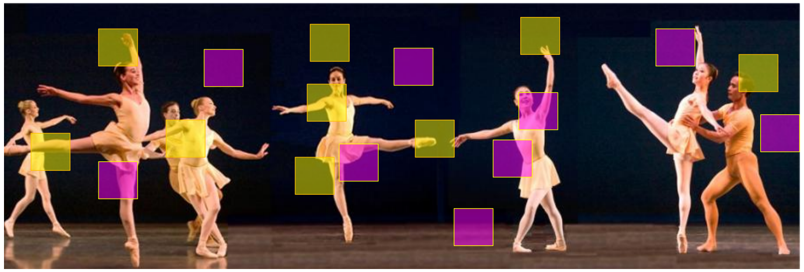
“Region Match”
Using Randomized Search



Extending “Patch Match”

[Barnes et al 2009]

“Region Match”
Using Randomized Search

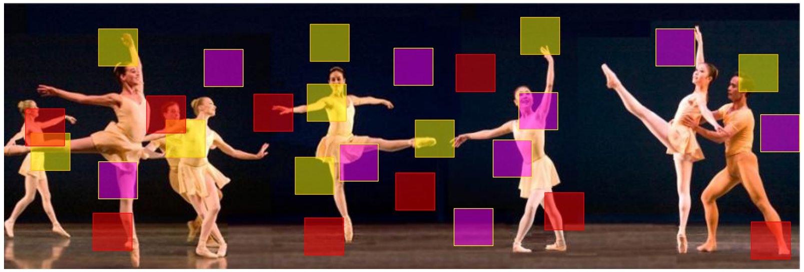


Extending “Patch Match”

[Barnes et al 2009]

"Region Match"

Using Randomized Search

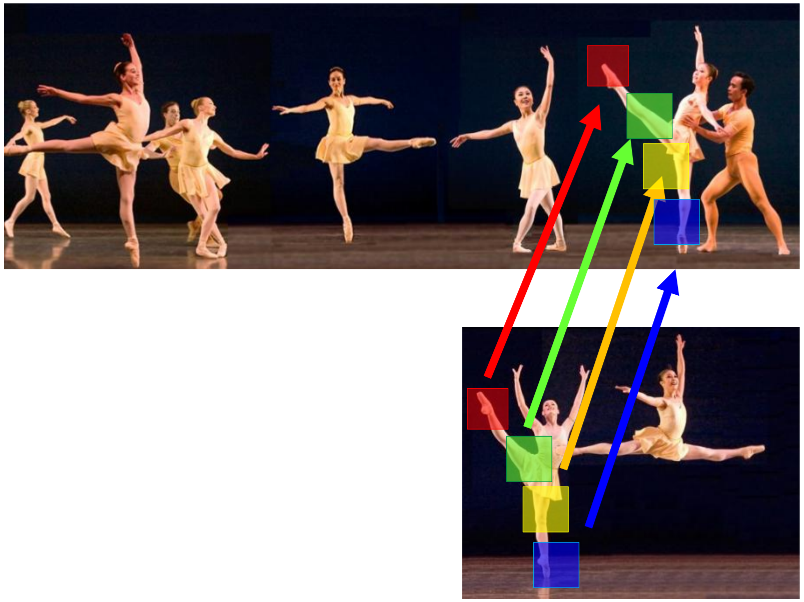


Extending "Patch Match"

[Barnes et al 2009]

"Region Match"

Using Randomized Search



Extending "Patch Match"

[Barnes et al 2009]

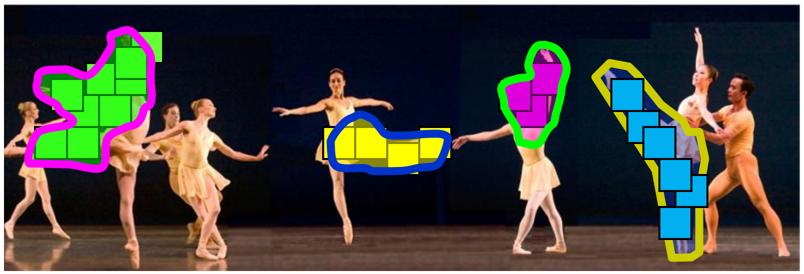
“Region Match”

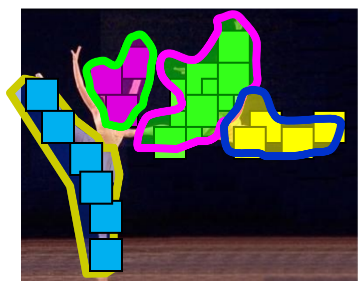
Using Randomized Search

Coherently mapped descriptors
➔ Large shared regions

"Region Match"

Using Randomized Search





CLAIM: ***40 samples per descriptor***

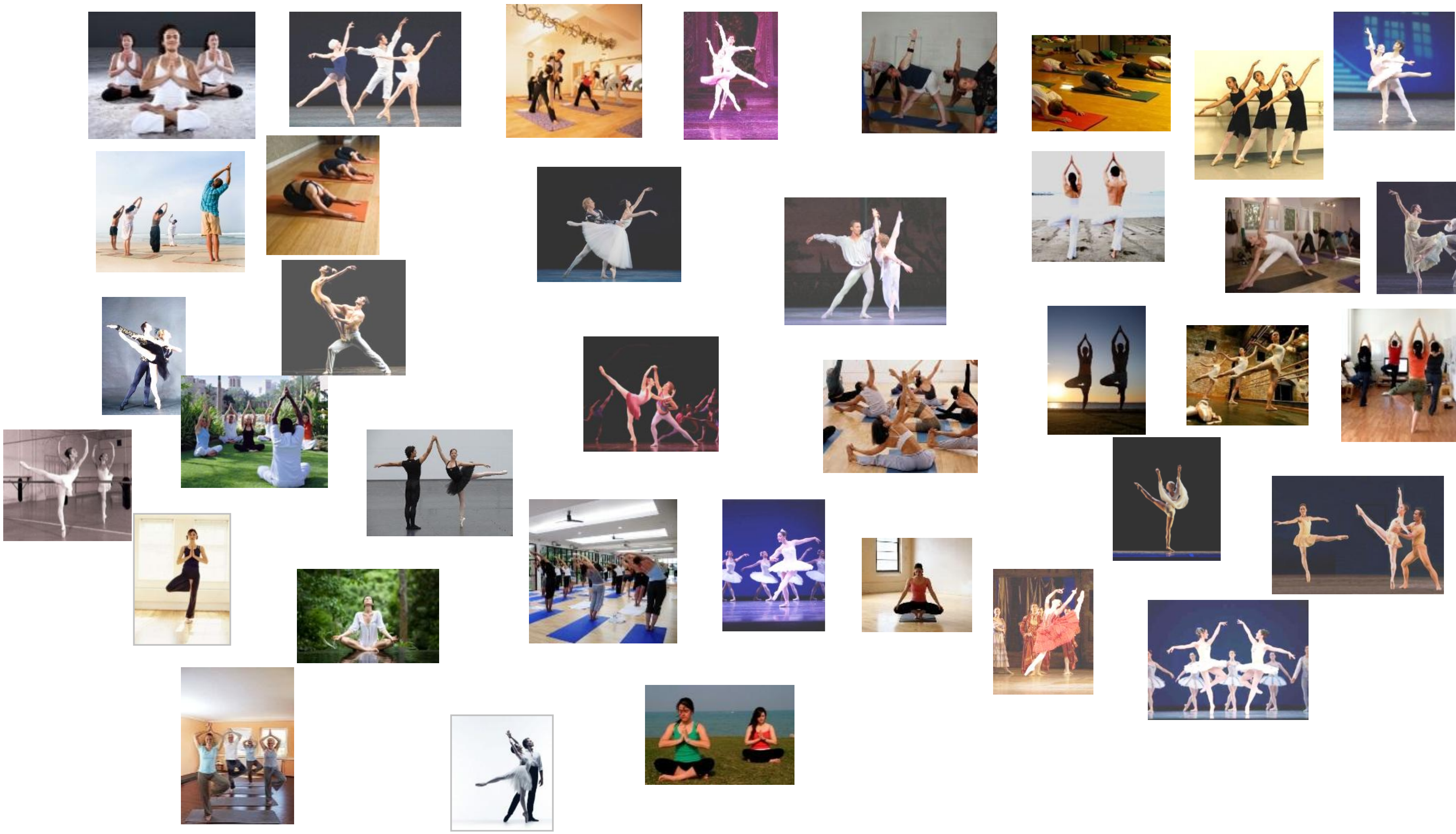
Find shared regions $|R| \geq 10\%$
with very high probability $\geq 98\%$

➔ ***linear complexity*** $O(n)$

Explicit segmentation ➔ ***NO NEED !***

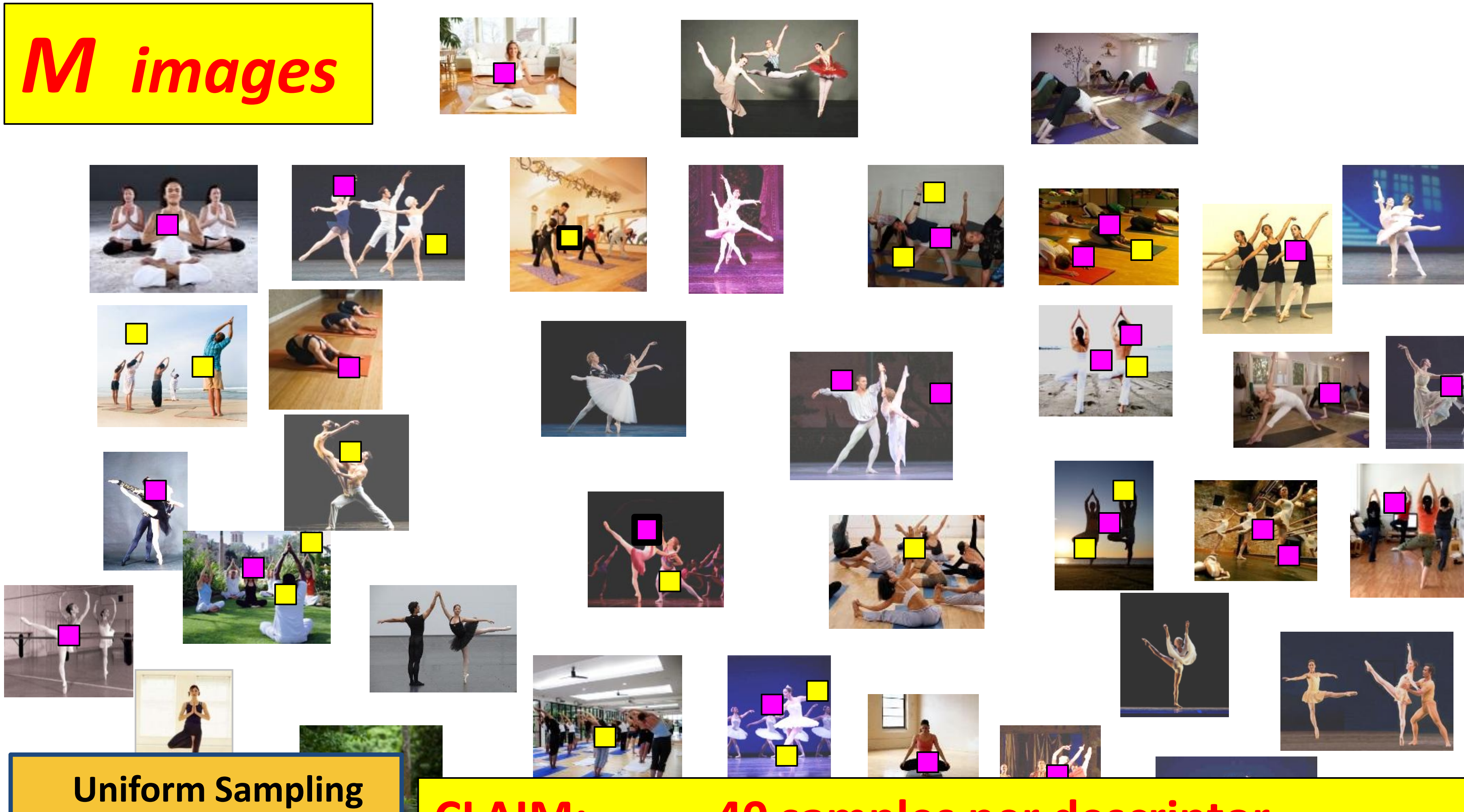
Going back to our Clustering problem…

M images



Going back to our Clustering problem…

M images



Uniform Sampling across all images

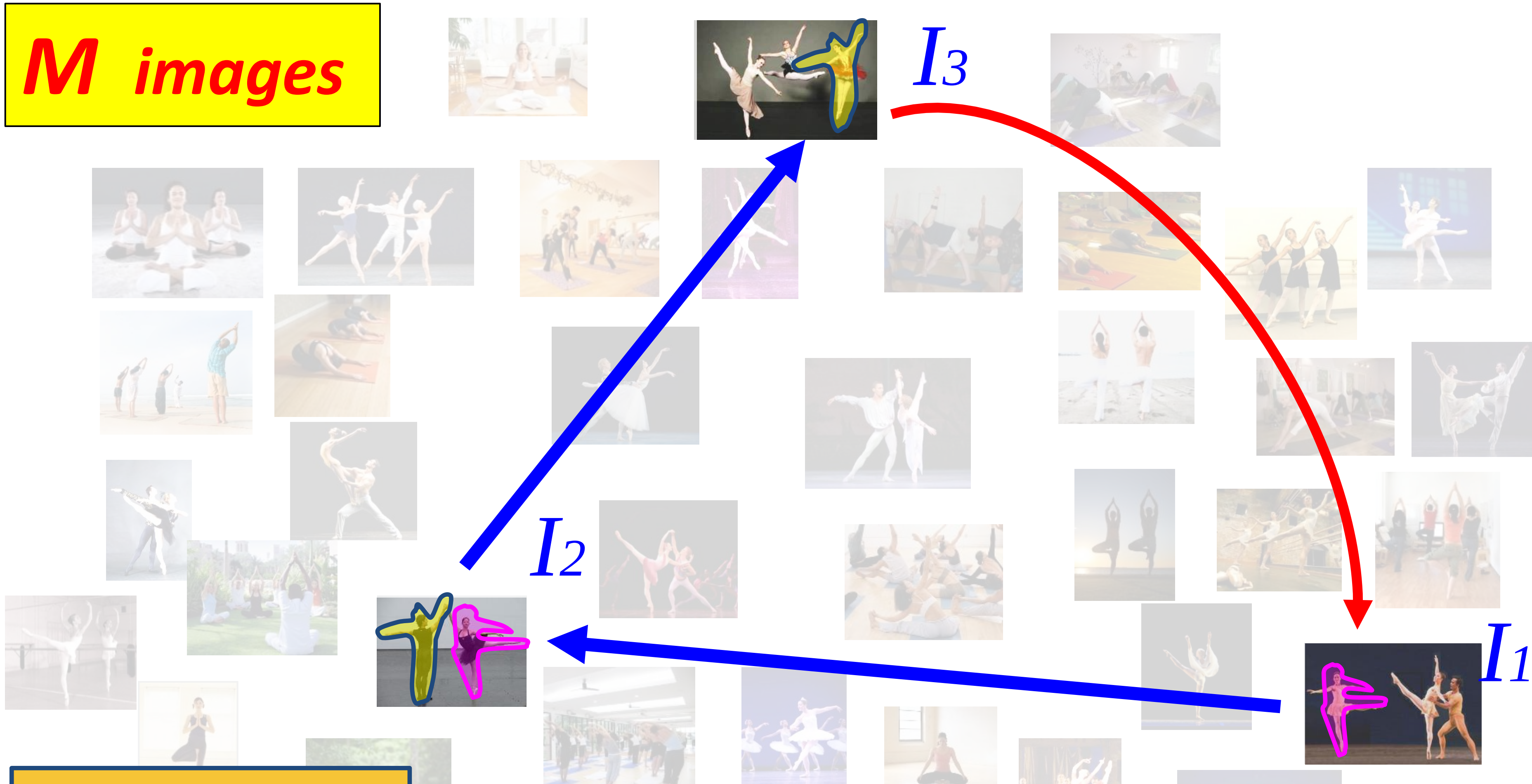
Images

CLAIM: 40 samples per descriptor **regardless of # images**

➔ *At least* one strong connection per image

Going back to our Clustering problem…

M images



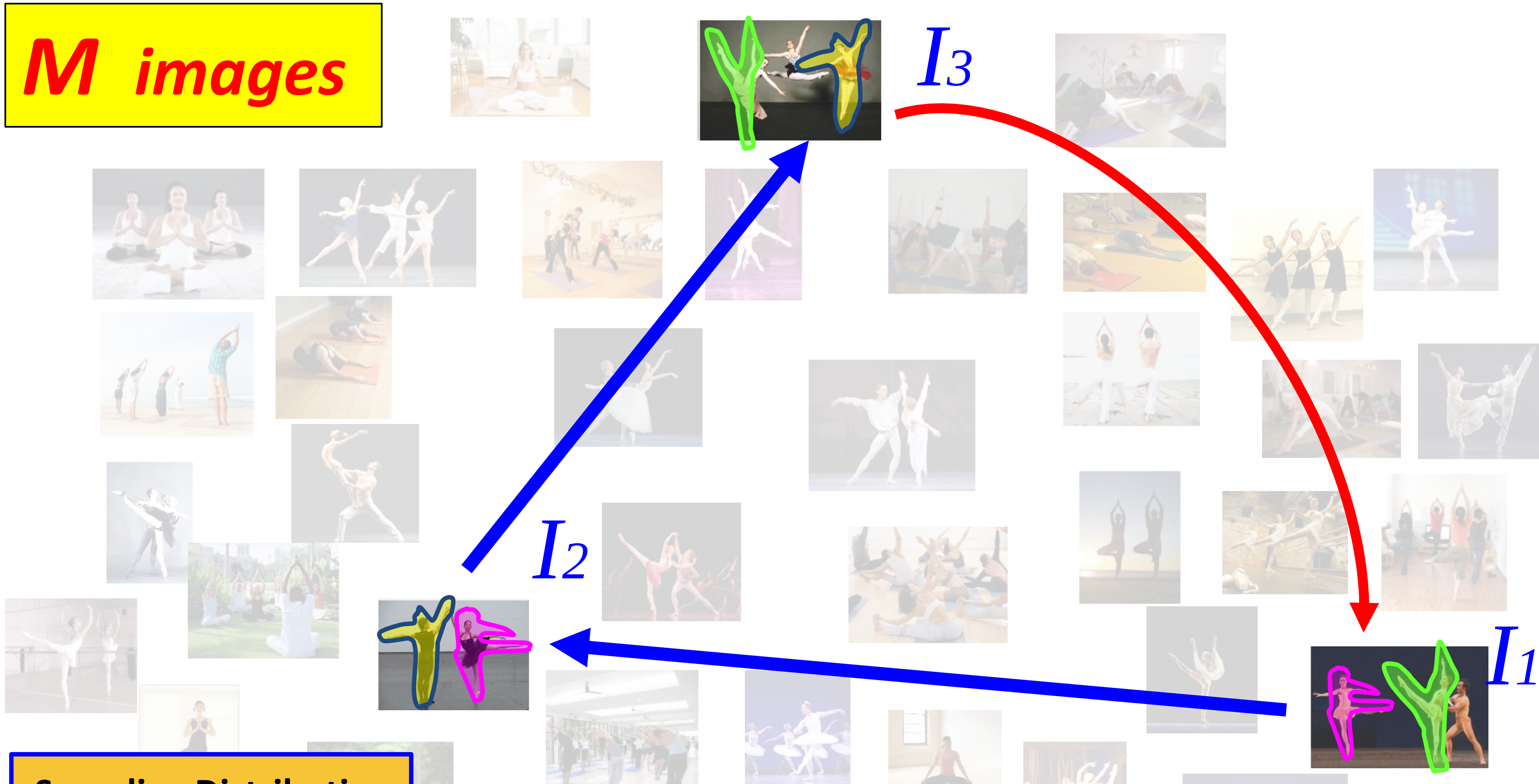
Uniform Sampling across all images

<u>CLAIM:</u> 40 samples per descriptor <u>regardless of # images</u>

➔ *At least* <u>one strong connection per image</u>

Going back to our Clustering problem…

M images



Sampling Distribution

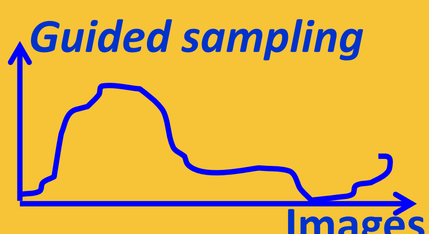


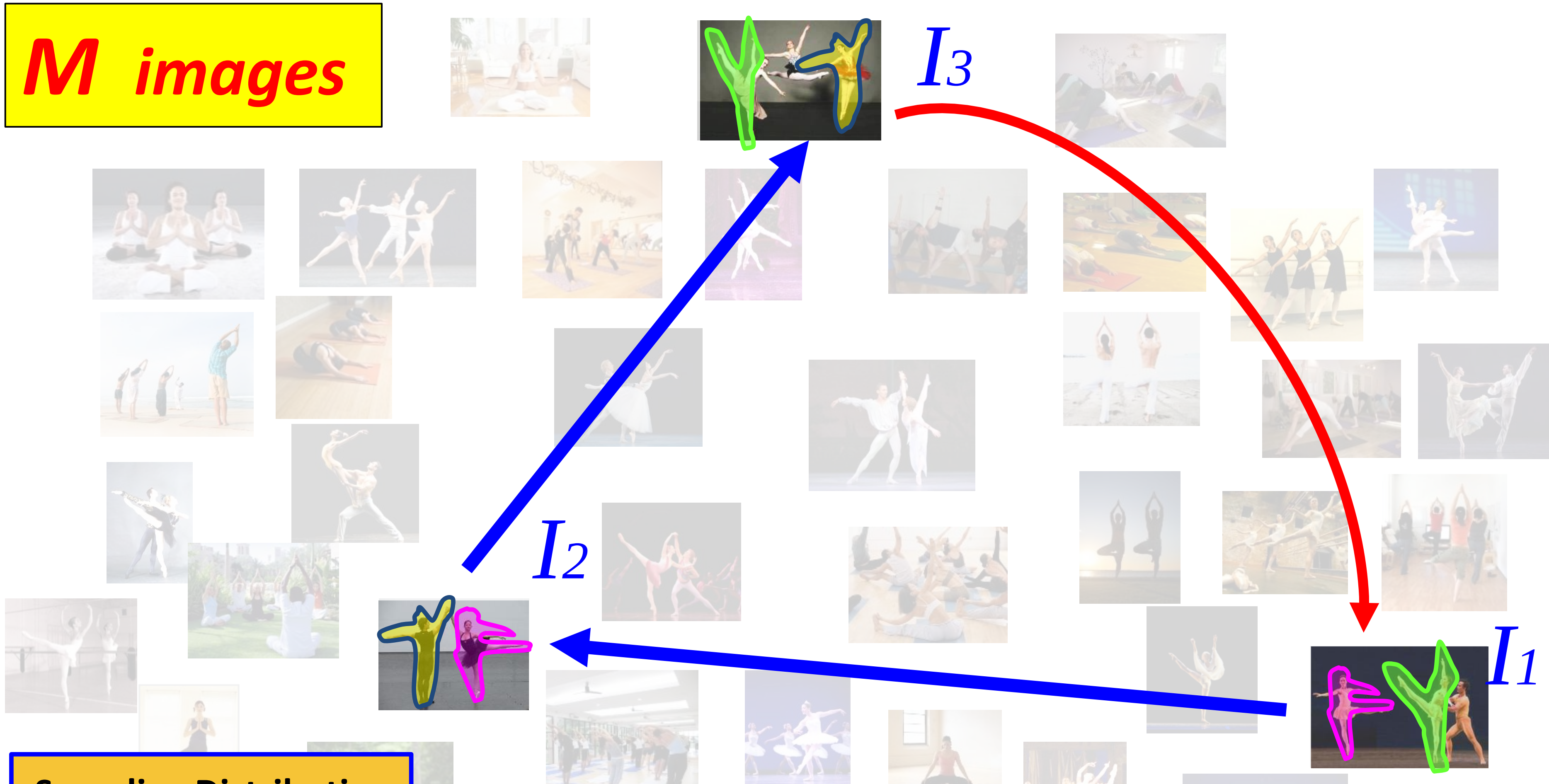
Image collaboration

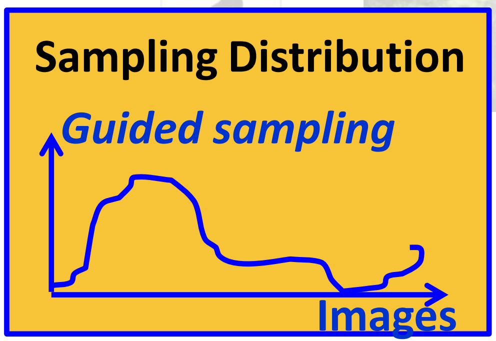
➔ **Sparse set of <u>meaningful</u> affinities**

➔ ***<u>Linear complexity</u>*** *(in size of dataset)*

Going back to our Clustering problem…

M images



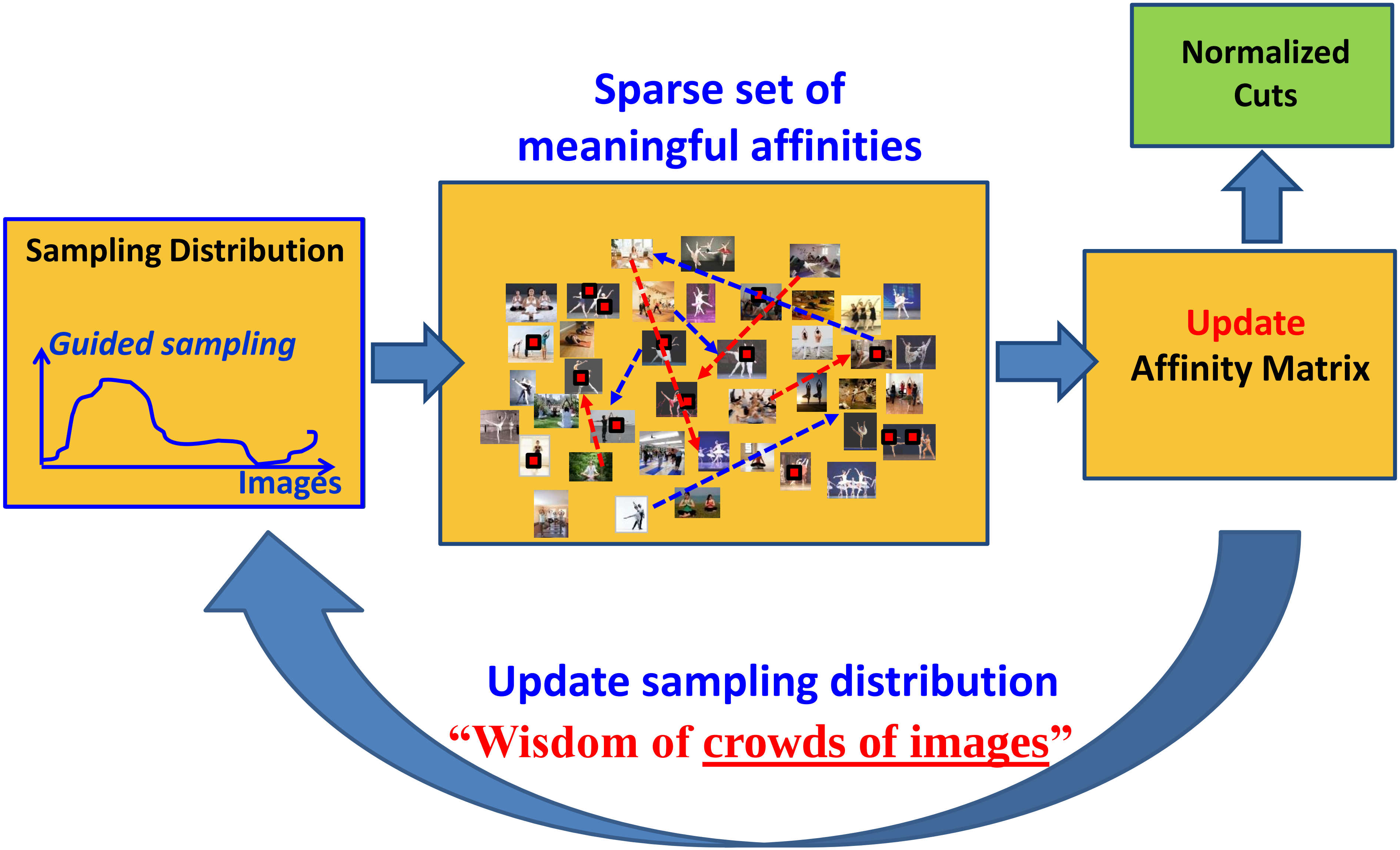


"Wisdom of crowds of images"

➔ **Sparse set of meaningful affinities**

➔ ***Linear complexity*** *(in size of dataset)*

Our Full Clustering Algorithm



Experiments

- ***Comparisons on Benchmark Datasets* (Caltech, ETHZ)**

  ➔ **Significant improvement over *state-of-the-art***

  (up to 30%)

- ***Experiments on <u>more challenging</u> datasets***

  ➔ Tiny datasets

  ➔ PASCAL-VOC

Experiments on Tiny Dataset

20 images (4 classes)



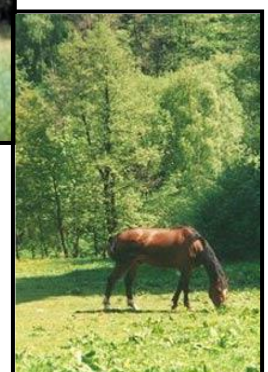
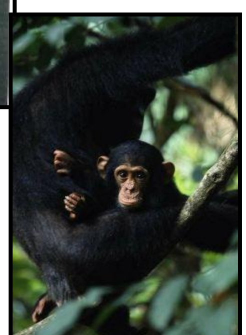






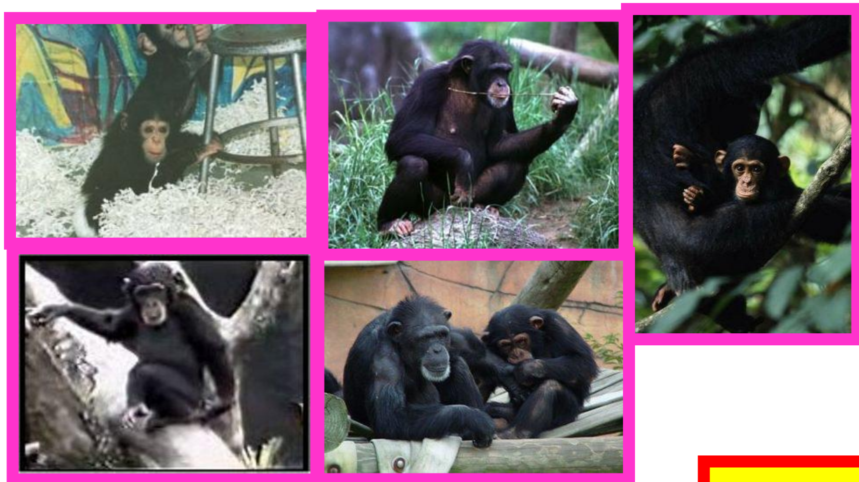




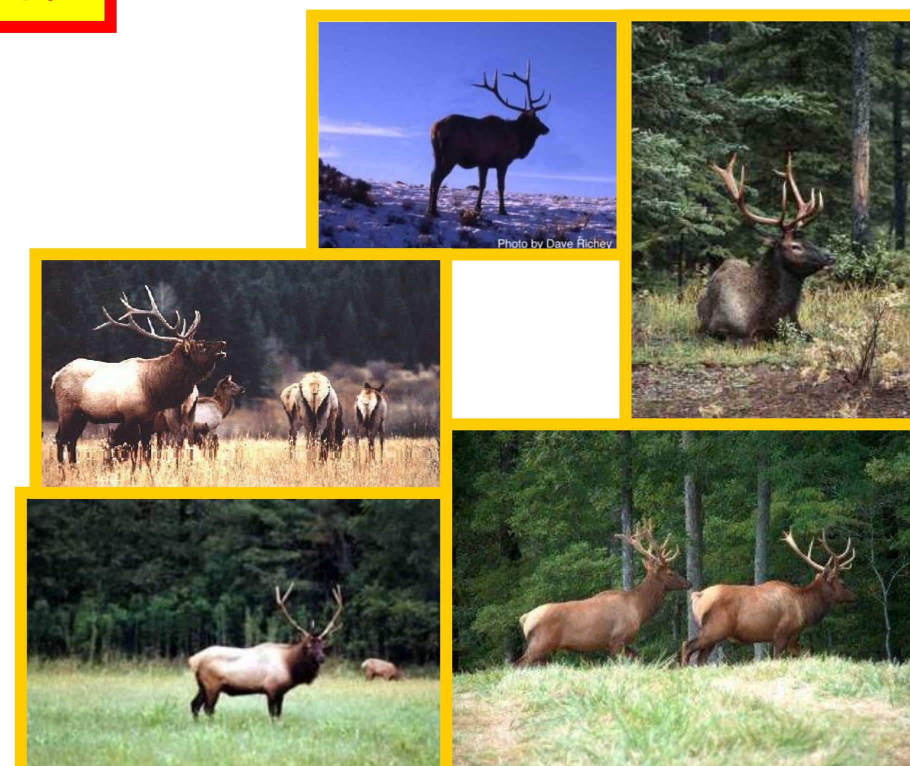






Experiments on Tiny Dataset



Purity = 100%



Experiments on Tiny Dataset

Statistically Significant Descriptors

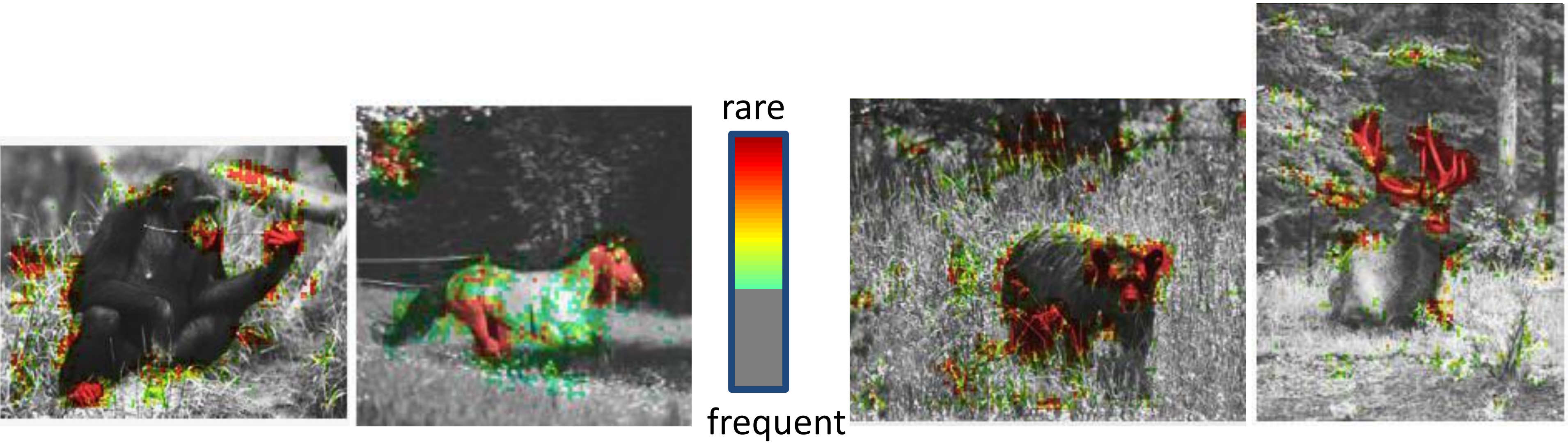
$$-\log p(d_i \mid H_0) \approx Err(d_i \mid Codebook)$$




Experiments on Tiny Dataset

Statistically Significant Descriptors

$$-\log p(d_i \mid H_0) \approx Err(d_i \mid Codebook)$$



The statistically significant descriptors → on the animals!

PASCAL subset (4 classes)

CARS, BICYCLES, CHAIRS, HORSES

PASCAL subset (4 classes)

CARS, BICYCLES, HORSES, CHAIRS

PASCAL subset (4 classes)

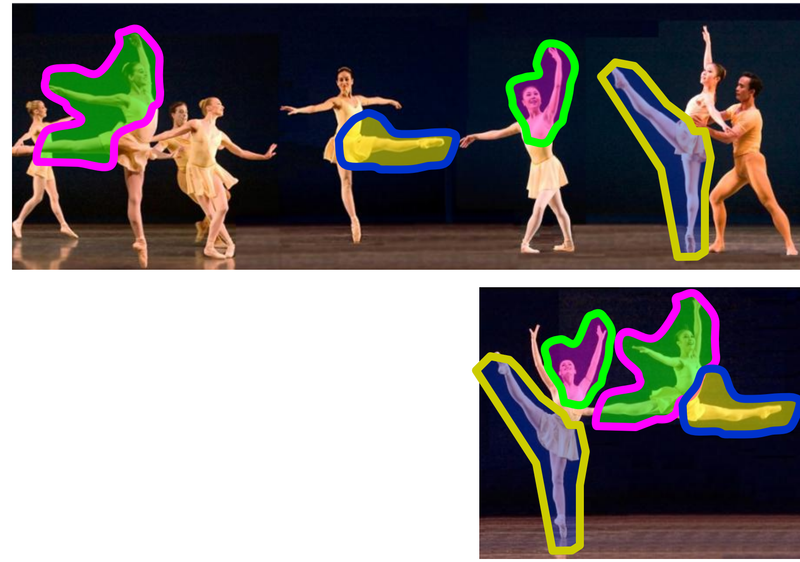
CARS, BICYCLES, HORSES, CHAIRS

Mean purity = 67%

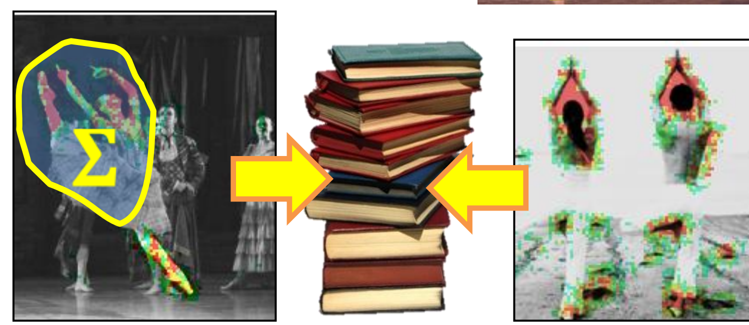
20% <u>better</u> than a baseline of:
SPM + Ncuts

Thank you!

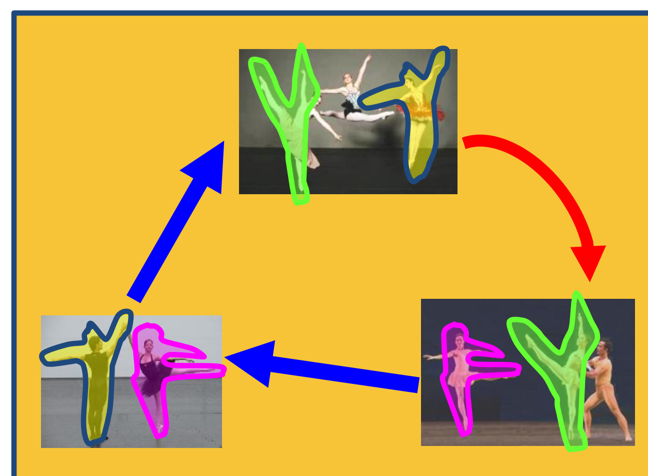
1. **"Affinity by composition"**
   ➔ Look for RARE shared regions



2. **Codebook Quantization Error**
   ➔ Estimate how rare a region is.



3. **Randomized search using the "Wisdom of Crowds of images"**
   ➔ Find shared regions
   ➔ Linear complexity



4. **State of the art results**
   - Benchmark datasets
   - New challenging datasets