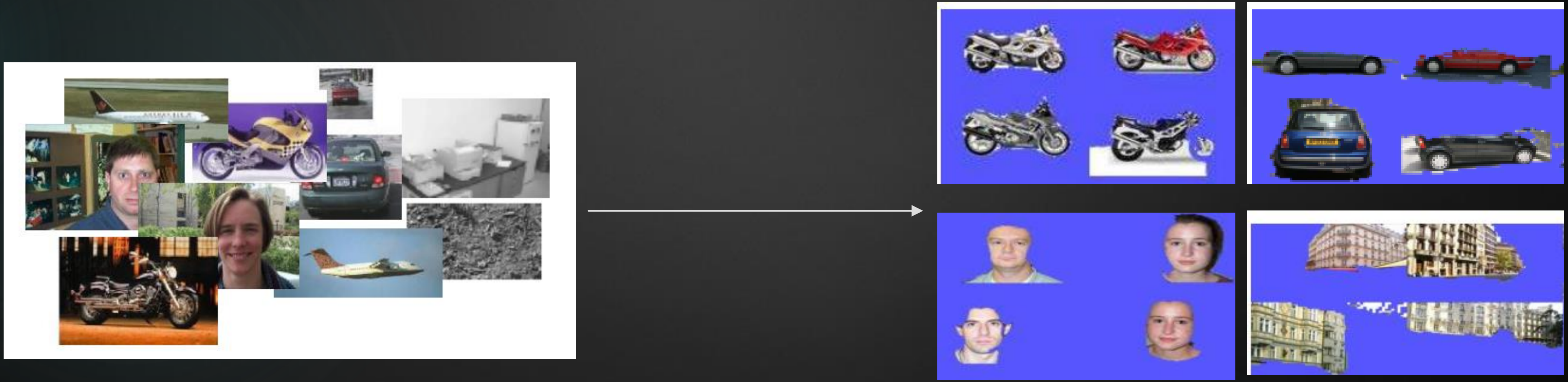
Using Multiple Segmentations to Discover Objects and their Extent in Image Collections

BRYAN RUSSELL, ALEXEI EFROS, JOSEF SIVIC, WILLIAM T. FREEMAN, ANDREW ZISSERMAN

YIYING LI

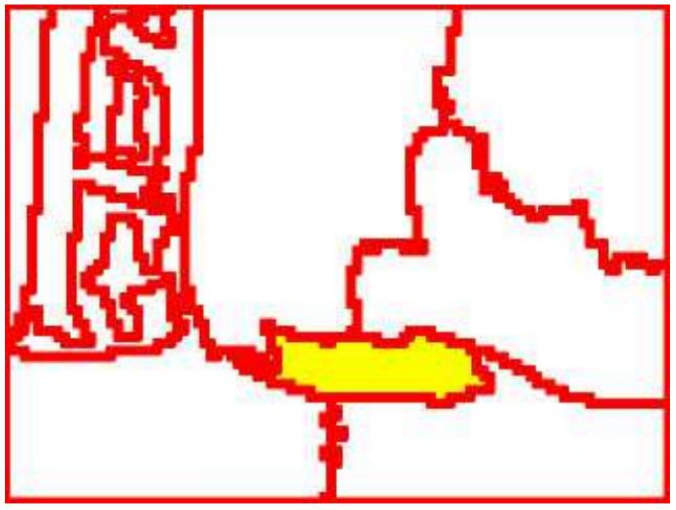
Problem

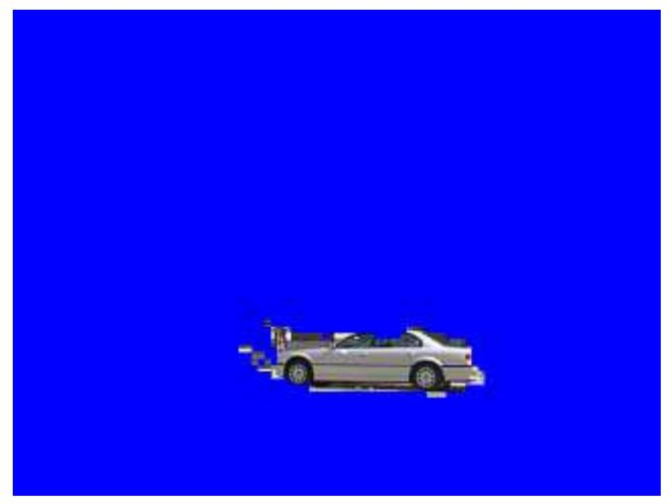
- "Is it possible to learn visual object classes simply from looking at images?"
- Given a collection of images that are unlabeled, discover object categories and segmentations automatically



Background

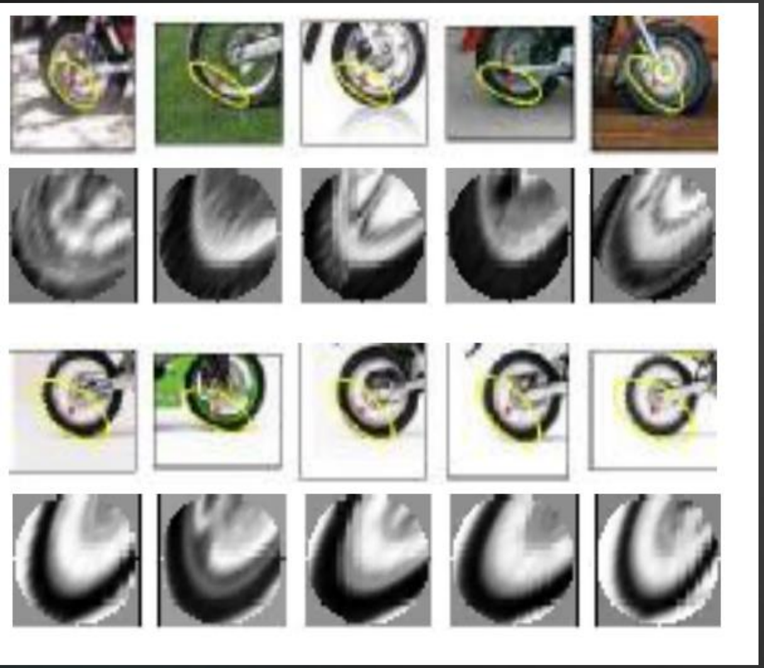
- General Approach
  - Segmentation
  - Visual Words
  - Clustering
    - K-means

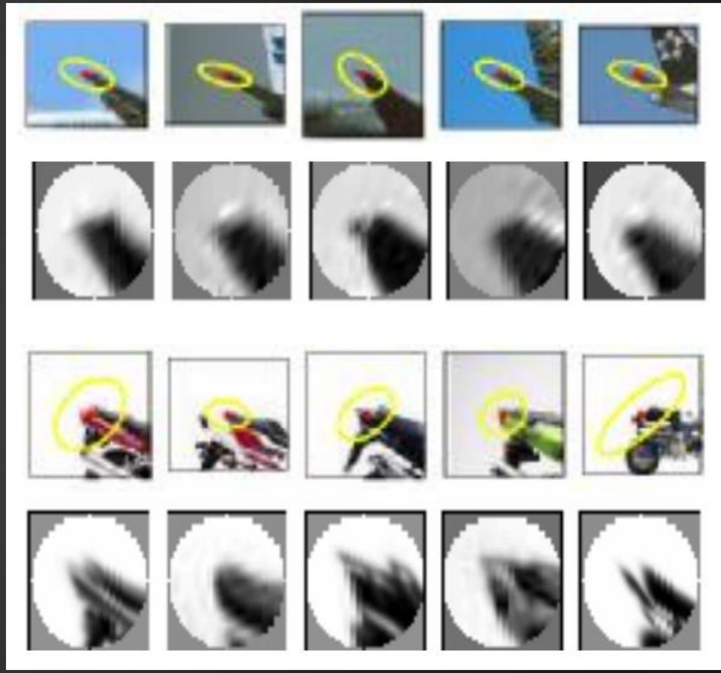




Background

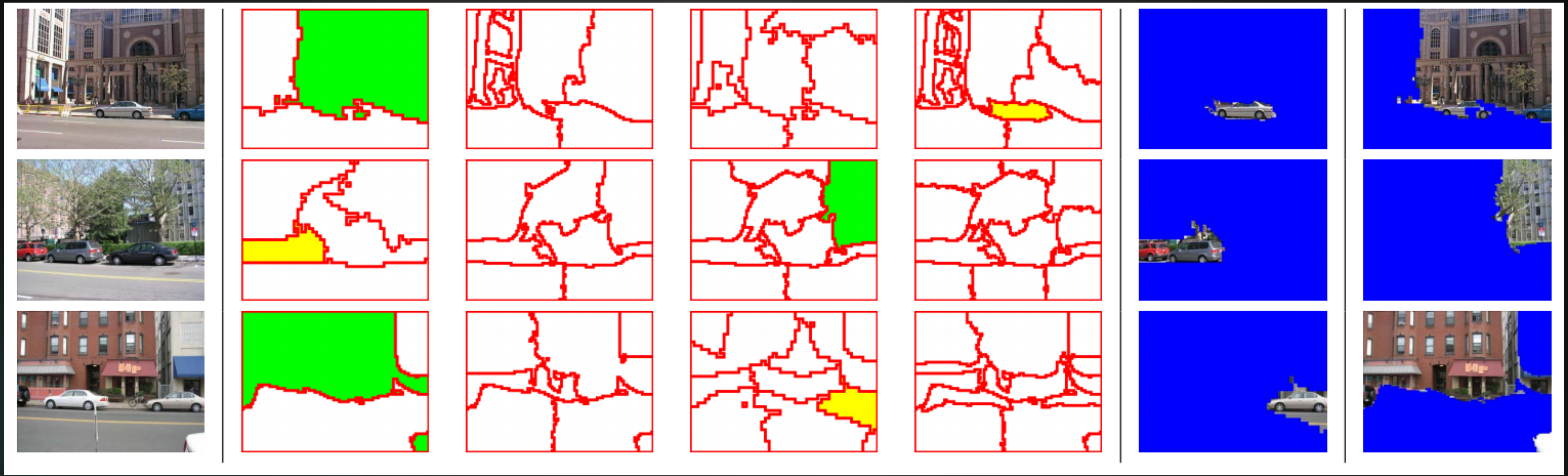
- Visual Words
  - Visual Synonyms and Polysemy





- No spatial information

Multiple Segmentations

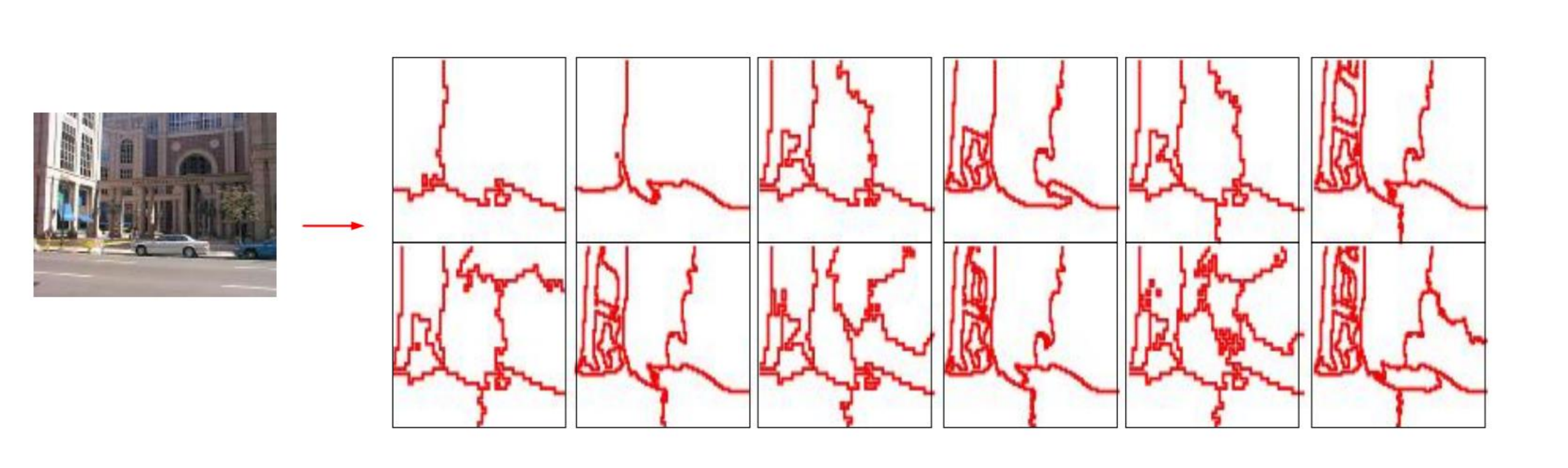


- None of the segmentations are perfect
- There are some segments that are okay
- “all good segments are alike, each bad segment is bad in its own way.”

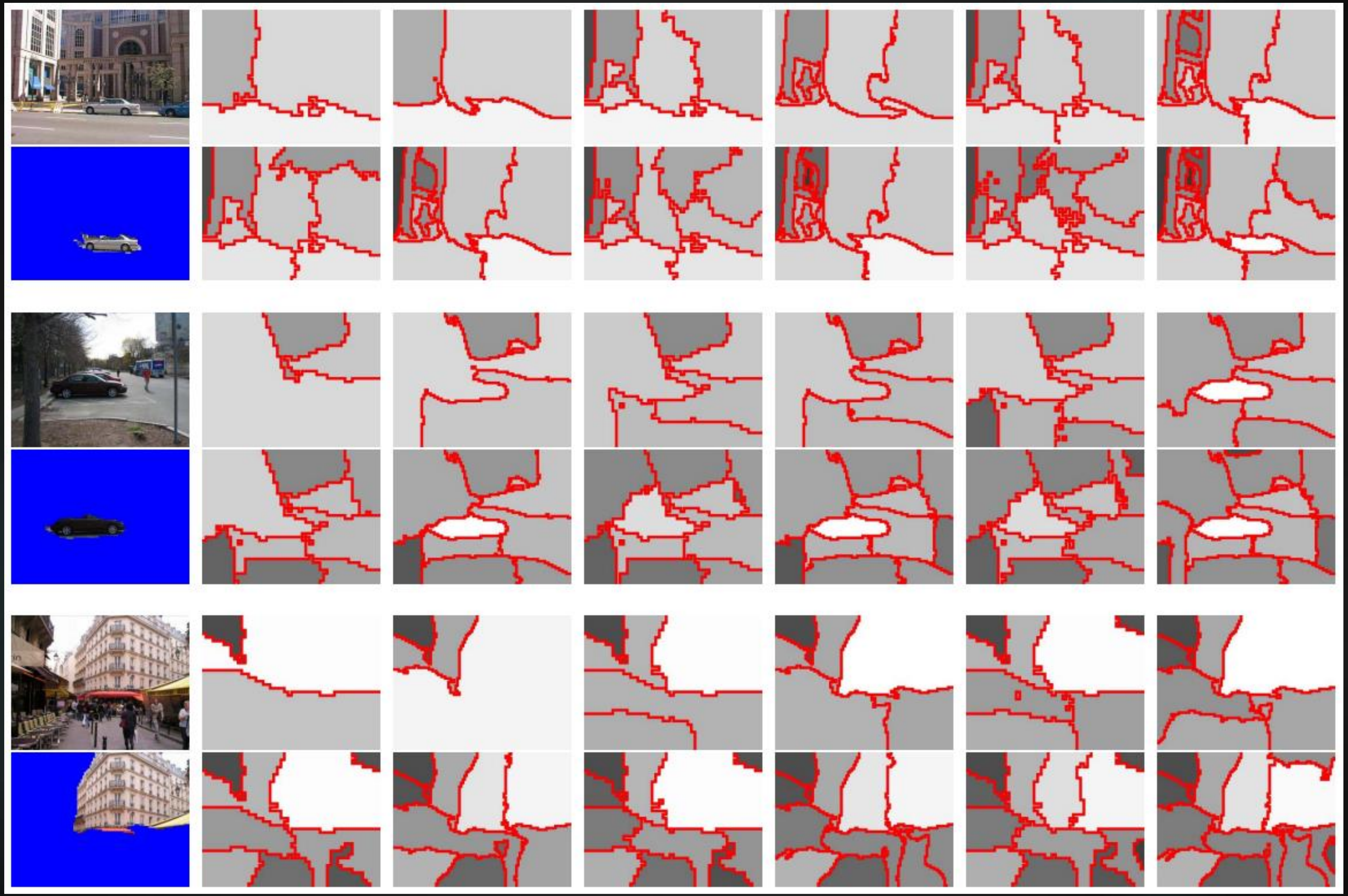
Algorithm

- Starting with a set of unlabeled images
  1. Compute multiple candidate segmentations using Normalized-Cuts
  2. Compute a histogram of visual words for each segmentation
  3. Perform topic discovery using Latent Dirichlet Allocation treating each segment as a document
  4. For each topic sort all segments by how well they describe the topic

Multiple Segmentations

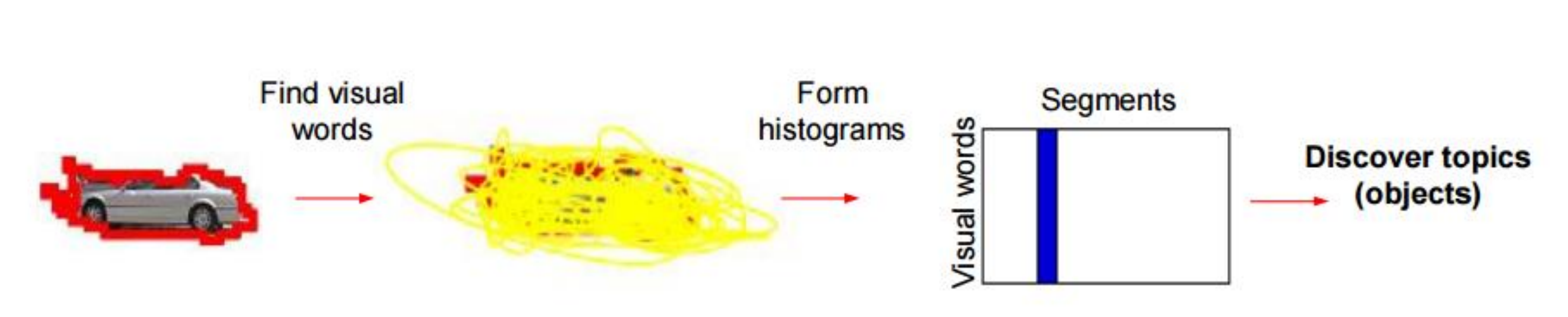


- Segmentation algorithm isn't the most important
- Use normalized cuts:
  - Vary number of segments and image scale



Topic Discovery

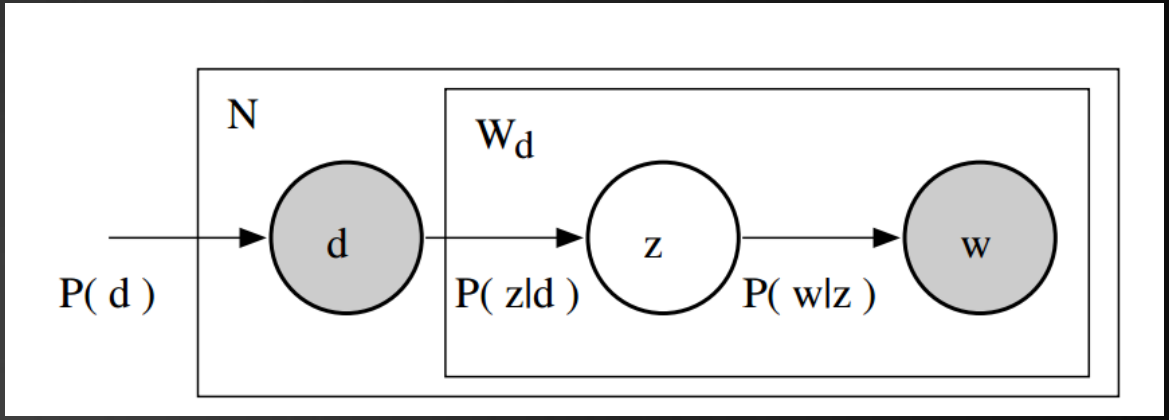
- Representing Segments

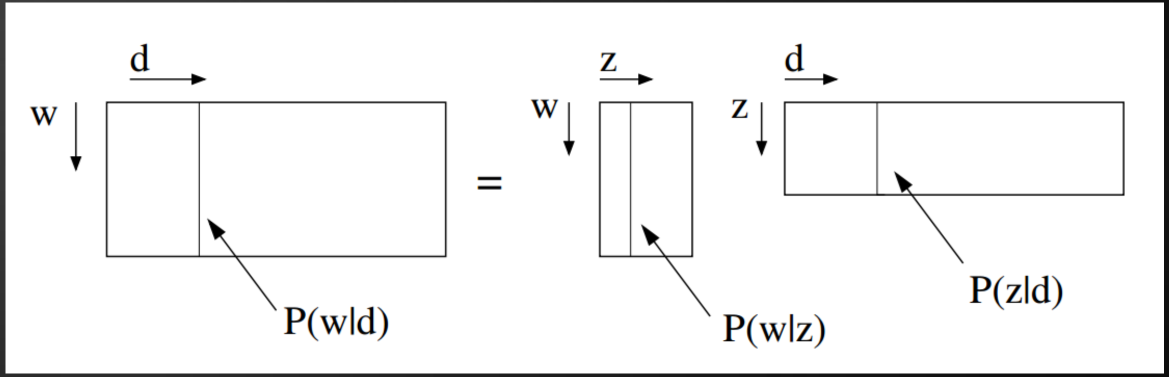


- Use a standard statistical text analysis technique
  - Probabilistic Latent Semantic Analysis
  - Latent Dirichlet Allocation
  - Introduce topic as a latent variable

pLSA

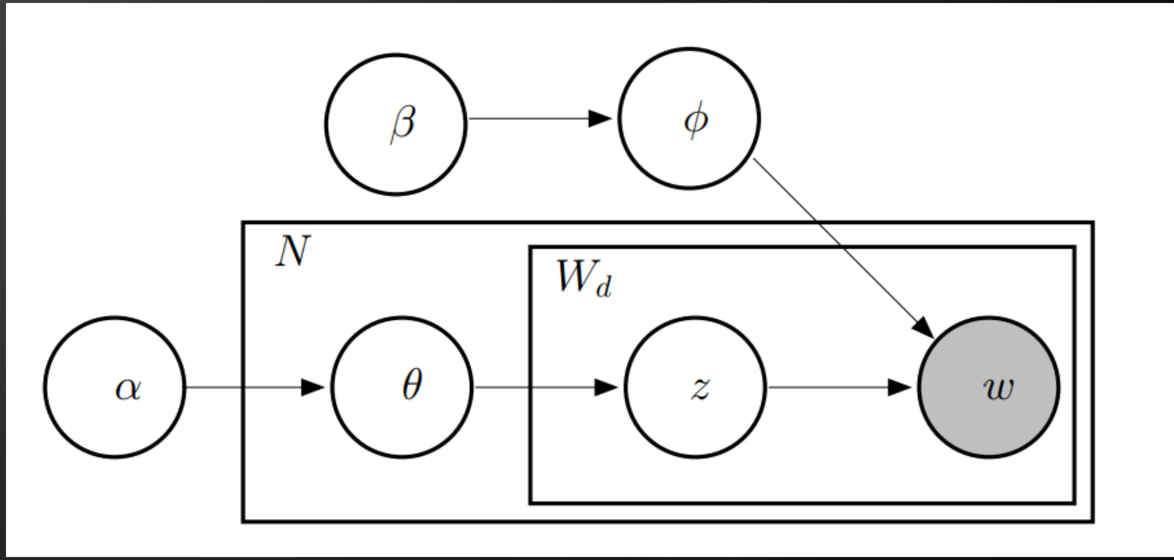
- $P(w_i|d_j) = \sum_{k=1}^{k} P(z_k|d_j)P(w_i|z_k)$
- Convex combination of topic vectors
- Each document is modeled as a mixture of topics





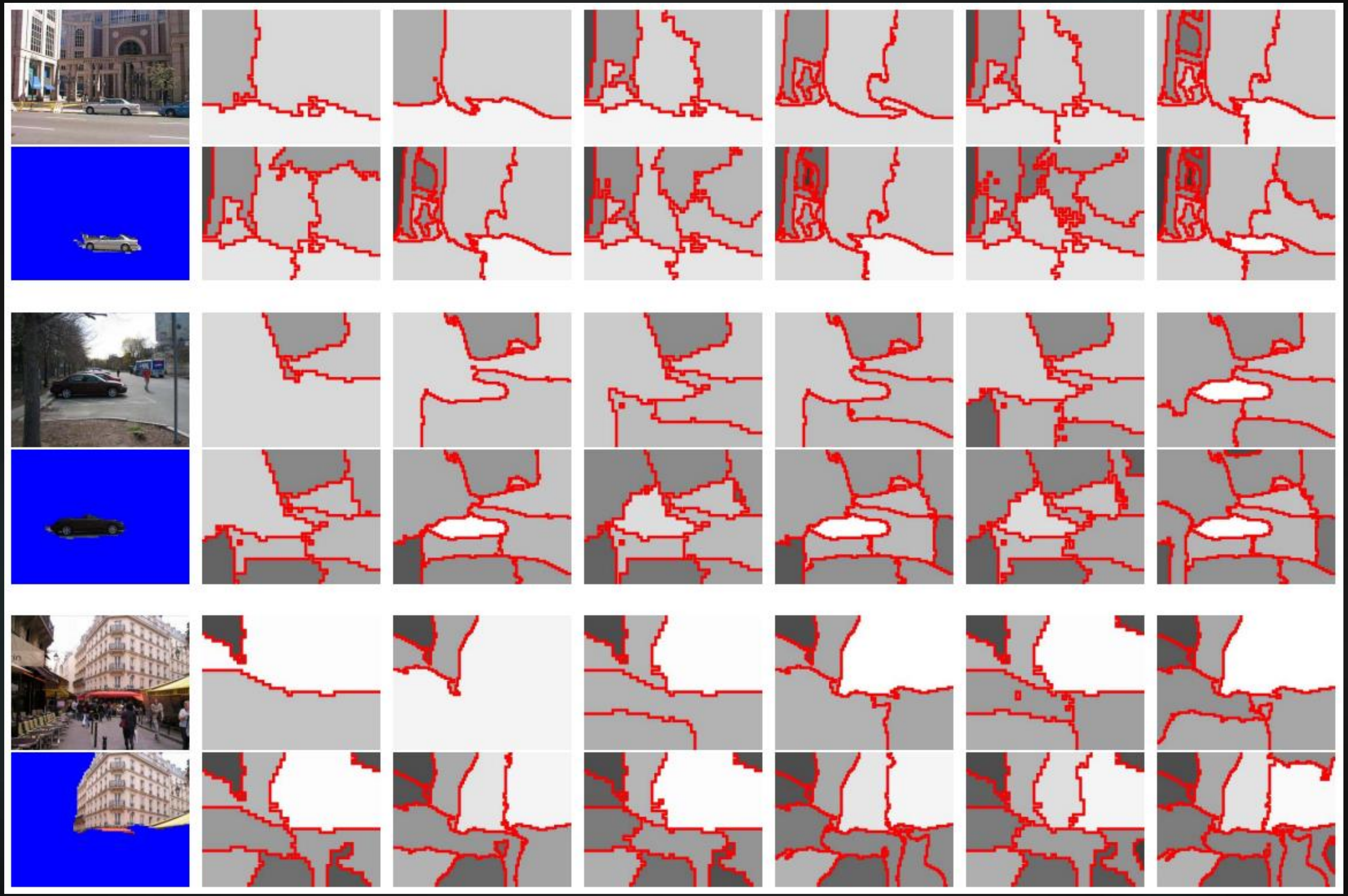
Latent Dirichlet Allocation

- $P(w|\phi, \alpha, \beta) = \int \sum p(w|z, \phi) p(z|\theta) p(\theta|\alpha) p(\phi|\beta) d\theta$
- LDA treats $P(z|d)$ as latent random variables
- Samples weights from a Dirichet distribution
  - Reduces overfitting
  - $P(\theta|\alpha)$ and $P(\phi|\beta)$
    - Parameterized
  - Use Gibbs sampling



Sorting Segments

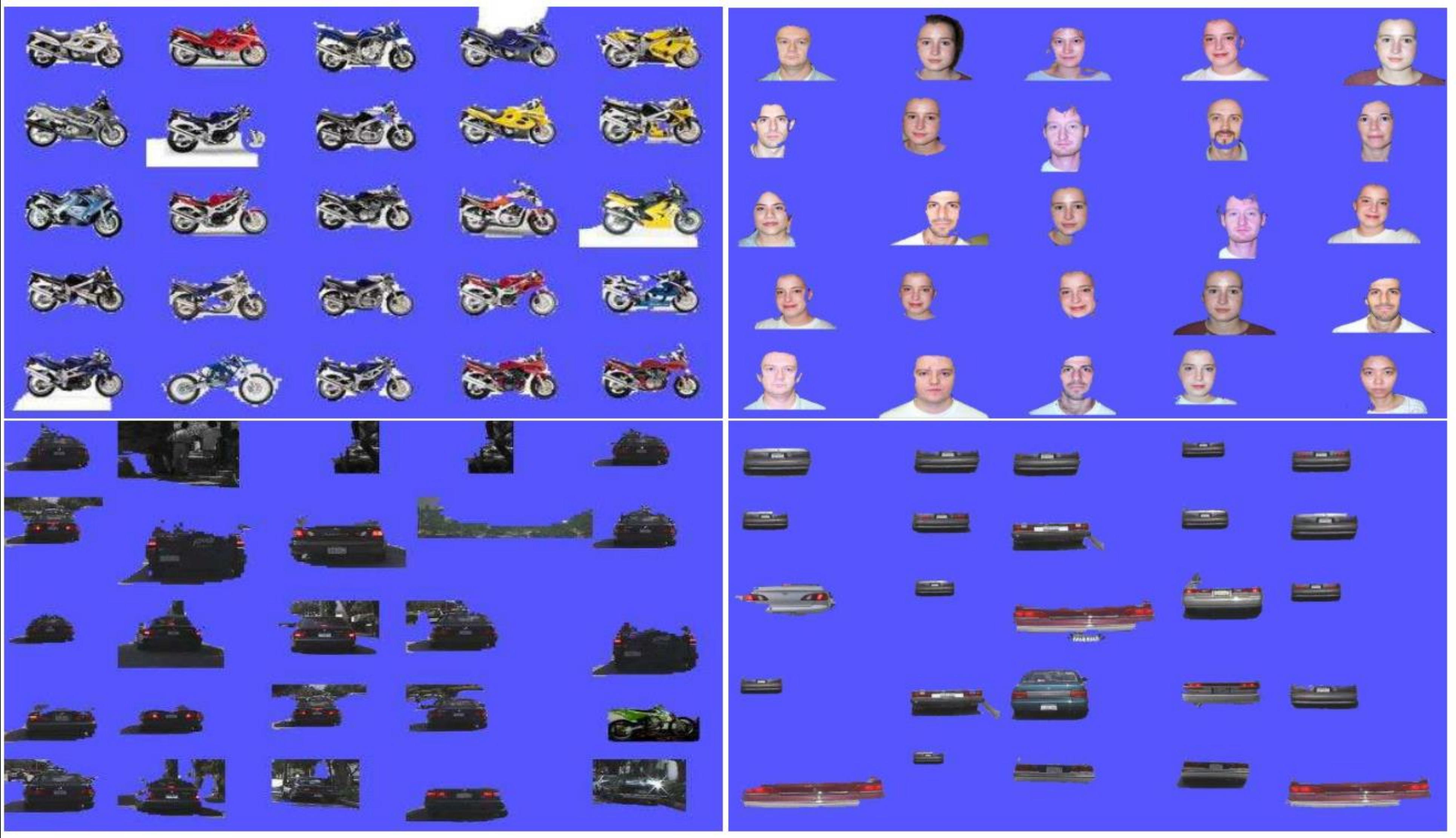
- How to find good segments in each topic?
- Compare visual words histograms with the learned multinomial weights
- Sort based on Kullback-Leibler divergence
  - $D_{KL}(P(w|s,\phi_s)||P(z,\phi_t))$



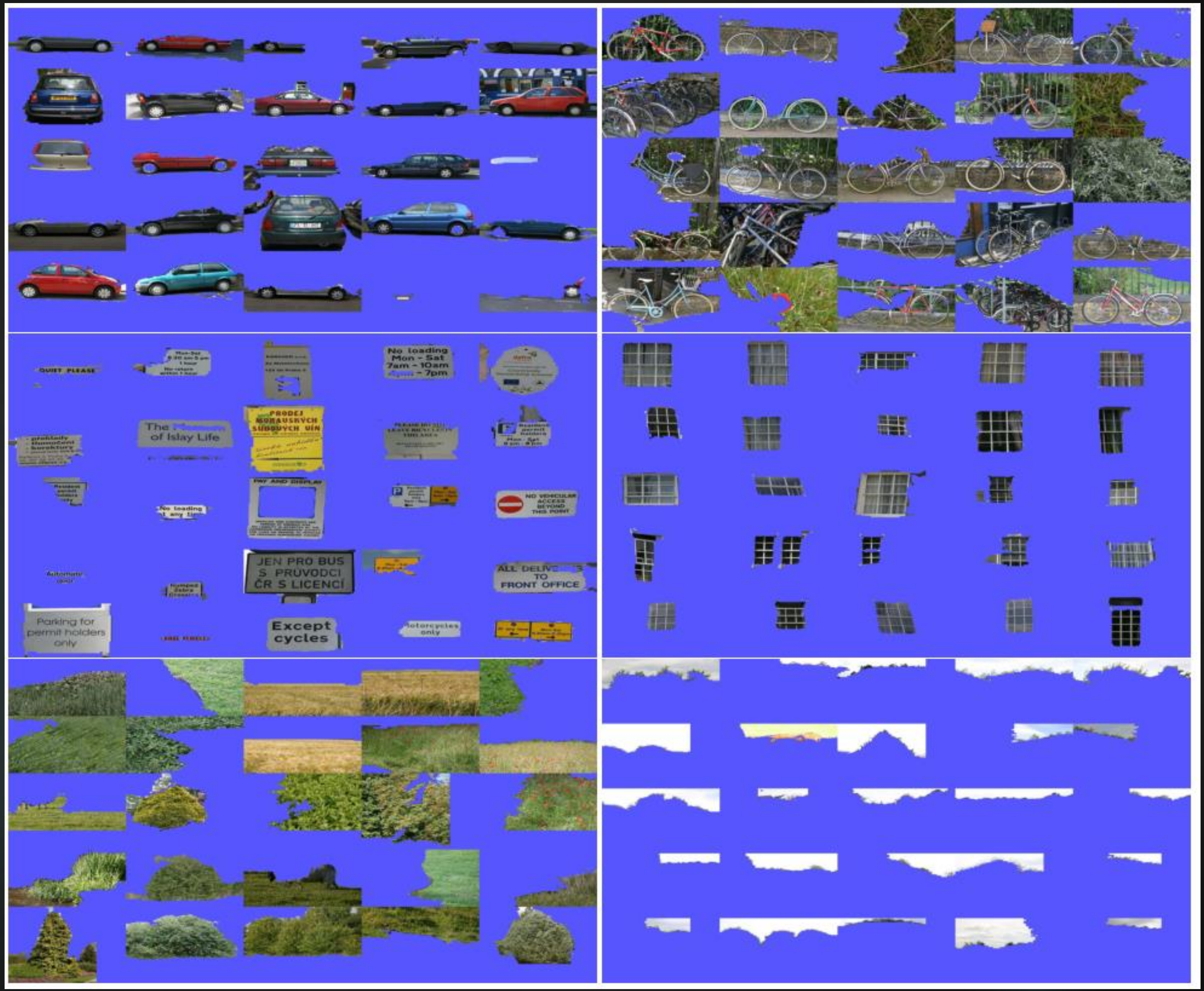
Results

- Datasets
  - Caltech
    - 4090
    - 4 categories
  - MSRC
    - 4325 images
    - 23 objects and scenes
  - LabelMe
    - 1554 images
    - Cars, buildings, trees

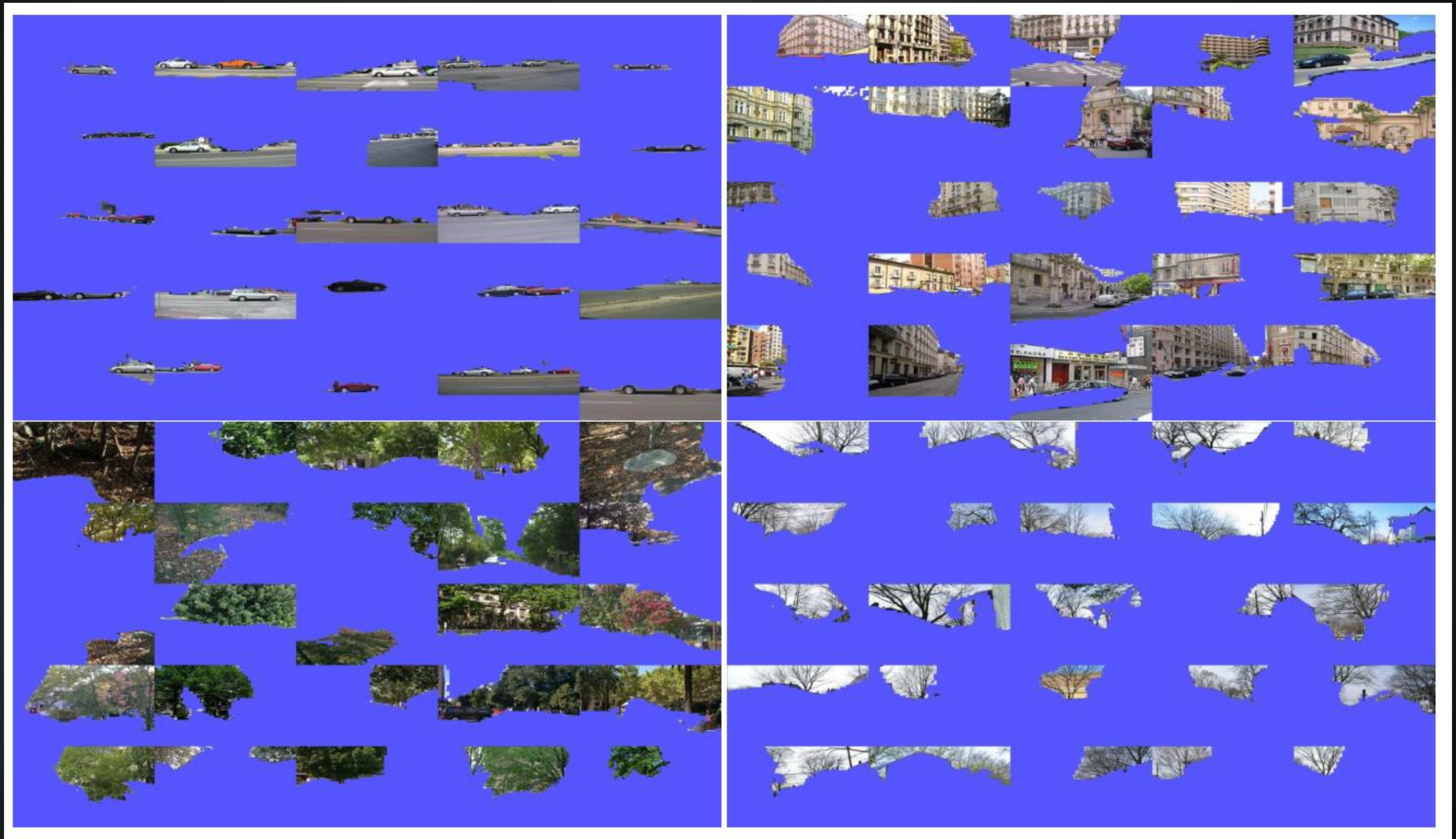
Caltech: 10 Topics



MSRC: 25 Topics



LabelMe: 20 Topics



Raw Numbers

Method	bicycles	cars	signs	windows
(a) Mult. seg. LDA	0.69	0.77	0.43	0.74
(b) Mult. seg. pLSA	0.67	0.28	0.34	0.57
(c) Sing. seg. LDA	0.67	0.73	0.46	0.72
(d) No seg. LDA	0.64	0.85	0.40	0.74
(e) Chance	0.06	0.12	0.04	0.15

MSRC Precision

Method	buildings	cars	roads	sky
(a) Mult. seg. LDA	0.53	0.21	0.41	0.77
(b) Mult. seg. pLSA	0.59	0.09	0.16	0.77
(c) Sing. seg. LDA	0.55	0.29	0.32	0.65
(d) No. seg. LDA	0.47	0.16	0.14	0.16

LabelMe Segmentation Score

Discussion