PS3-34

On Understanding Visual Relationships by Visualizing Bag-of-Visual-Words Models

E-mail: kastnerm@murase.is.i.nagoya-u.ac.jp

Marc A. Kastner Ichiro Ide Yasutomo Kawanishi Takatsugu Hirayama Daisuke Deguchi Hiroshi Murase

Nagoya University

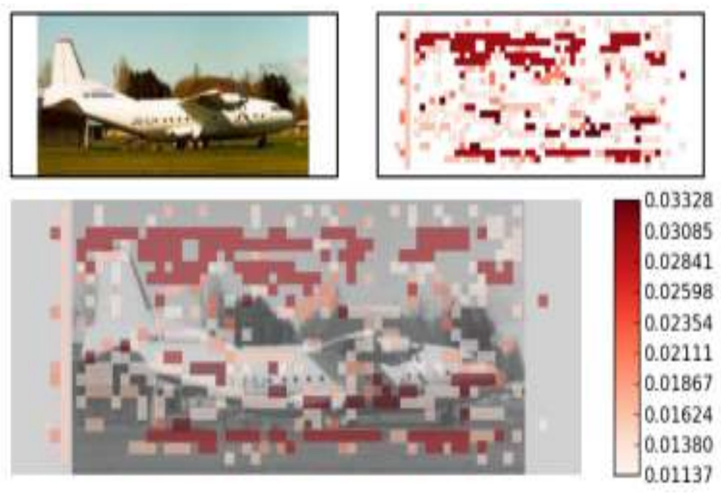
Motivation

- Background: Black-boxed machine learning algorithms demand for more data knowledge
  - For multi-modal applications, semantics and human perception needed to understand semantic gap
  - Try to understand how a computer understands

Newly established field "Explainable AI"[1] asks for more understanding of machine learned results

- Idea: Visualize similarities across related concepts
  - Show similar image regions
  - Show visual-spatial distribution of images

Related work



- Visualize contents of Bag-of-Visual-Words models[2]
  - Which image regions are crucial for correct classification?
  - Create probability heat maps for image regions

Interesting findings: Sometimes very unexpected image regions are most relevant for the classifiers

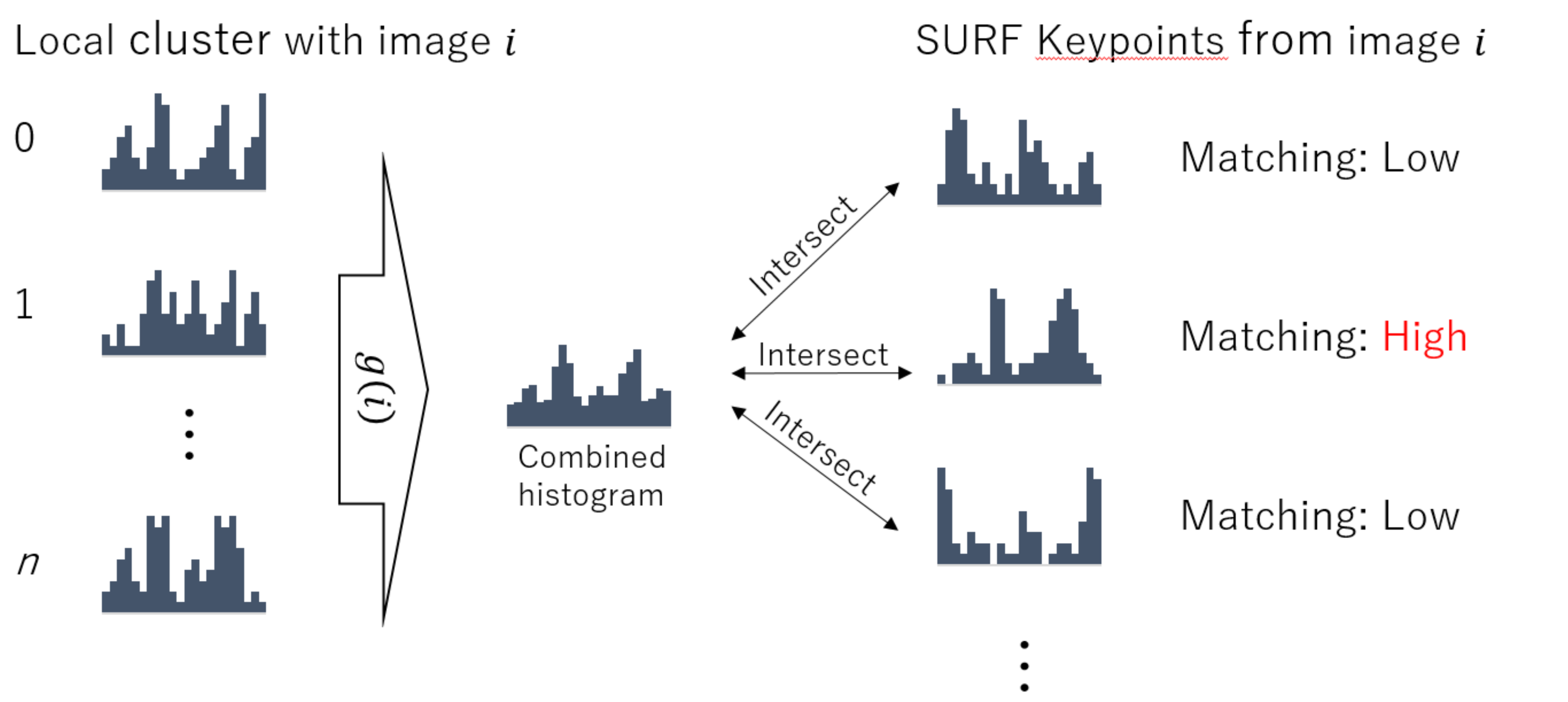
- Reconstruct Bag-of-Visual-Words models[3]
  - Identify which features are retained in a visual model
  - Highlights which regions were crucial for encoding

Highlight keypoints

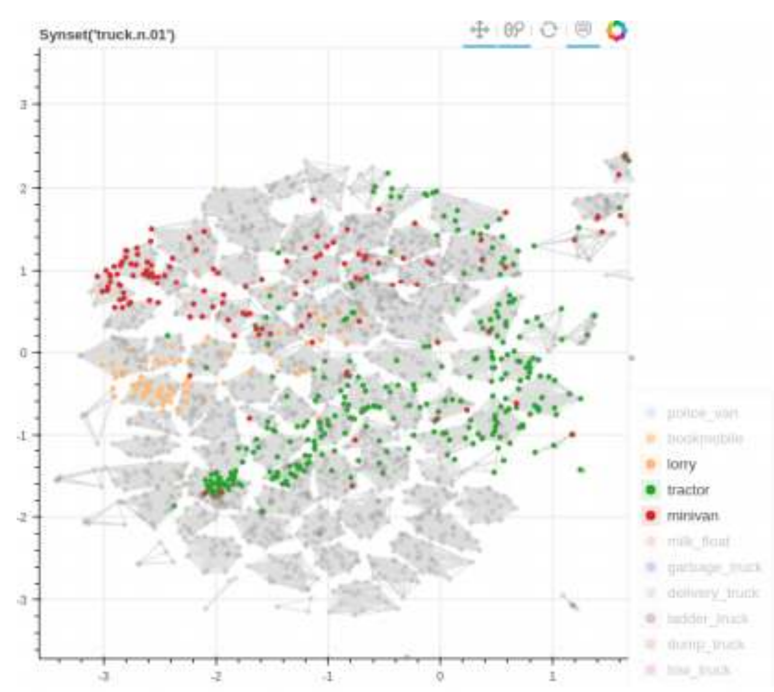
$$g(x) = \frac{\prod_{i=0}^{n}(f_i(x)+1)}{2^n}$$

- Find common keypoints

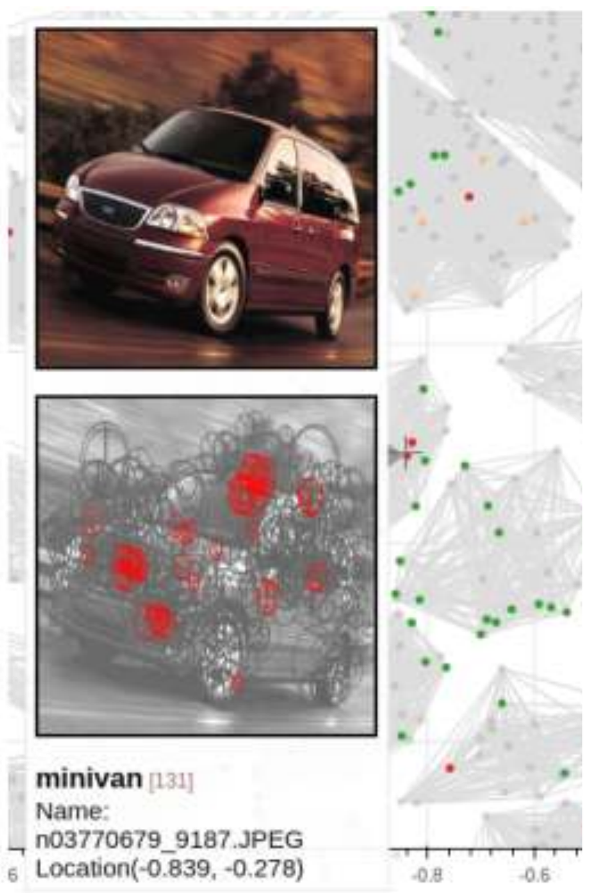
1. Extracting BoW histogram $f(x)$ for each image
2. Combine images of local cluster to histogram $g(x)$
3. For each image, intersect $g(x)$ with the histogram of each SURF keypoint to find visually similar keypoints in each image.



Visualization tool



- Visualize the visual model using UMAP[4] and Bokeh[5]
  - BoVW histogram decides location of each sample
- Highlight images based on sub-concept
  - Visually narrow sub-concepts are clustered in a corner
  - Visually open sub-concepts are spatially scattered



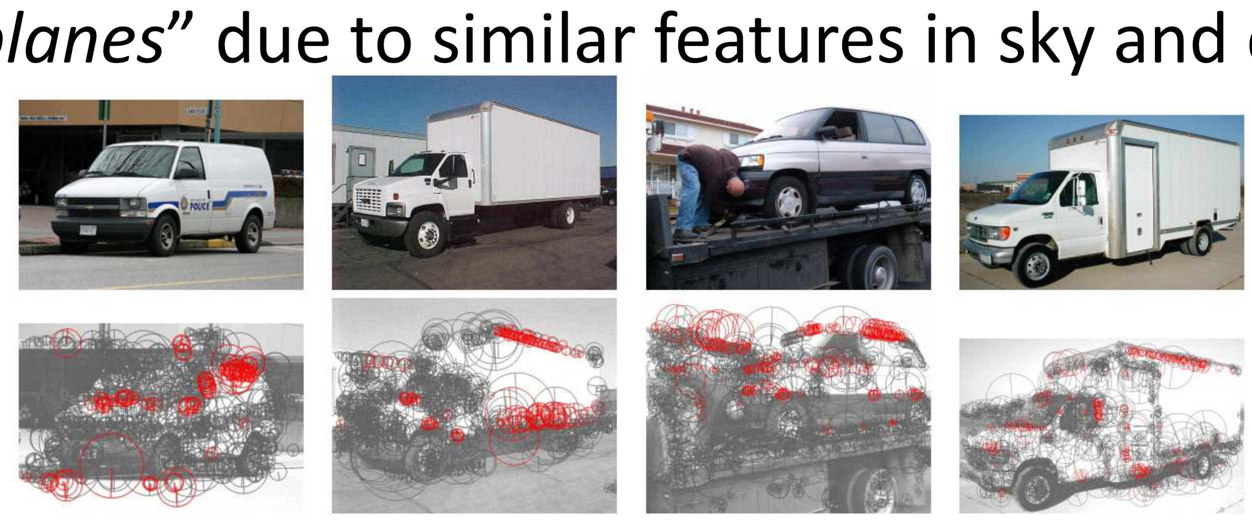
Labeling can be set to children nodes or sibling sub-trees

- Mouse-over shows extended information on data samples
  - Raw image, sub-concept name, imageability labels (if available)
  - Bottom image highlights SURF keypoints visually common between neighboring images

Tool is designed to browse ImageNet concepts based on visual characteristics in sub-concepts

Discussion

- Find visually-related groups even if they belong to different concepts
  - In "*street vehicles*", "*trucks with company logos*" are clustered close to "*cars with text*" due to text patterns
- Importance of backgrounds
  - In "*vehicles*", "*helicopters*" are clustered close to "*airplanes*" due to similar features in sky and clouds



Spatial clustering finds visual semantics in a unsupervised way

Future work

- Correlate visual feature space to visual variety
  - Compare results to imageability and concreteness
  - Infer imageability from area of visual feature space
- Cross comparison of ImageNet subtrees
  - Compare variety in car types to variety of plants
- Create live demo for interactive browsing
  - If possible, publically available though Web app

[1] W. Samek et al. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv 2017
[2] C. Hentschel and H. Sack. What image classifiers really see – visualizing bag-of-visual words models. MMM 2015
[3] H. Yue et al. Visualizing bag-of-words for high-resolution remote sensing image classification. J Appl Remote Sens 2016
[4] L. McInnes and J. Healy. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv 2018
[5] Bokeh development team. Bokeh: Python library for interactive visualization. 2014