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

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

Motivation
- For multi-modal applications, semantics and human perception needed to understand semantic gap
- Try to understand how a computer understands

Motivation
- Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv 2017
- What image classifiers really see – visualizing bag-of-visual words models. MMM 2015
- Uniform Manifold Approximation and Projection for dimension reduction. arXiv 2018
- Bokeh: Python library for interactive visualization. 2014

Highlight keypoints
- 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.

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

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

Visualization tool
- Visualize the visual model using UMAP⁴ and Bokeh⁵
  - 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

Future work
- toolbox to browse ImageNet concepts based on visual characteristics in sub-concepts

Future work
- Create probability heat maps for image regions
  - “vehicles”, “helicopters” are clustered close to “airplanes” due to similar features in sky and clouds

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

Future work
- 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

Future work
- Spatial clustering finds visual semantics in an unsupervised way

Related work
- Visualize contents of Bag-of-Visual-Words models²
  - 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 models³
  - Identify which features are retained in a visual model
  - Highlights which regions were crucial for encoding