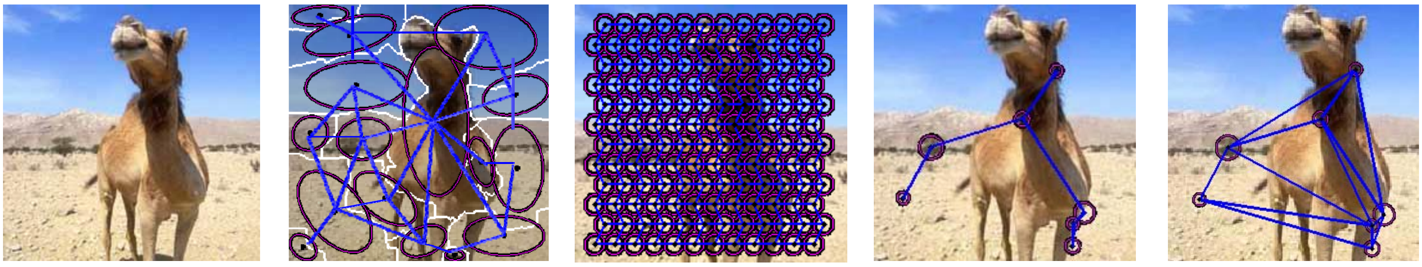
Structural Image Classification with Graph Neural Networks

Alyssa Quek

Supervisors: Zhiyong Wang, Jian Zhang and David Feng
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES



(a) Original Image (b) Region Adjacency Graph (RAG) using Edge Flow (c) Grid structure using dense sampling (d) Minimum Spanning Tree (MST) using Hessian-Laplace (e) Delaunay Triangulation using Hessian-Laplace

Figure 1: *Various structural image representations modelled as undirected graphs*

Image Classification:

the process of assigning a given image into one of several predefined categories.

1. Project Aims

- Exploit the patterns and relationships between regions by representing images as undirected graphs.
- Employ the recently-proposed Graph Neural Networks [1] model to handle the classification of structured data:

2. Image Structure as Graphs

An image's structural context can be represented as a graph $G=\{N, E\}$, where N (nodes) correspond to interest regions and E (edges) correspond to the connections between two distinct regions. Figure 1 illustrates the structural representations considered.

Region Adjacency Graph (RAG):

- Segment regions based on colour and texture using Edge Flow [2].
- Fit ellipses and connect adjacent regions.

Grid structure:

- Dense sample every 16th pixel, offsetting by half every odd row.
- Join NSEW neighbours where available.

Minimum Spanning Tree (MST):

- Detect regions with scale-invariant blob detector, Hessian-Laplace [3]. Identifies around 200-3000 regions per image.
- MST incurs the lowest total cost in forming a connected path to all nodes.

Delaunay Triangulation:

- Same interest regions as MST.
- Delaunay triangulation for a set *P* of points is a triangulation such that no point in *P* is inside the circumcircle of any triangle

3. Node and Edge Labels

$n_i = (x_i, y_i, s_i)$ — l_{n_i} = SIFT desc.

$n_j = (x_j, y_j, s_j)$ — l_{n_j} = SIFT desc.

$l_{(n_i, n_j)}$ = (norm. xy distance, norm. scale diff., norm. feat. distance)

Figure 2: *Node and edge information diagram*

n is a node with center co-ordinates (x, y) and detected at scale s.
l_n is a node label and $l_{(n,m)}$ is an edge label.
SIFT = Scale Invariant Feature Transform [4] descriptor.

4. Filtering Nodes detected with Hessian-Laplace

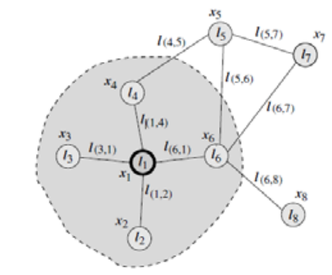
(1) Form a representative node set using *k*-means clustering in feature space. This is performed on a subset of images from the entire dataset.
(2) Apply further heuristics to remove outliers and non-meaningful clusters.
(3) For each image, associate all nodes to an appropriate cluster centroid.
(4) Keep the node set in which each node is closet to its associated centroid. This yields a filtered node set containing at most *k* nodes.

(a) Unfiltered regions (b) Filtered regions

Figure 3: *Images before and after filtering Hessian-Laplace regions*

5. Classification with Graph Neural Networks (GNN)

A state x_n is attached to node n, based on the information contained in the neighbourhood.



$$x_1 = f_w(\underbrace{l_1, l_{(1,2)}, l_{(3,1)}, l_{(1,4)}, l_{(6,1)}}_{l_{co[1]}}, \underbrace{x_2, x_3, x_4, x_6}_{x_{ne[1]}}, \underbrace{l_2, l_3, l_4, l_6}_{l_{ne[n]}})$$

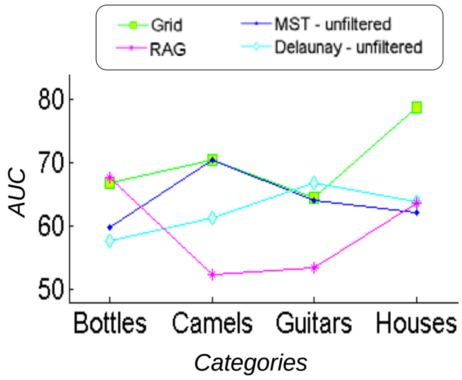
Figure 4: *The state of node* 1 *(x_1) depends on its label (l_1), the labels of its edges ($l_{co[n]}$), the states of neighbouring nodes ($x_{ne[l]}$) and their labels ($l_{ne[n]}$).*

The node state contains a representation of the category and can be used to produce a decision output value o_n. At every iteration t, the state of $x_n(t+1)$ is calculated using some function f_w. Finally, the output o_n is calculated using another function g_w. In the GNN model, f_w and g_w are implemented by feedfoward neural networks.

6. Results

We perform a repeated holdout on 350 images using the one-against-all approach.

GNN Structural Image Classification Performances



	RAG	Grid	MST Unfilt.	Delau Unfilt.	MST Filt.	Delau Filt.
Ave AUC	59.35	70.21	64.16	62.43	TBA	TBA

7. Future Work

- More features: invariant colour descriptors
- Combine various local invariant detectors to form a more representative set of regions
- Evaluate on a larger dataset with more image categories
- Better centroid proposals and try other clustering techniques: hierarchical k-means.
- More complex graph structures

8. References

[1] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.
[2] Ma, W.-Y. and Manjunath, B. Edgeflow: a technique for boundary detection and image segmentation. *IEEE Transactions on Image Processing*, 9(8):1375 –1388, August 2000.
[3] Mikolajczyk, K. and Schmid, C. Scale & affine invariant interest point detectors. *International Journal on Computer Vision*, 60(1):63–86, 2004.
[4] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2):91–110, 2004.

THIS RESEARCH IS SPONSORED BY
BMIT NICTA