

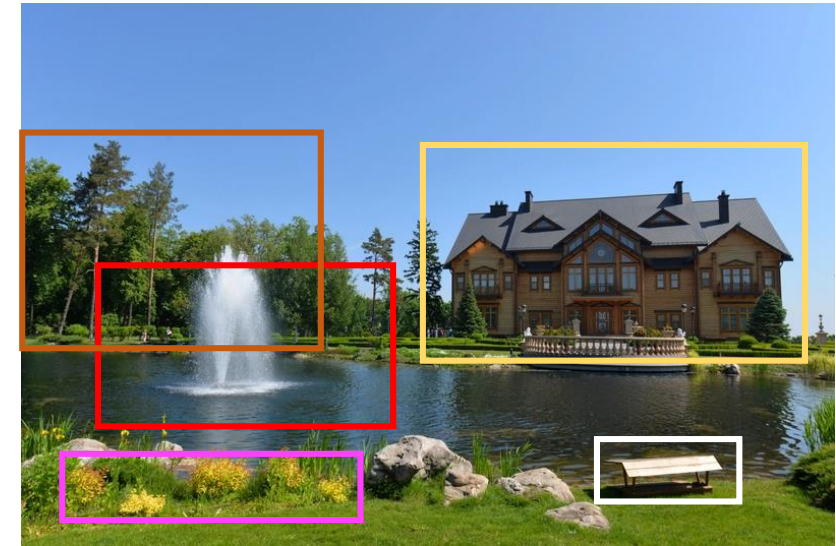
# Learning Semantic-Specific Graph Representation for Multi-Label Image Recognition

Chae Hun Shin

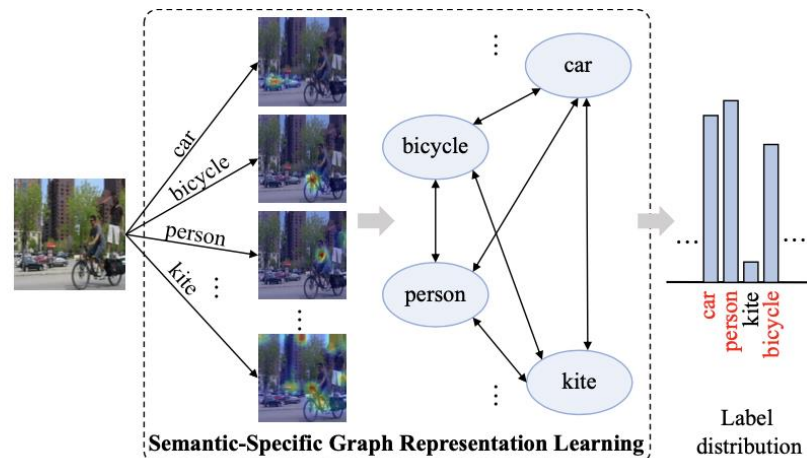
Seoul National University

# Multi-Label Image classification

- Identifies all objects in Image
  - Employing object localization
  - Resort to visual attention networks.



- Most of existing algorithms do not consider the dependencies between categories or regions.
- Modeling the dependency between semantics of categories with GNN.



# Category feature extraction

- Extract feature map of images
- Extract semantic embedding vector for each categories
- Semantic guided attention mechanism
  - Guide to focusing more on the semantic-aware regions.
  - Attention weights for each position
- Final feature map of images for each class

$$\mathbf{f}^I = f_{cnn}(I),$$

$$\mathbf{x}_c = f_g(w_c),$$

$$\tilde{\mathbf{f}}_{c,wh}^I = \mathbf{P}^T \left( \tanh \left( (\mathbf{U}^T \mathbf{f}_{wh}^I) \odot (\mathbf{V}^T \mathbf{x}_c) \right) \right) + \mathbf{b},$$

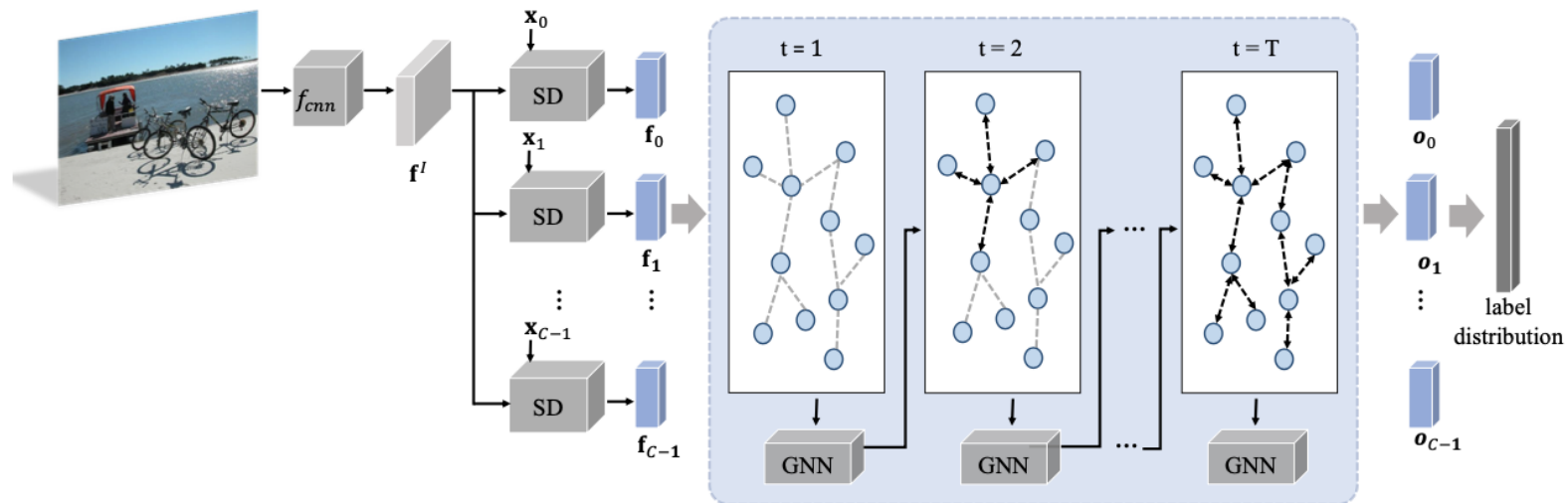
$$\tilde{a}_{c,wh} = f_a(\tilde{\mathbf{f}}_{c,wh}^I).$$

$$a_{c,wh} = \frac{\exp(\tilde{a}_{c,wh})}{\sum_{w',h'} \exp(\tilde{a}_{c,w'h'})}.$$

$$\mathbf{f}_c = \sum_{w,h} a_{c,wh} \mathbf{f}_{c,wh}$$
$$\{\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_{C-1}\}$$

# Graph construction

- Construct the graph  $\mathcal{G} = \{V, A\}$  based on the [statistical label co-occurrence](#).
  - Node set  $V = \{v_0, v_1, \dots, v_{c-1}\}$  represents each category.
  - Edge set  $A = \{a_{00}, a_{01}, \dots, a_{(c-1)(c-1)}\}$ , where  $a_{cc'}$  means the probability of the existence of object belonging to category  $c'$  in the presence of object belonging to category  $c$ .
- Each hidden features of nodes are initialized with the feature vectors.  $\mathbf{h}_c^0 = \mathbf{f}_c$ .



# Graph propagation

- Propagate as aggregating message from its neighbor nodes.
  - Nodes **interact with each other** under the guidance of the statistical label co-occurrence.

$$\mathbf{a}_c^t = \left[ \sum_{c'} (a_{cc'}) \mathbf{h}_c^{t-1}, \sum_{c'} (a_{c'c}) \mathbf{h}_c^{t-1} \right].$$

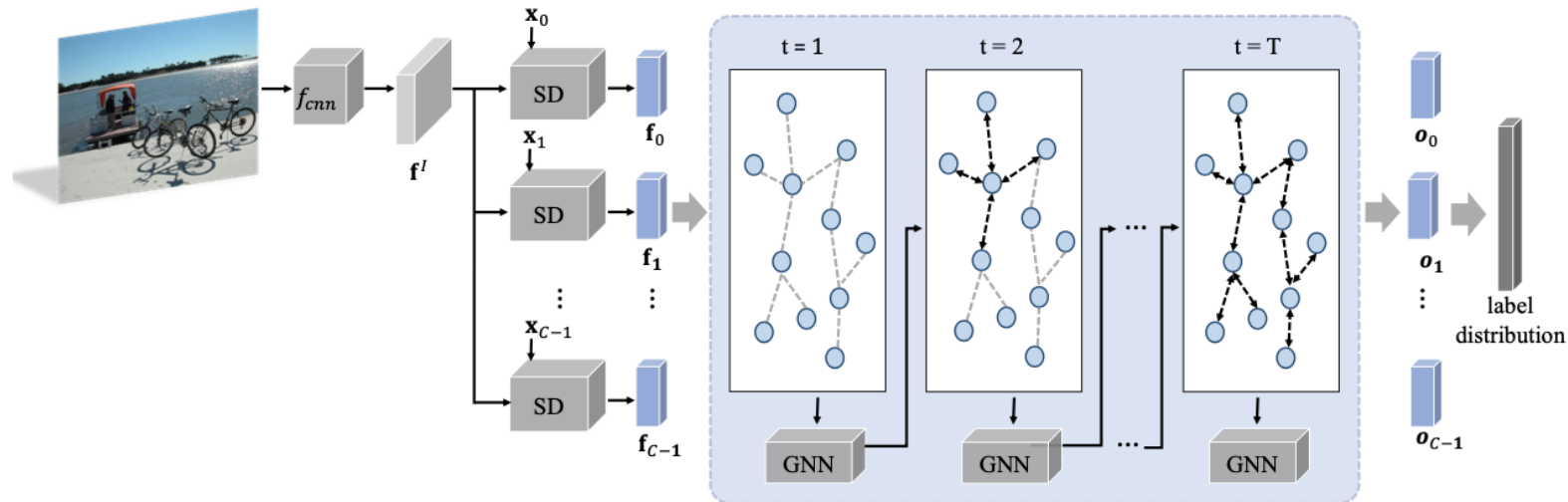
- Updates the hidden states of nodes via a gated mechanism.

$$\mathbf{z}_c^t = \sigma(\mathbf{W}^z \mathbf{a}_c^t + \mathbf{U}^z \mathbf{h}_c^{t-1})$$

$$\mathbf{r}_c^t = \sigma(\mathbf{W}^r \mathbf{a}_c^t + \mathbf{U}^r \mathbf{h}_c^{t-1})$$

$$\widetilde{\mathbf{h}}_c^t = \tanh(\mathbf{W} \mathbf{a}_c^t + \mathbf{U}(\mathbf{r}_c^t \odot \mathbf{h}_c^{t-1}))$$

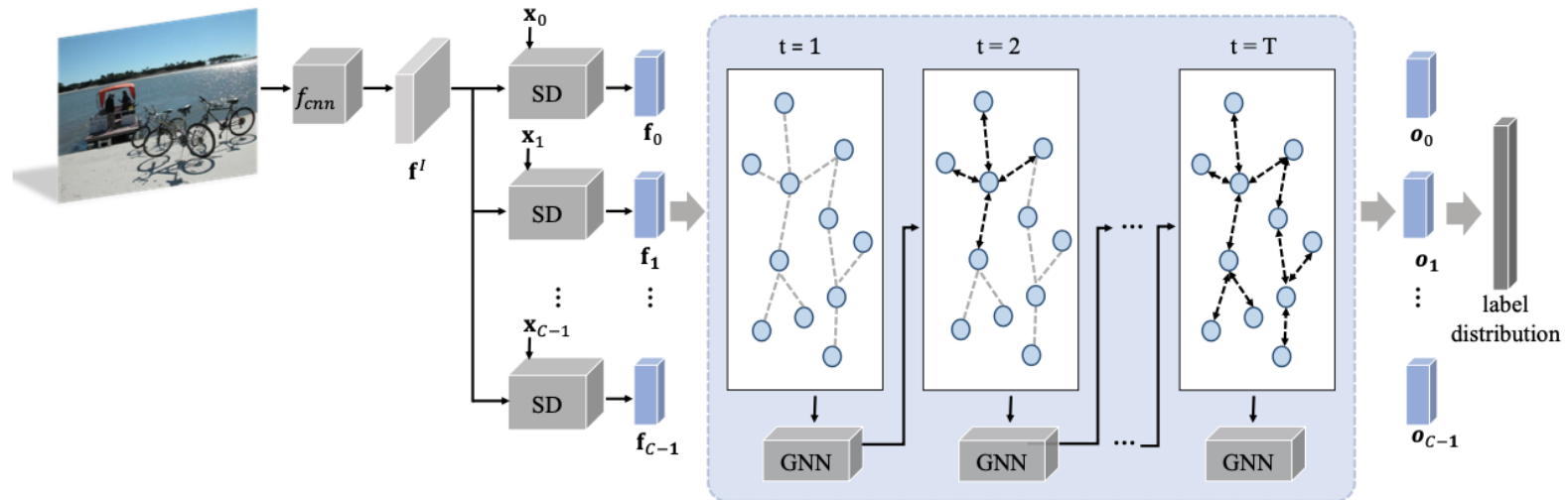
$$\mathbf{h}_c^t = (1 - \mathbf{z}_c^t) \odot \mathbf{h}_c^{t-1} + \mathbf{z}_c^t \odot \widetilde{\mathbf{h}}_c^t$$



# Multi-label classification

- Predict presence of category based on the final hidden states and initial hidden states

$$\mathbf{o}_c = f_o(\mathbf{h}_c^T, \mathbf{h}_c^0)$$
$$s_c = f_c(\mathbf{o}_c)$$



# Semantic-Specific Graph Representation Learning(SSGRL)

