

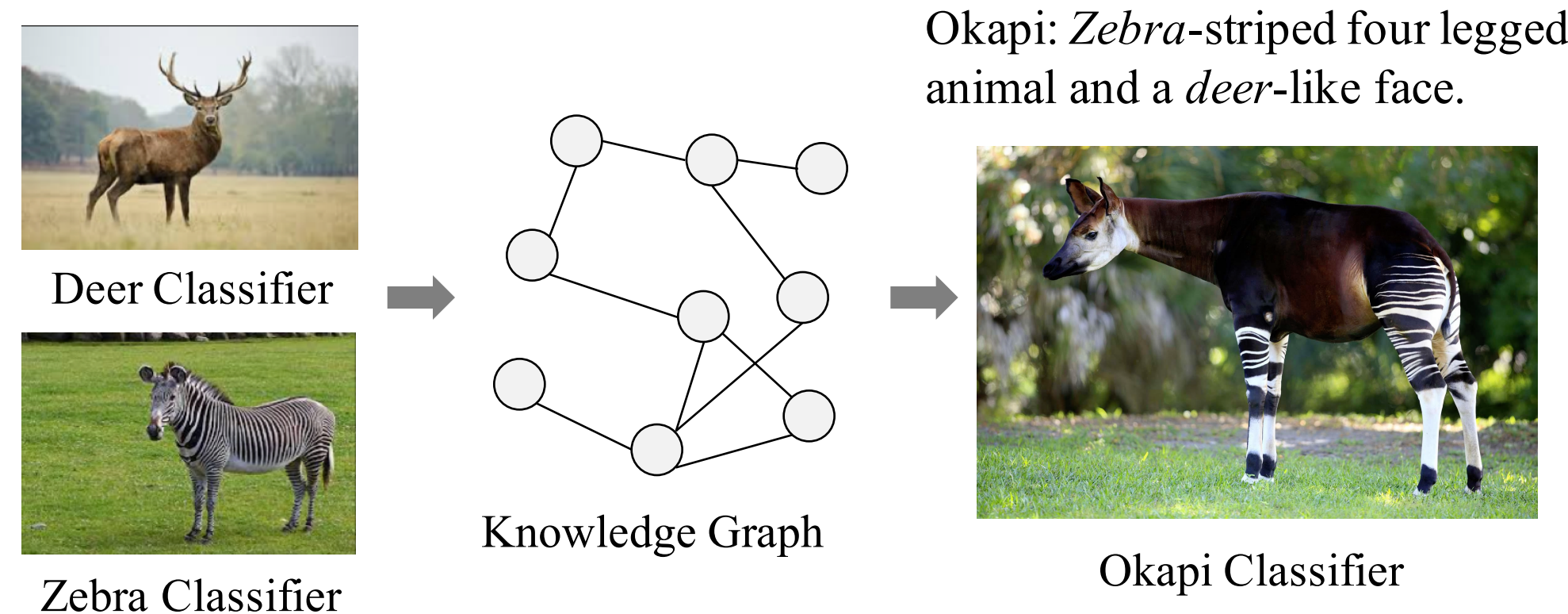


Motivations

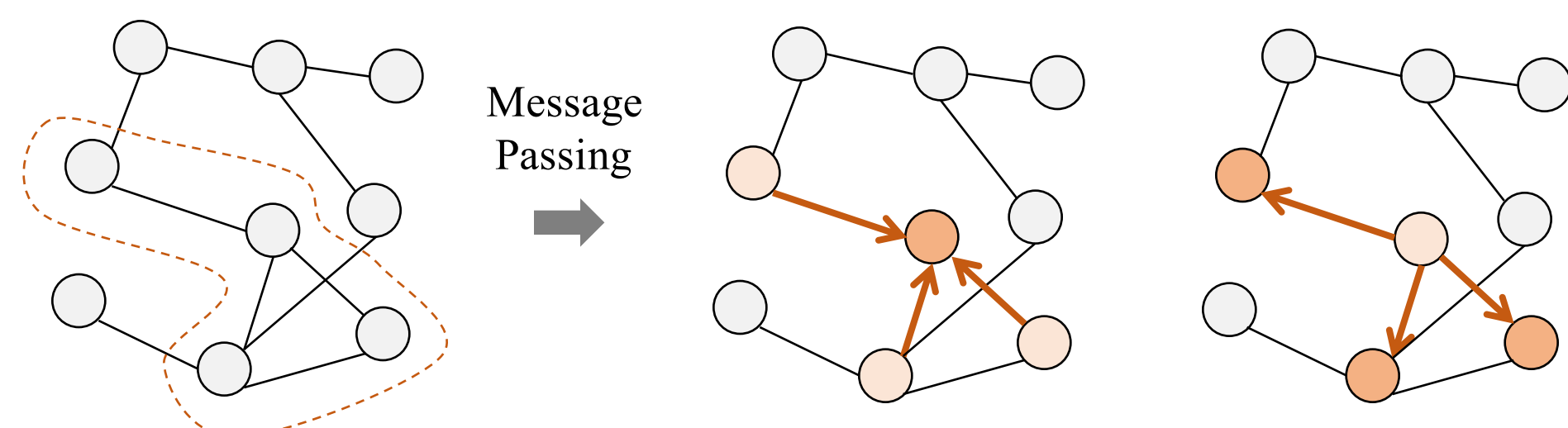
Zero-shot Recognition: Learning a visual classifier for a category with *zero* training images.

Key: Transfer knowledges from familiar classes to describe the unseen ones:

- *Implicit Knowledges:* Word embeddings
- *Explicit Knowledges:* Knowledge Graphs



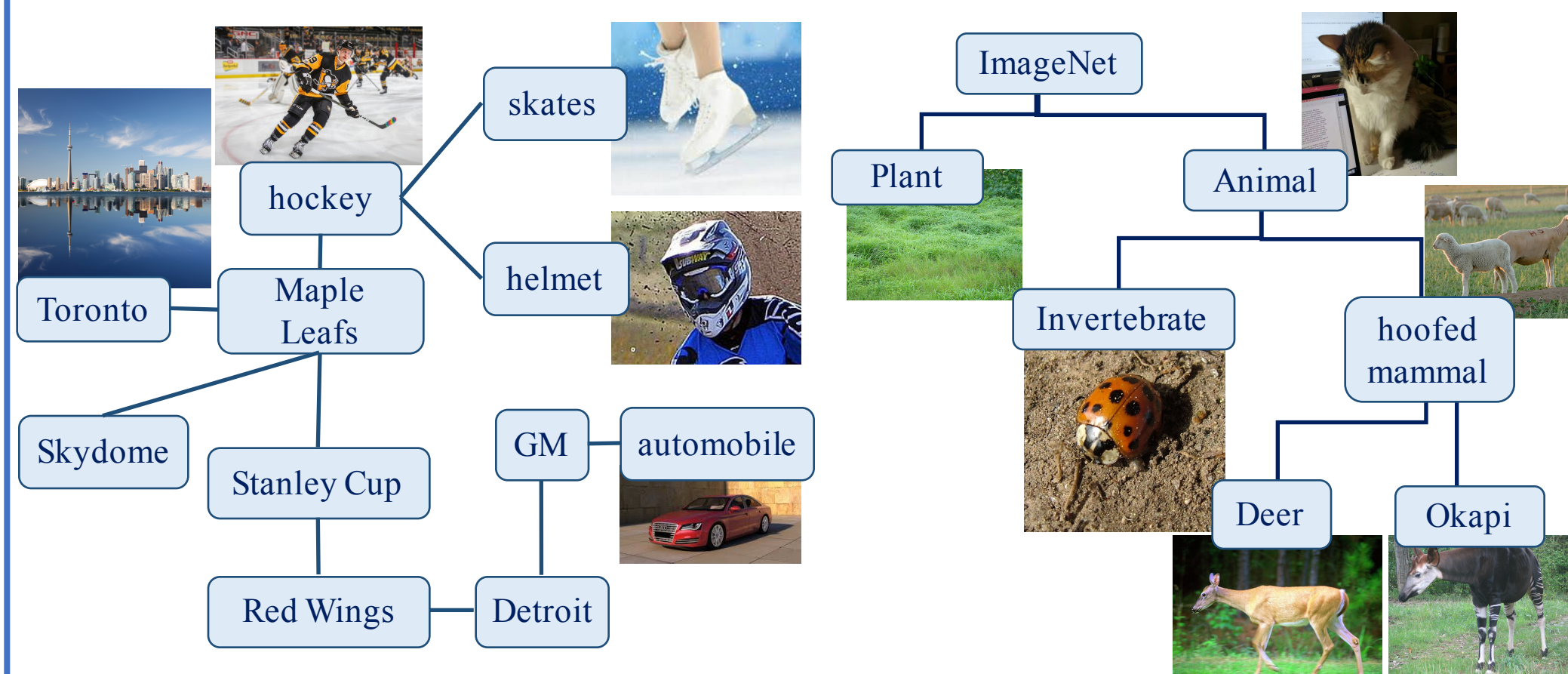
Graph Convolution Network:



Knowledge Graph Examples

NELL Knowledge Graph

WordNet Knowledge Graph

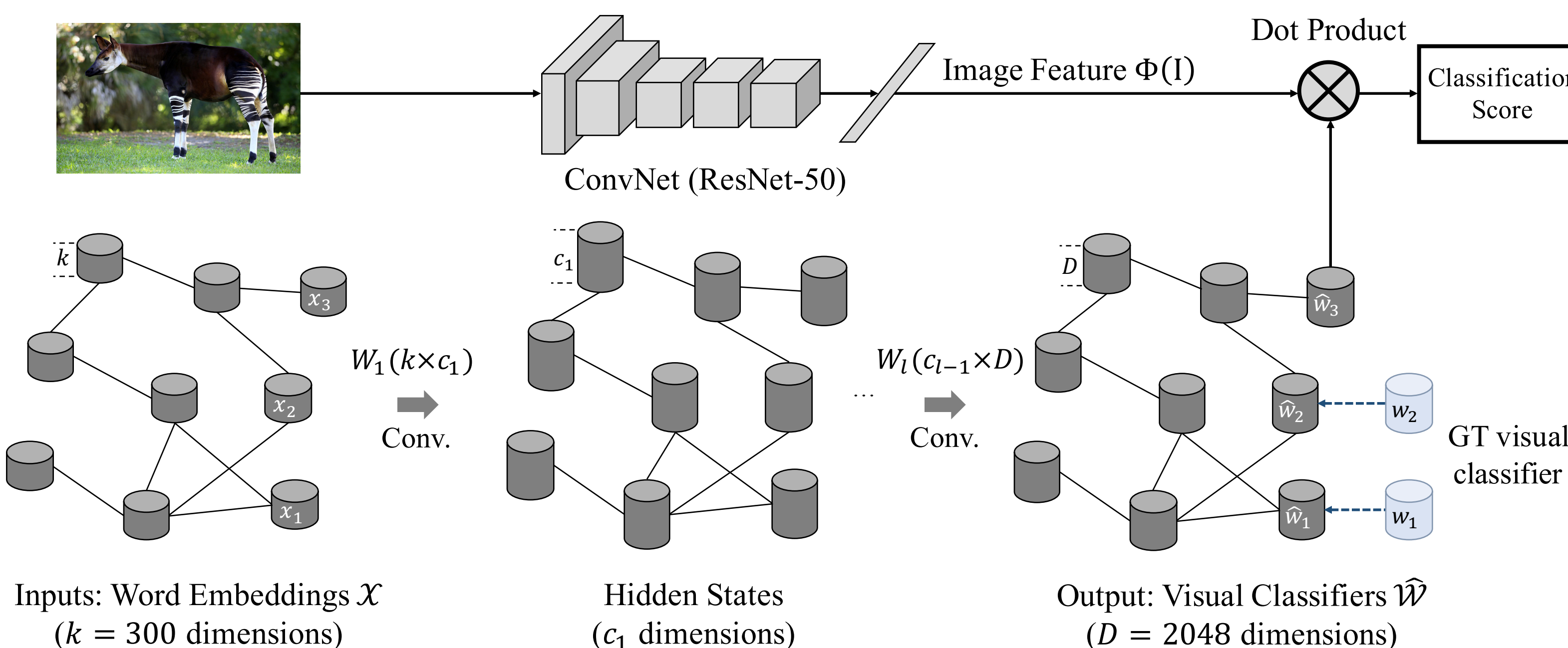


Approach

The framework contains 2 parts:

- ConvNet: *Input* the image and *output* the deep features of the image.
- Graph ConvNet: *Input* the word embeddings and *output* the visual classifiers.

Visual Classifiers: last FC layer weights of the ConvNet, fixed during training.

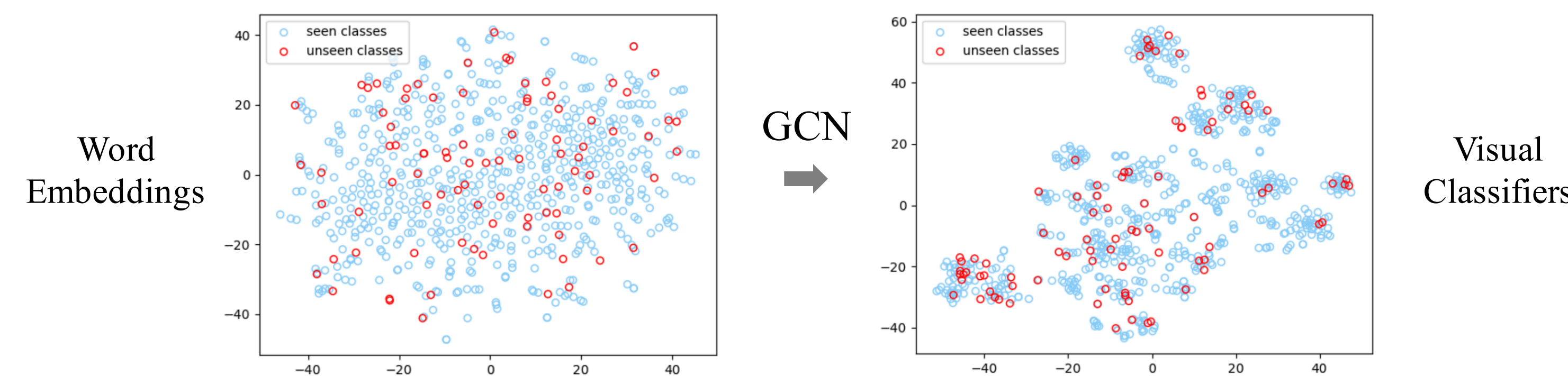


Formulations:

We assume there are N classes in total and only m classes out of them have training images.

- Training Loss:
$$\mathcal{L} = \frac{1}{m} \sum_{i=1}^m (\hat{w}_i - w_i)^2$$
- Zero-shot Inference: $\text{argmax}_i \langle \hat{w}_i, \Phi(I) \rangle, i = m + 1, \dots, N$
- Generalized Inference: $\text{argmax}_i \langle \hat{w}_i, \Phi(I) \rangle, i = 1, \dots, N$

t-SNE Visualizations:



Experiments on NEIL + NELL

There are 14K object nodes in the post-processed NELL graph.

We train the GCN model on 616 classes, generate novel classifiers and evaluate on other 88 classes.

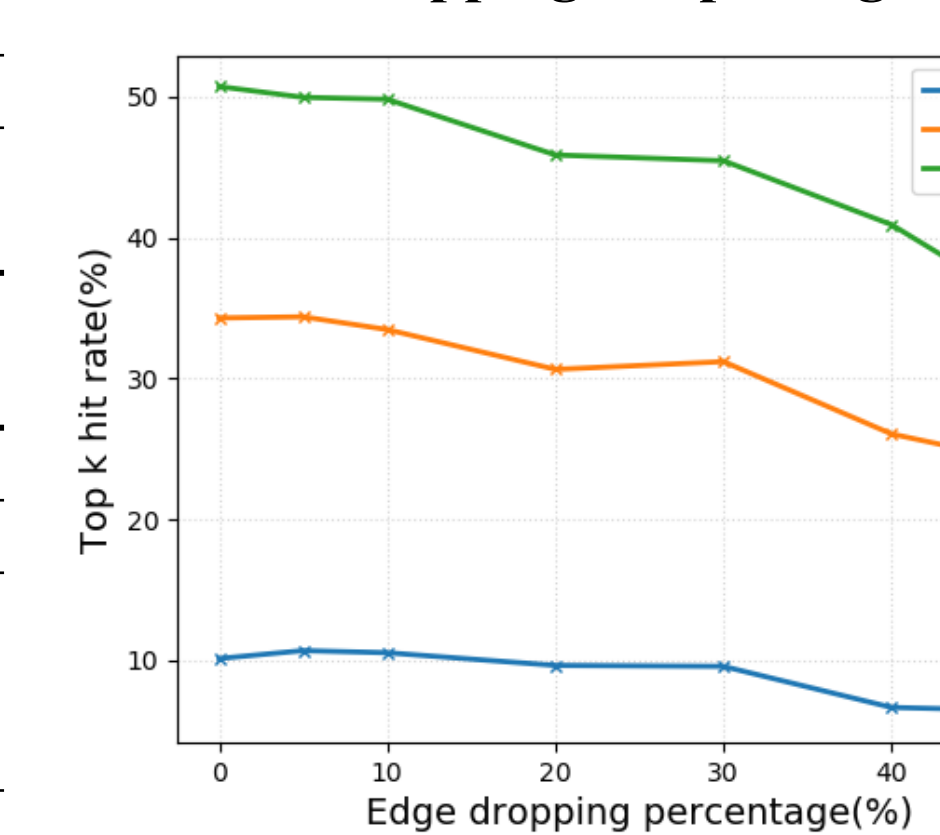
Zero-shot Classification Accuracy

Test Set	Model	Hit@k(%)		
		1	5	10
NEIL	ConSE	7.7	14.7	20.5
	Ours	10.8	33.7	49.0

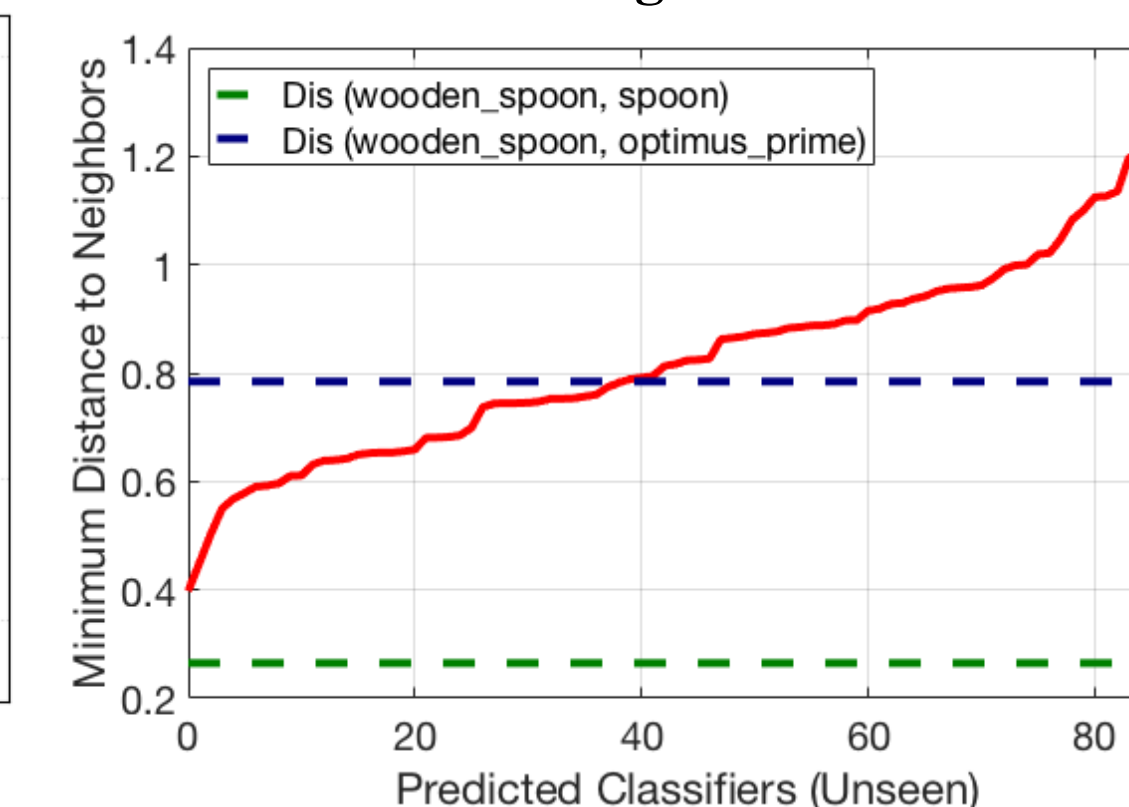
Random Graphs Do NOT Work

Graph	Hit@k(%)		
	1	5	10
Fully Connected	1.0	5.9	12.1
Star	1.1	4.8	9.7
Random	1.0	5.6	11.3
Knowledge Graph	10.8	33.7	49.0

Dropping Graph Edges



Minimum distance between predicted and training classifiers



Experiments on ImageNet + WordNet

There are 30K object nodes in the WordNet graph.

Training: the standard ImageNet-1K (classes) set.

Testing: the whole ImageNet set: 2-hops (1.5K classes), 3-hops (7.8K classes), All (21K classes).

Zero-shot Classification Accuracy

Test Set	Model	Hit@k(%)		
		1	5	10
2-hops	SYNC	10.5	28.6	40.1
	Ours	21.0	52.7	64.8
3-hops	SYNC	2.9	9.2	14.2
	Ours	4.3	14.2	20.4
All	SYNC	1.4	4.5	7.1
	Ours	1.9	6.4	9.3

Generalized Zero-shot Classification

Test Set	Model	Hit@k(%)		
		1	5	10
2-hops (+1K)	ConSE	0.1	24.3	29.1
	Ours	10.2	42.1	56.2
3-hops (+1K)	ConSE	0.3	7.3	10.0
	Ours	2.4	12.0	18.2
All (+1K)	ConSE	0.1	3.5	4.9
	Ours	1.1	5.4	8.3

2-hops Accuracy with Different Word Embeddings

Word Embedding	Model	Hit@k(%)		
		1	5	10
GloVe	DEM	7.8	17.2	21.2
	Ours	18.5	50.1	62.4
word2vec	DEM	9.8	27.8	37.6
	Ours	18.7	49.6	62.0
FastText	DEM	13.0	33.5	44.1
	Ours	18.3	51.1	63.4

