Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

Xiaolong Wang et al.
CVPR, 2018

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- Intro.

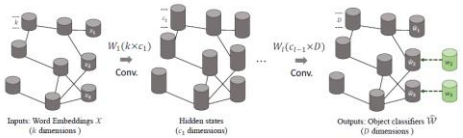
Figure 1. Can you find "okapi" in these images? Okapi is " zebra-striped four legged animal with a brown torso and a deer-like face". In this paper, we focus on the problem of zero-shot learning where visual classifiers are learned from semantic embeddings and relationships to other categories.

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- Intro.

Figure 1. Can you find "okapi" in these images? Okapi is " zebra-striped four legged animal with a brown torso and a deer-like face". In this paper, we focus on the problem of zero-shot learning where visual classifiers are learned from semantic embeddings and relationships to other categories.

Semantic embeddings
Knowledge graph



Graph Convolutional Network (GCN)

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN
  - semi-supervised entity classification 을 위해 제안된 모델
  - Ex. Training

    entity : dog, cat

    label : mammal

    Testing

    entity : lion

    label : ? (expected to be mammal)
  - Data set: $\{(x_i, y_i)\}_{i=1}^{n}$

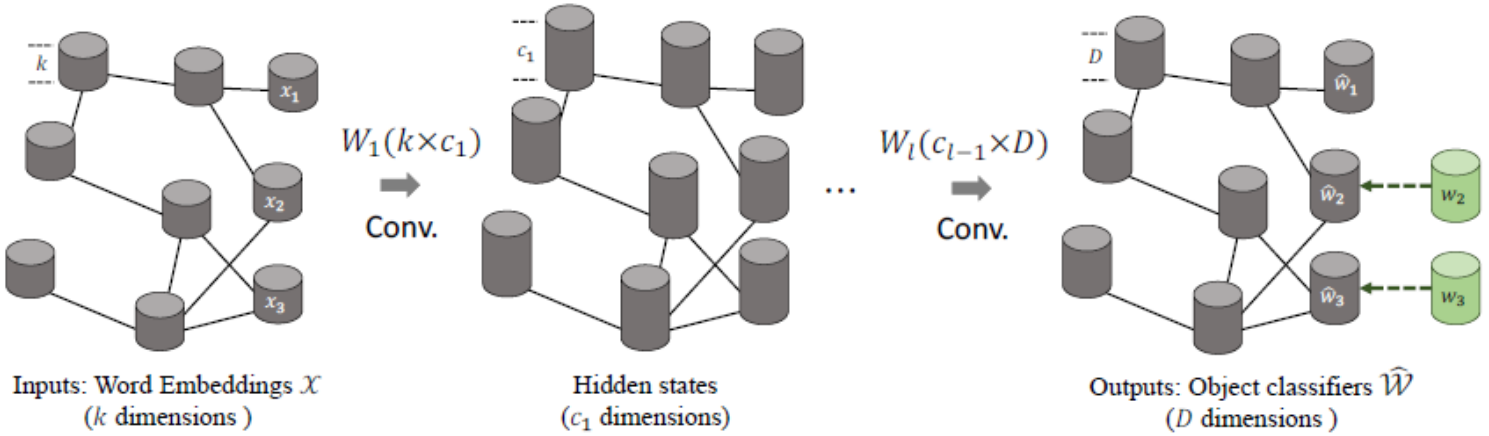
    n개의 entities (중 m개만 ground-truth를 안다고 가정)

    C개의 labels , $y_i \in \{1, ..., C\}$
  - Objective : m개 entity 로 학습시켜서

    n-m개의 entities에 대해 label(1~C)을 예측

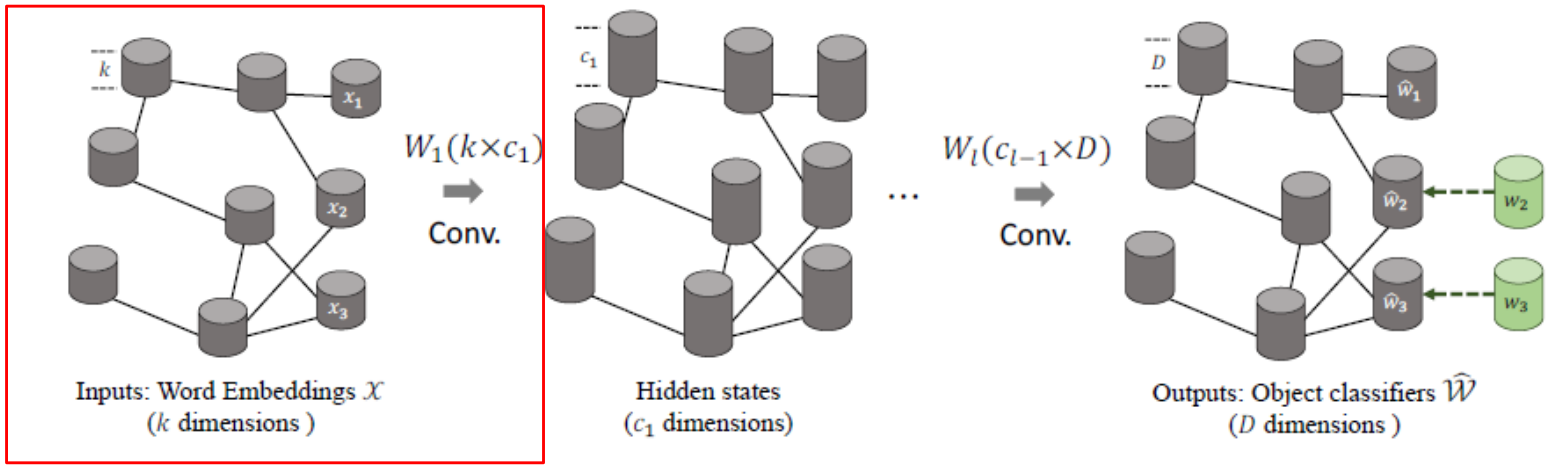
Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN



Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN

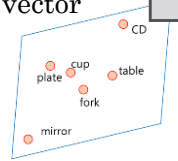


n x n $\quad$ n x k $\quad$ k x c_1

A · X · W

binary 인접행렬 $\quad$ word vector

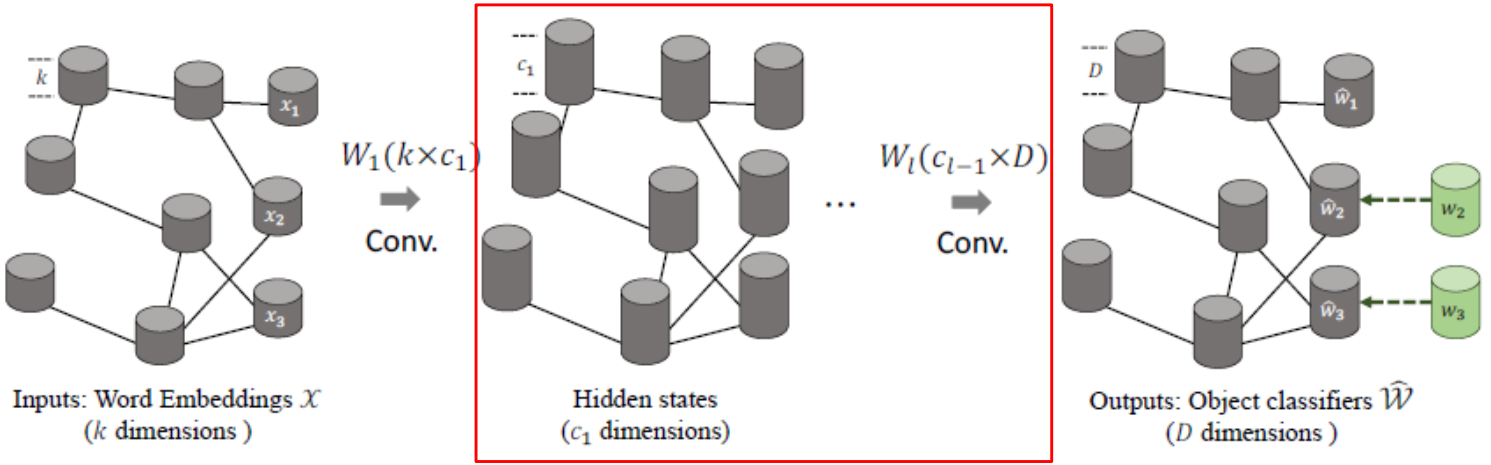
Knowledge graph



Semantic embeddings

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN

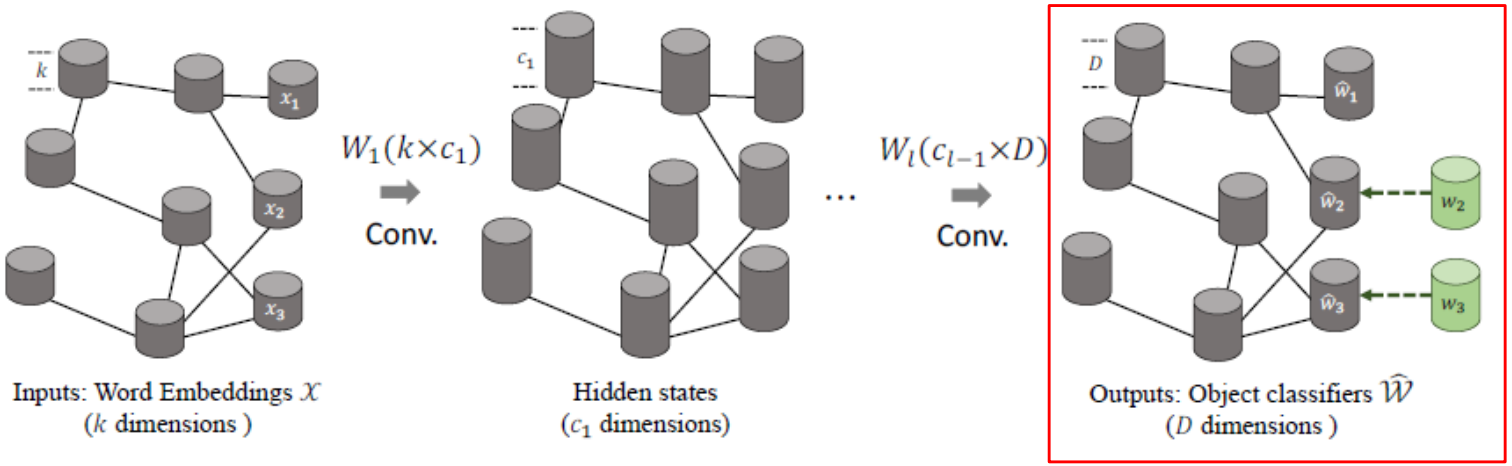


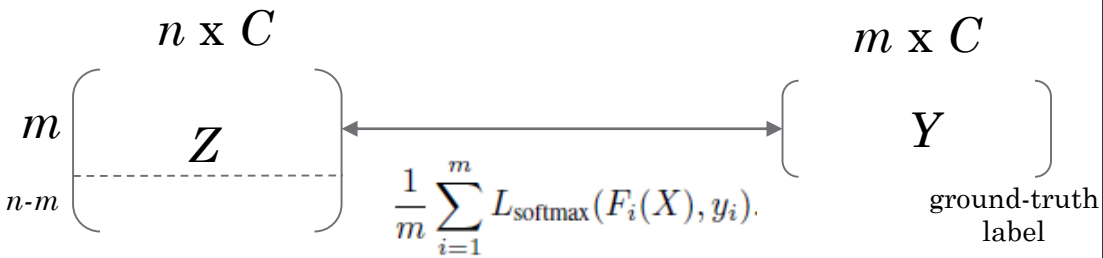
$$\cdots \left[\begin{matrix} n \times n \\ \mathrm{A} \end{matrix}\right] \cdot \left(\left[\begin{matrix} n \times n \\ \mathrm{A} \end{matrix}\right] \cdot \left[\begin{matrix} n \times k \\ \mathrm{X} \end{matrix}\right] \cdot \left[\begin{matrix} k \times c_1 \\ \mathrm{W}_1 \end{matrix}\right] \right) \left[\begin{matrix} c_1 \times c_2 \\ \mathrm{W}_2 \end{matrix}\right] \cdots$$

$$\mathrm{X}' \Rightarrow Z = \hat{A} X' W$$

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN



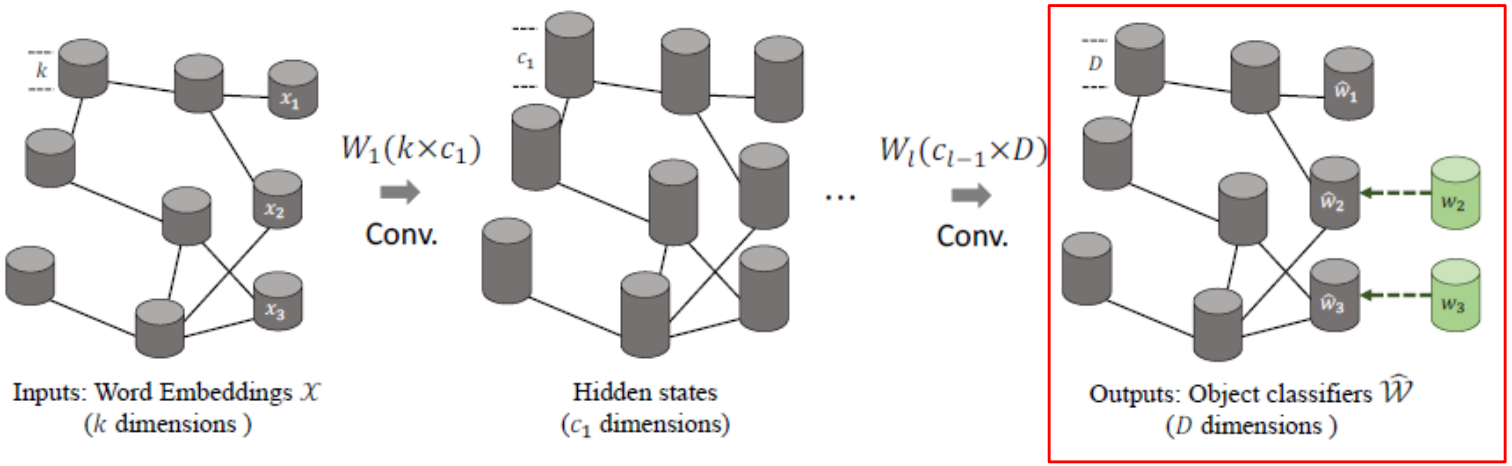


Back prop.

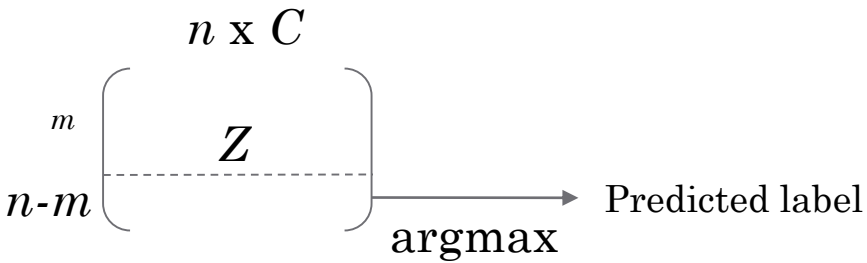
$$Z = \hat{A}X'W$$

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN



Inference :

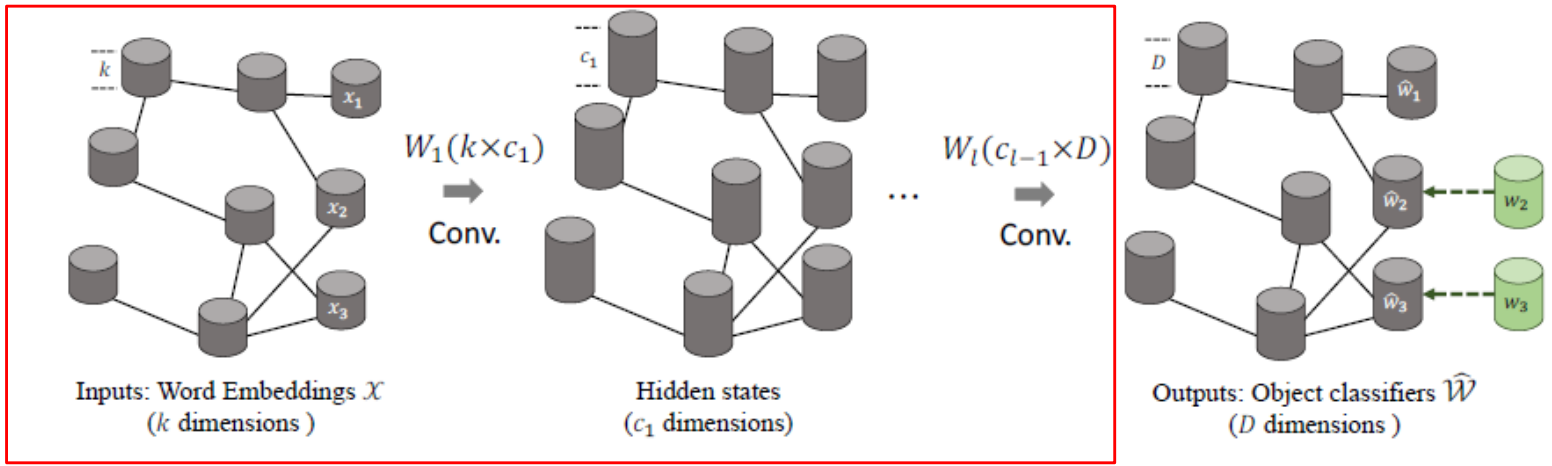


Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

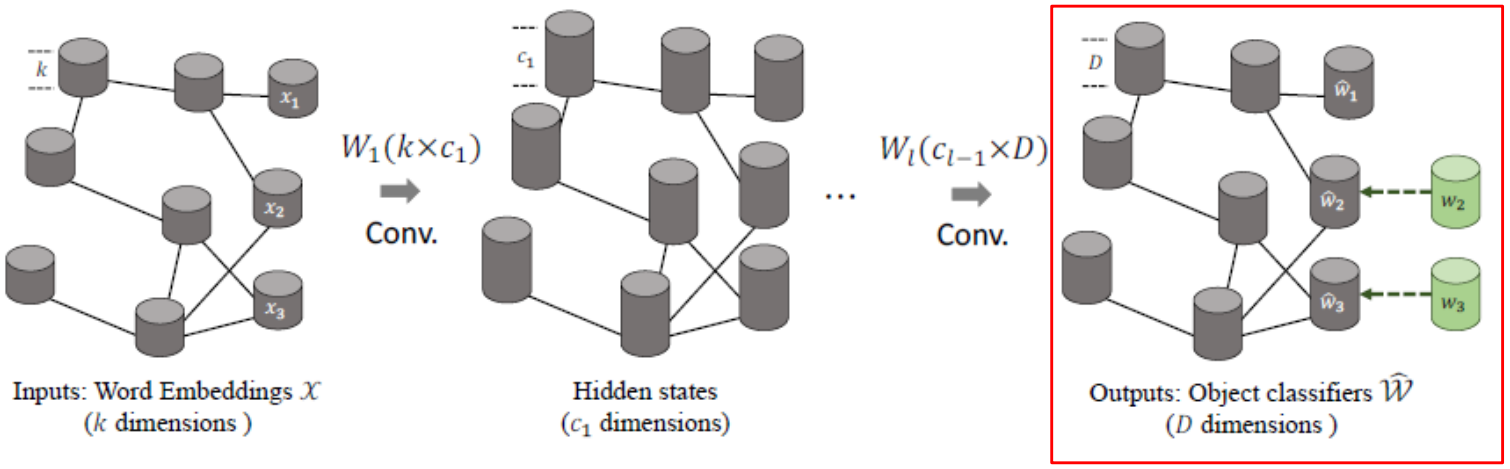
- GCN based Zero-shot Recognition



똑같아요

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

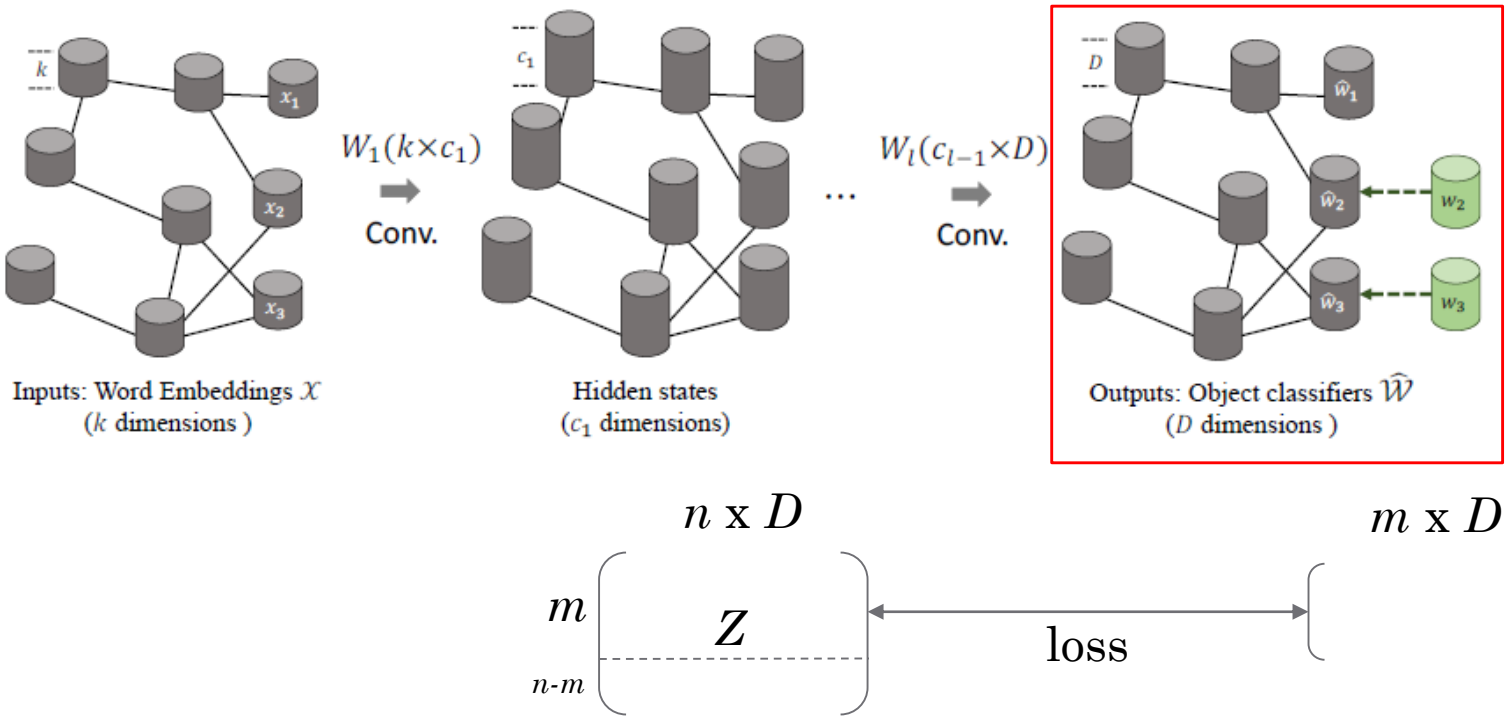
- GCN based Zero-shot Recognition



달라요

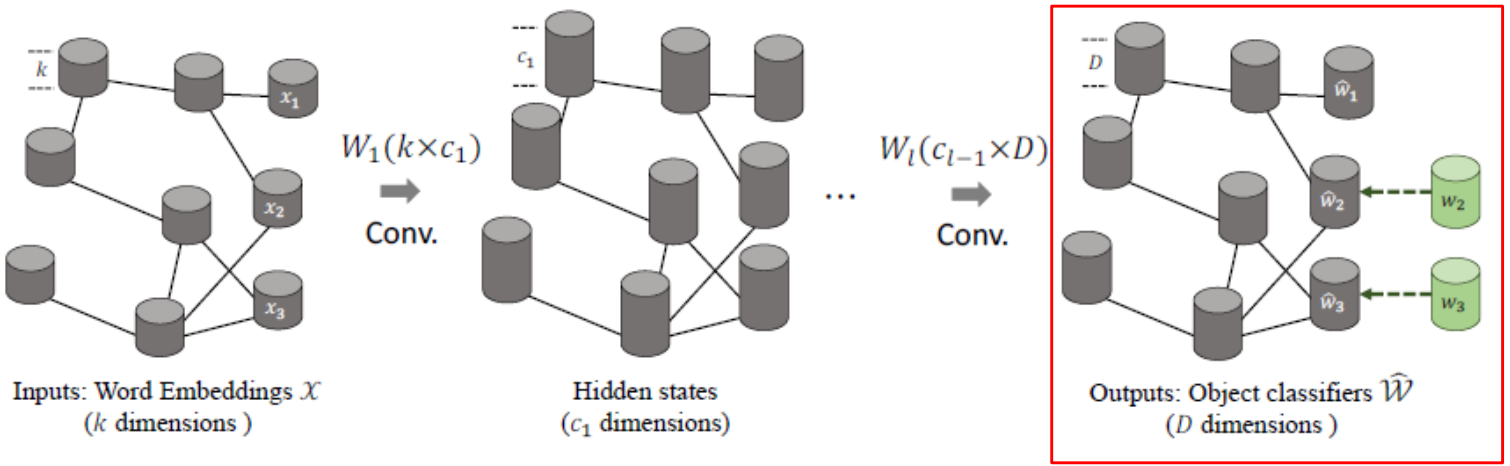
Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

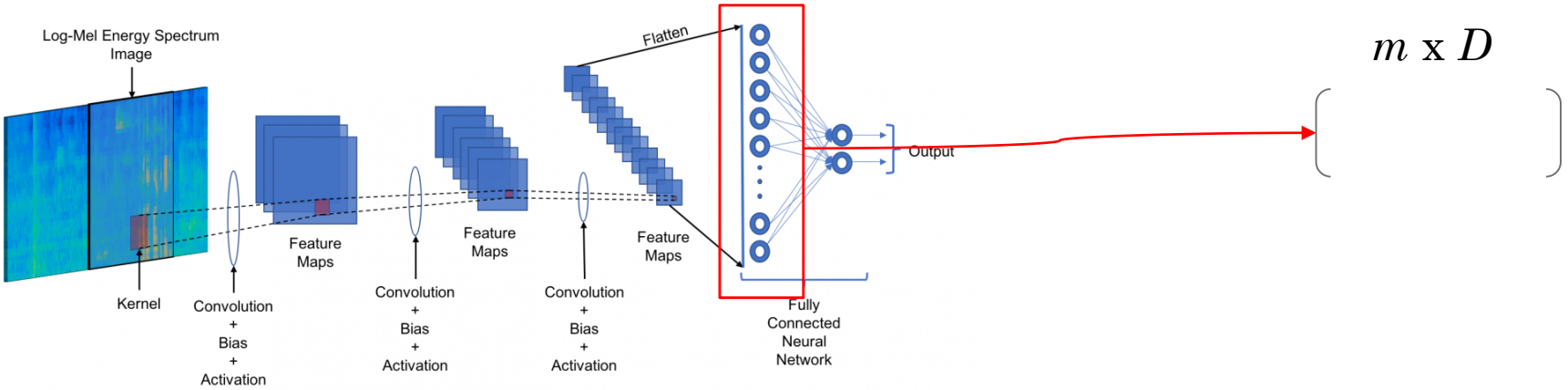
- GCN based Zero-shot Recognition



Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

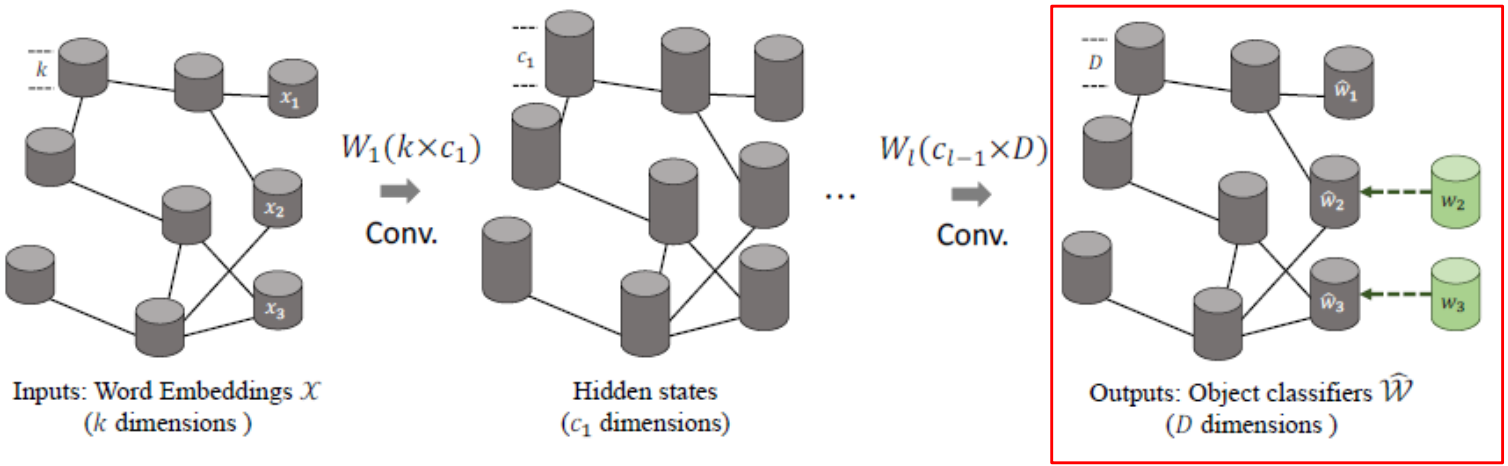
- GCN based Zero-shot Recognition

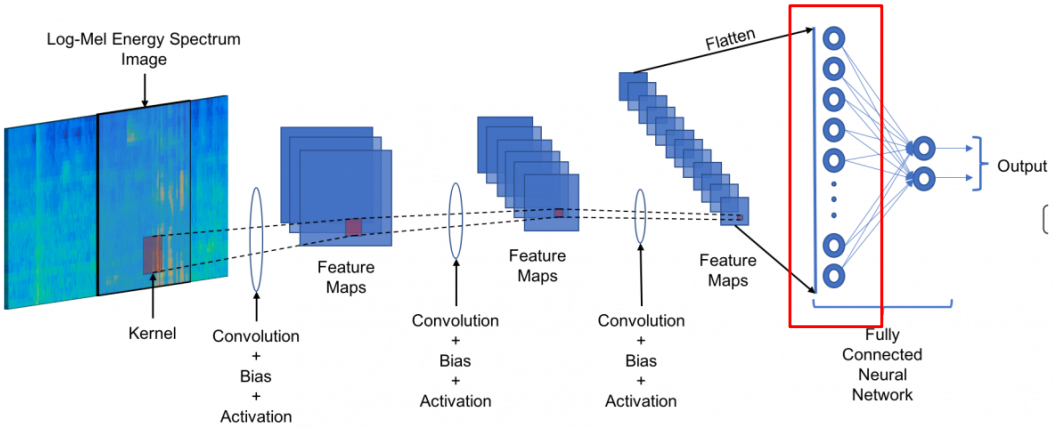




Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition





D x m

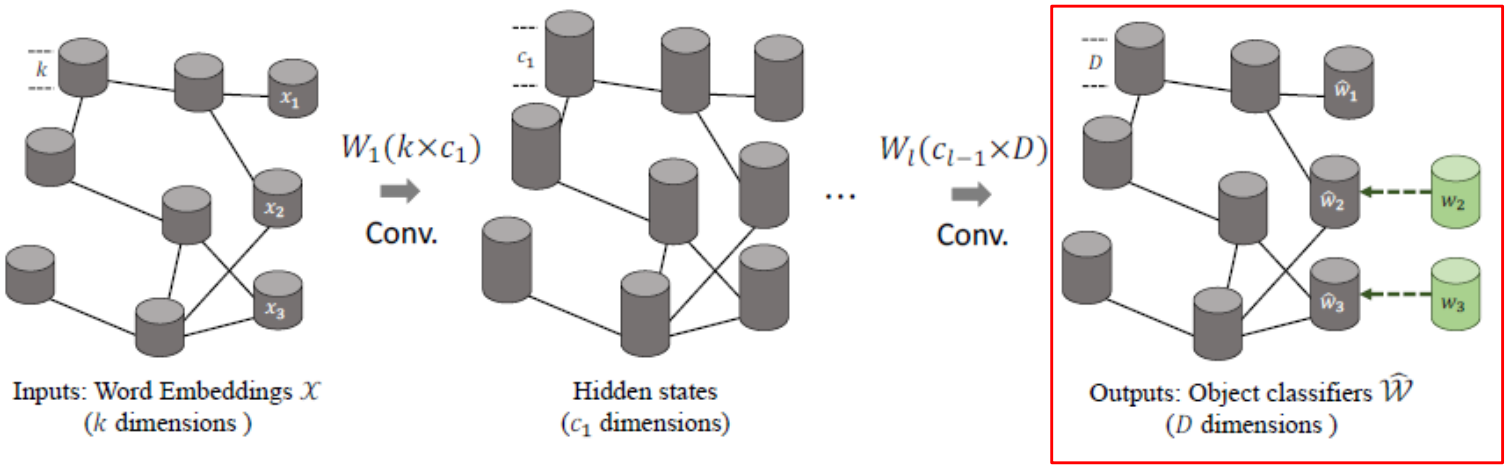
1 x D [image feature] • [classifier weight]

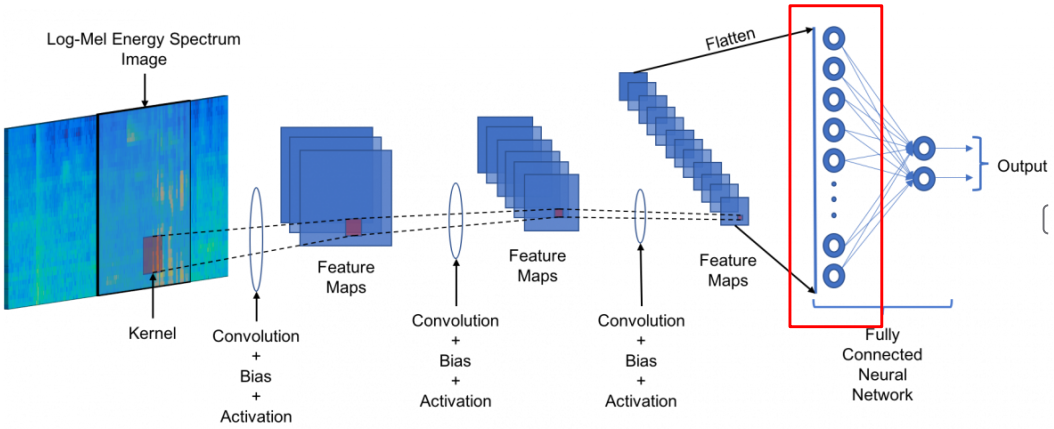
[score per label]

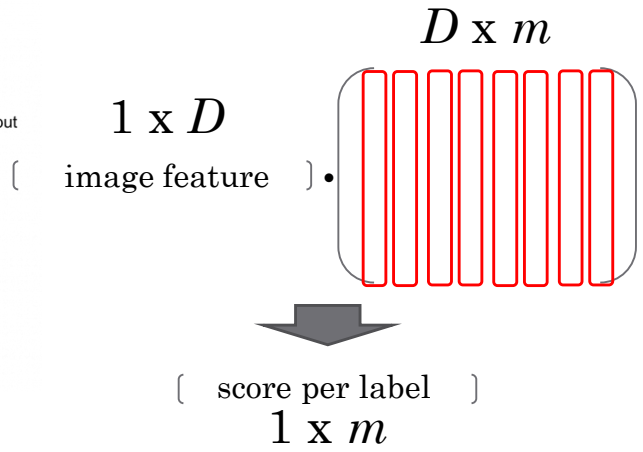
1 x m

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition

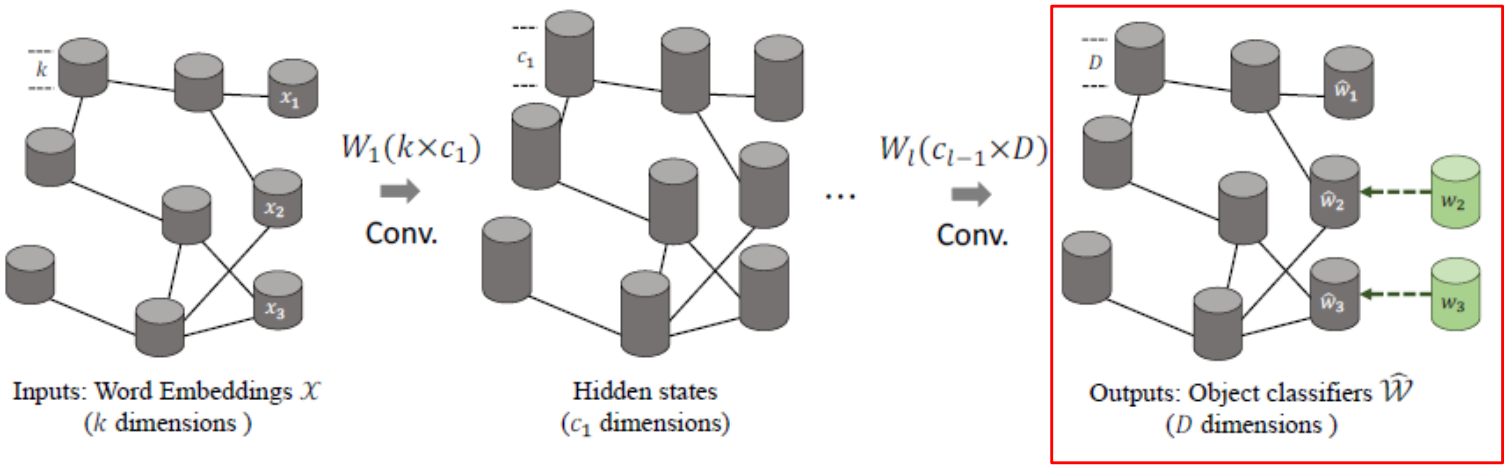


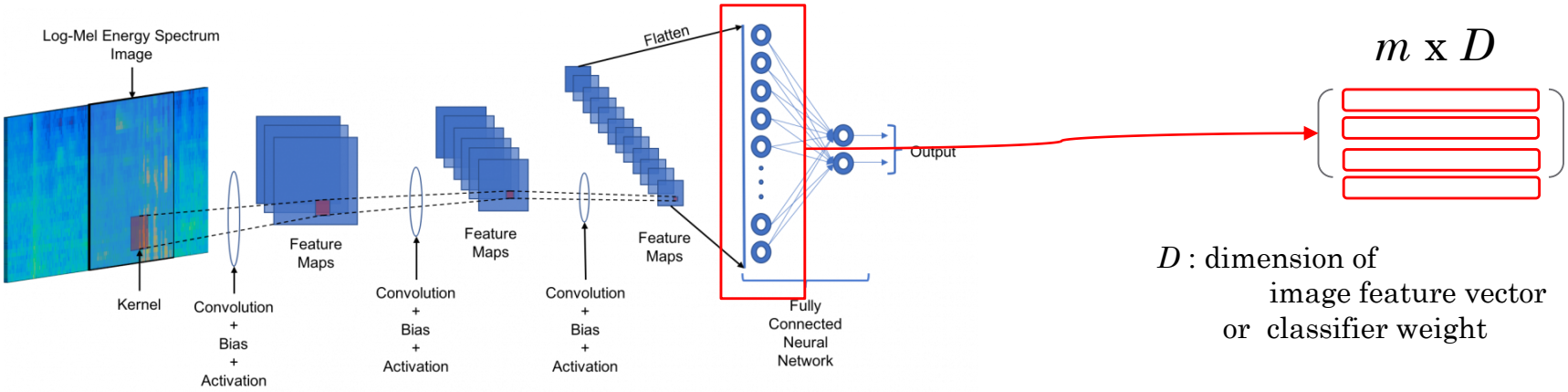




Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

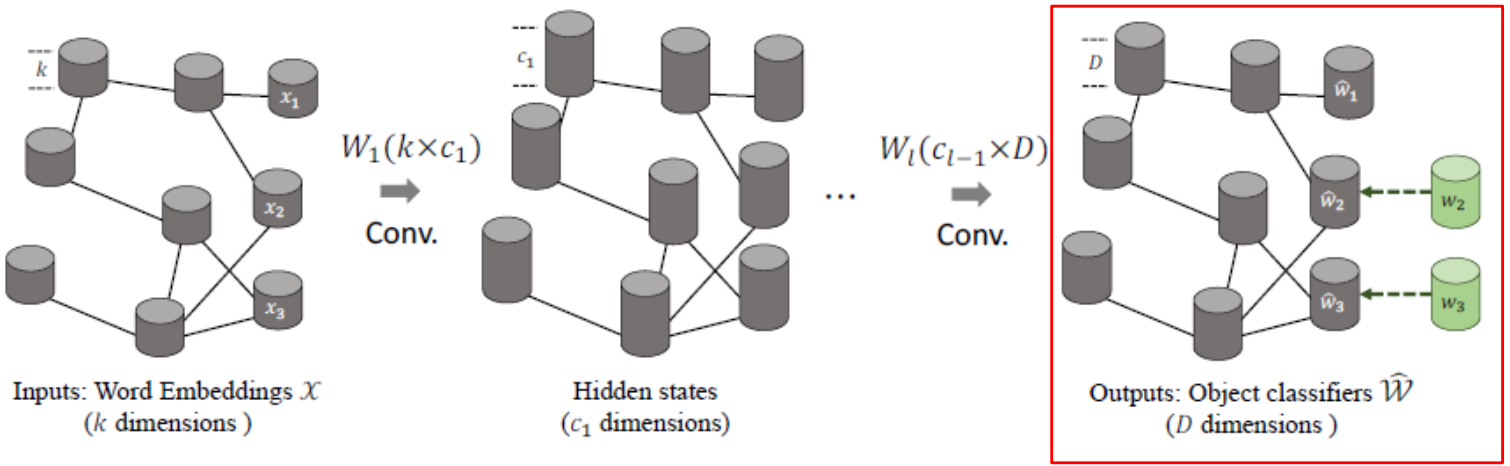
- GCN based Zero-shot Recognition

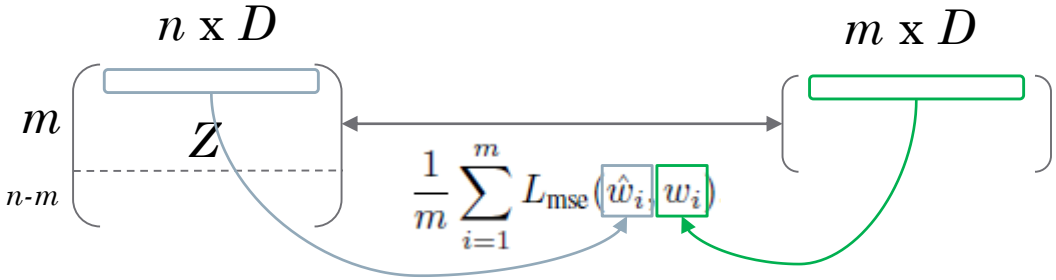




Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

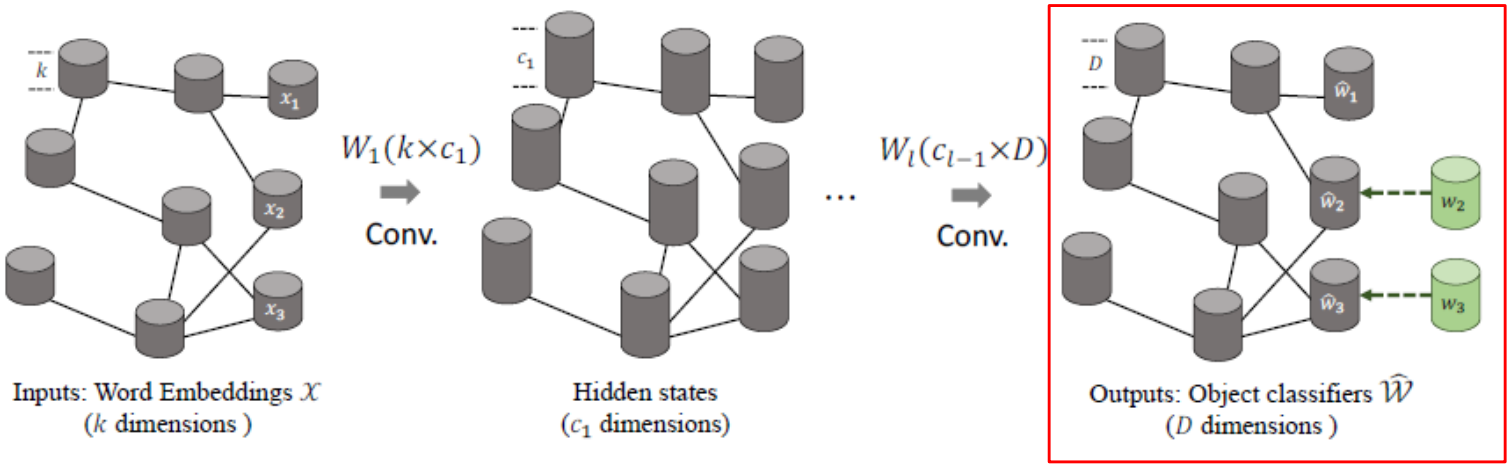
- GCN based Zero-shot Recognition

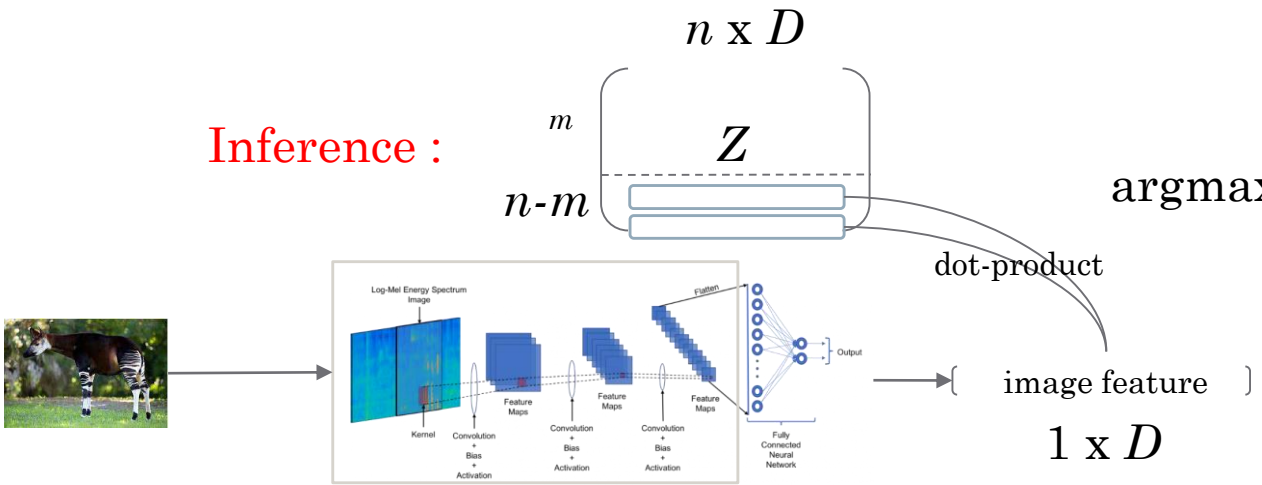




Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition





Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition
  - Trainset : ImageNet 1K , testset : ImageNet 21K (or 21K + 1K)
  - Semantic emb. : GloVe (GoogleNews) , knowledge graph : WordNet

Test Set	Model	ConvNets	Hit@k (%) 1	2	5	10	20
2-hops	ConSE [4]	Inception-v1	8.3	12.9	21.8	30.9	41.7
	ConSE(us)	Inception-v1	12.4	18.4	25.3	28.5	31.8
	SYNC [4]	Inception-v1	10.5	17.7	28.6	40.1	52.0
	EXEM [5]	Inception-v1	12.5	19.5	32.3	43.7	55.2
	Ours	Inception-v1	18.5	31.3	50.1	62.4	72.0
	Ours	ResNet-50	**19.8**	**33.3**	**53.2**	**65.4**	**74.6**
3-hops	ConSE [4]	Inception-v1	2.6	4.1	7.3	11.1	16.4
	ConSE(us)	Inception-v1	3.2	4.9	7.6	9.7	11.4
	SYNC [4]	Inception-v1	2.9	4.9	9.2	14.2	20.9
	EXEM [5]	Inception-v1	3.6	5.9	10.7	16.1	23.1
	Ours	Inception-v1	3.8	6.9	13.1	18.8	26.0
	Ours	ResNet-50	**4.1**	**7.5**	**14.2**	**20.2**	**27.7**
All	ConSE [4]	Inception-v1	1.3	2.1	3.8	5.8	8.7
	ConSE(us)	Inception-v1	1.5	2.2	3.6	4.6	5.7
	SYNC [4]	Inception-v1	1.4	2.4	4.5	7.1	10.9
	EXEM [5]	Inception-v1	**1.8**	2.9	5.3	8.2	12.2
	Ours	Inception-v1	1.7	3.0	5.8	8.4	11.8
	Ours	ResNet-50	**1.8**	**3.3**	**6.3**	**9.1**	**12.7**

(a) Top-k accuracy for different models when testing on only unseen classes.

Test Set	Model	ConvNets	Hit@k (%) 1	2	5	10	20
2-hops (+1K)	DeViSE [13]	AlexNet	0.8	2.7	7.9	14.2	22.7
	ConSE [34]	AlexNet	0.3	6.2	17.0	24.9	33.5
	ConSE(us)	Inception-v1	0.2	7.8	18.1	22.8	26.4
	ConSE(us)	ResNet-50	0.1	11.2	24.3	29.1	32.7
	Ours	Inception-v1	7.9	18.6	39.4	53.8	65.3
	Ours	ResNet-50	**9.7**	**20.4**	**42.6**	**57.0**	**68.2**
3-hops (+1K)	DeViSE [13]	AlexNet	0.5	1.4	3.4	5.9	9.7
	ConSE [34]	AlexNet	0.2	2.2	5.9	9.7	14.3
	ConSE(us)	Inception-v1	0.2	2.8	6.5	8.9	10.9
	ConSE(us)	ResNet-50	0.2	3.2	7.3	10.0	12.2
	Ours	Inception-v1	1.9	4.6	10.9	16.7	24.0
	Ours	ResNet-50	**2.2**	**5.1**	**11.9**	**18.0**	**25.6**
All (+1K)	DeViSE [13]	AlexNet	0.3	0.8	1.9	3.2	5.3
	ConSE [34]	AlexNet	0.2	1.2	3.0	5.0	7.5
	ConSE(us)	Inception-v1	0.1	1.3	3.1	4.3	5.5
	ConSE(us)	ResNet-50	0.1	1.5	3.5	4.9	6.2
	Ours	Inception-v1	0.9	2.0	4.8	7.5	10.8
	Ours	ResNet-50	**1.0**	**2.3**	**5.3**	**8.1**	**11.7**

(b) Top-k accuracy for different models when testing on both seen and unseen classes (a more practical and generalized setting).

Table 5. Results on ImageNet. We test our model on 2 different settings over 3 different datasets.

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition

Model	Word Embedding	Hit@k (%)				
		1	2	5	10	20
[53]	GloVe	7.8	11.5	17.2	21.2	25.6
Ours	GloVe	**18.5**	**31.3**	**50.1**	**62.4**	**72.0**
[53]	FastText	9.8	16.4	27.8	37.6	48.4
Ours	FastText	**18.7**	**30.8**	**49.6**	**62.0**	**71.5**
[53]	GoogleNews	13.0	20.6	33.5	44.1	55.2
Ours	GoogleNews	**18.3**	**31.6**	**51.1**	**63.4**	**73.0**

Table 6. Results with different word embeddings on ImageNet (2 hops), corresponding to the experiments in Table 5a.

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

- GCN based Zero-shot Recognition

Test Image	ConSE (10)	Ours
	panthera tigris(train) tiger cat (train) felis onca (train) leopard (train) tiger shark (train)	tigress (test) **bengal tiger (test)** panthera tigris (train) tiger cub (test) tiger cat (train)
	rock beauty (train) ringlet (train) flagpole (train) large slipper (test) yellow slipper (train)	**butterfly fish (test)** rock beauty (train) damselfish (test) atoll (test) barrier reef (test)
	tractor (train) reaper (train) thresher (train) trailer truck (train) motortruck (test)	tracked vehicle (test) tractor (train) propelled vehicle (test) reaper (train) forklift (train)

Figure 6. Visualization of top 5 prediction results for 3 different images. The correct prediction results are highlighted by red bold characters. The unseen classes are marked with a red "test" in the bracket. Previously seen classes have a plain "train" in the bracket.