

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

Xiaolong Wang, Yufei Ye, Abhinav Gupta

CVPR 2018

조건희

Index

- Introduction
- Related work
- Approach
- Experiment
- Conclusion

Introduction

Introduction

Zero-shot recognition 태스크를
GCN 기법을 활용하여 해결

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier

for a category with zero training examples,

just using the word embedding of the category and its relationship to other categories,

which visual data are provided

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier
이미지 분류 모델을 학습할 거임

for a category with zero training examples,

just using the word embedding of the category and its relationship to other categories,

which visual data are provided

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier
이미지 분류 모델을 학습할 거임

for a category with zero training examples,
훈련 데이터에 이미지 샘플이 없는 카테고리를 분류해주는 모델임

just using the word embedding of the category and its relationship to other categories,

which visual data are provided

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier
이미지 분류 모델을 학습할 거임

for a category with zero training examples,
훈련 데이터에 이미지 샘플이 없는 카테고리를 분류해주는 모델임

just using the word embedding of the category and its relationship to other categories,
그 카테고리 and 다른 카테고리 사이의 관계와 워드임베딩만 이용해서 해볼거임

which visual data are provided

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier
이미지 분류 모델을 학습할 거임

for a category with zero training examples,
훈련 데이터에 이미지 샘플이 없는 카테고리를 분류해주는 모델임

just using the word embedding of the category and its relationship to other categories,
그 카테고리 and 다른 카테고리의 관계와 워드임베딩만 이용해서 해볼거임

which visual data are provided
다른 카테고리들은 이미지 데이터가 주어짐

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

learning a visual classifier
이미지 분류 모델을 학습할 거임

for a category with zero training examples,
훈련 데이터에 이미지 샘플이 없는 카테고리를 분류해주는 모델임

just using the word embedding of the category and its relationship to other categories,
그 카테고리 and 다른 카테고리의 관계와 워드임베딩만 이용해서 해볼거임

which visual data are provided
다른 카테고리들은 이미지 데이터가 주어짐

unseen

seen

Introduction

Zero-shot recognition

Zero-shot recognition 태스크가 뭐지?

familiar category(seen)로부터 얻은 정보를

unfamiliar category(unseen)가 무엇인지 알아내는 데에 활용

→ 일종의 Transfer learning

Introduction

Zero-shot recognition



→ 개



→ 고양이

train

test



→ 개

일반적인 이미지 분류

Introduction

Zero-shot recognition



→ 개



→ 고양이

train

test



→ ?

Zero-shot recognition

Introduction

Zero-shot recognition

“오카피”

동물

4개의 다리는 얼룩말 무늬

몸통은 갈색

얼굴은 사슴을 닮음

Introduction

Zero-shot recognition



Introduction

Zero-shot recognition

“오카피”

동물

4개의 다리는 얼룩말 무늬

몸통은 갈색

얼굴은 사슴을 닮음

Introduction

Zero-shot recognition

We have labeled data, why bother?

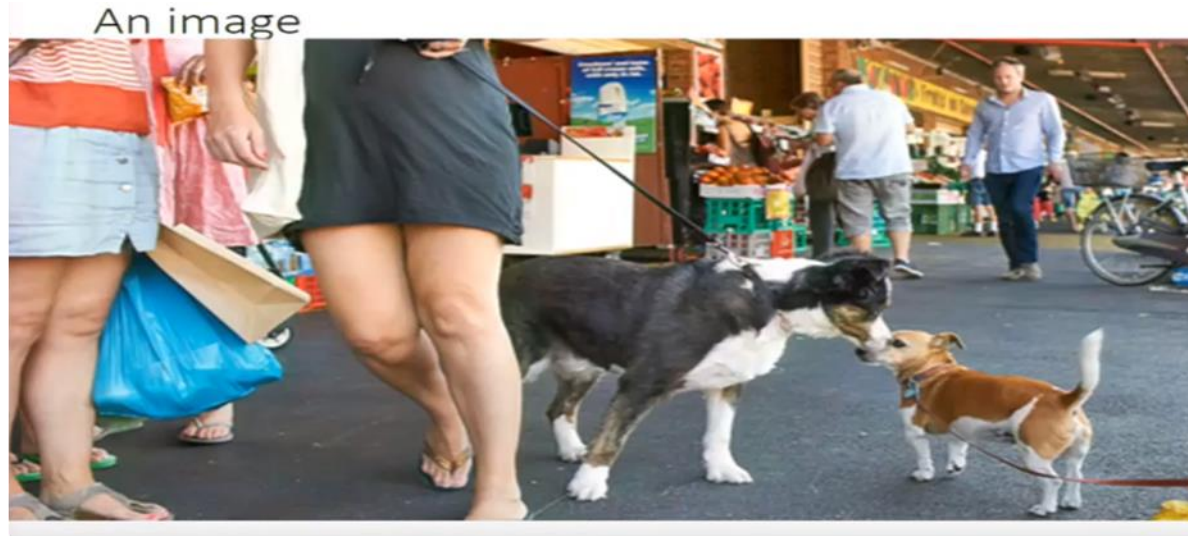


Imagenet: ~15,000,000 images
Open Images: ~9,000,000 images
Places: ~2,500,000 images

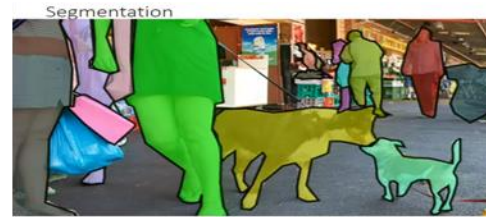
Zero-shot recognition 이 왜 필요할까?

Introduction

Zero-shot recognition



Classification



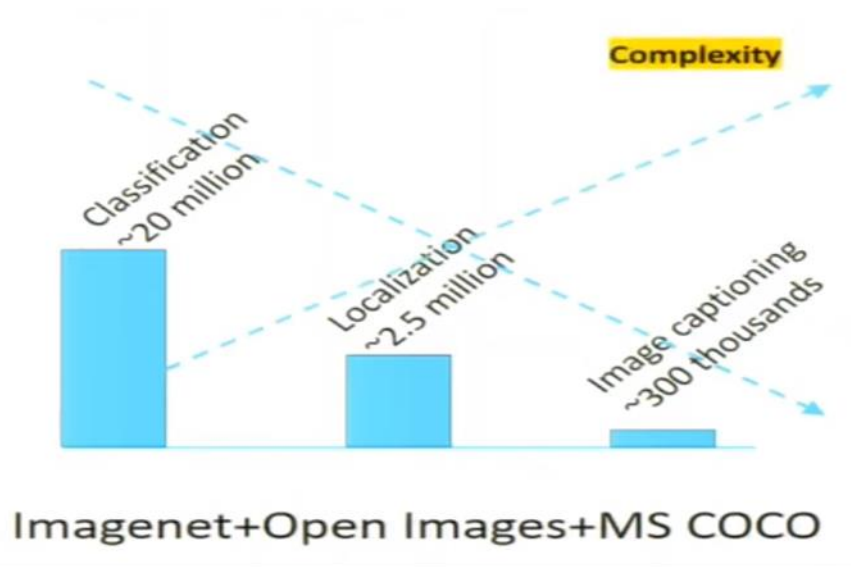
Segmentation



Captioning

Introduction

Zero-shot recognition



더 복잡한 annotation 일수록 데이터셋을 구축하는 것이 더 어려움.

The more complex task we target,
the fewer annotations we have,
the more relevant zero-shot learning is.

Related work

Related work

unseen 과 seen 사이의 관계를 모델이 알 수 있도록 하는 것이 핵심

기존 연구의 2가지 대표적 방법

- (1) Explicit knowledge base 를 이용하는 방법
- (2) Implicit knowledge representation 을 이용하는 방법

Related work

(1) Explicit knowledge base

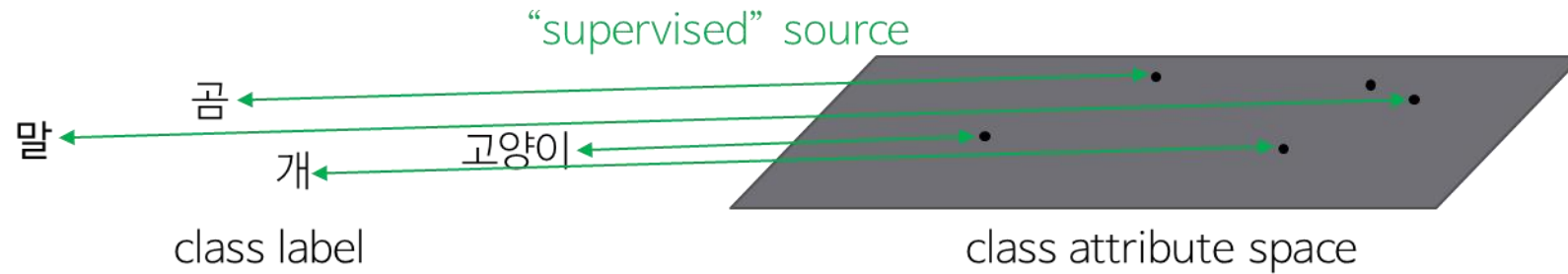
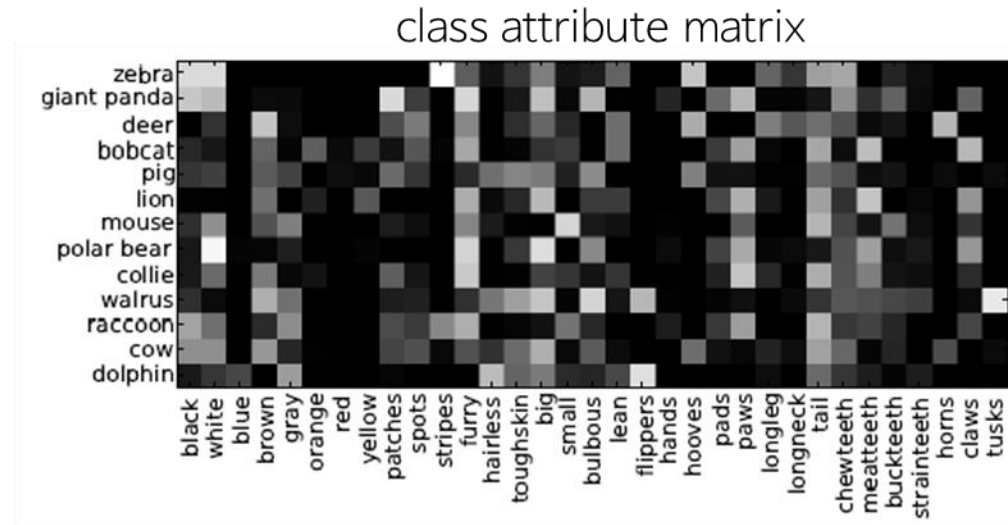
(1) Explicit knowledge base 를 이용하는 방법

seen, **unseen** 전부 포함된 지식 그래프

→ 두 카테고리 간의 관계가 명시적으로 표현됨.

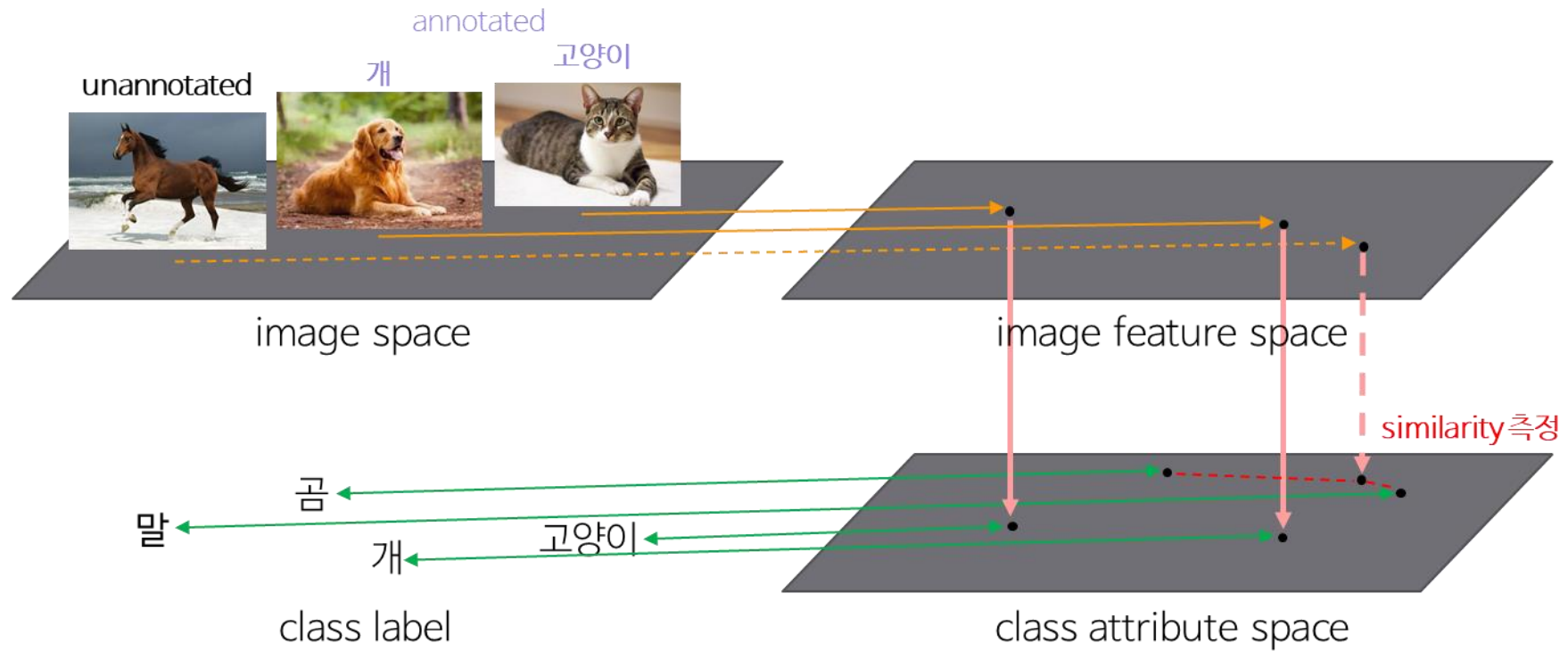
Related work

(1) Explicit knowledge base



Related work

(1) Explicit knowledge base



Related work

(2) Implicit knowledge representation

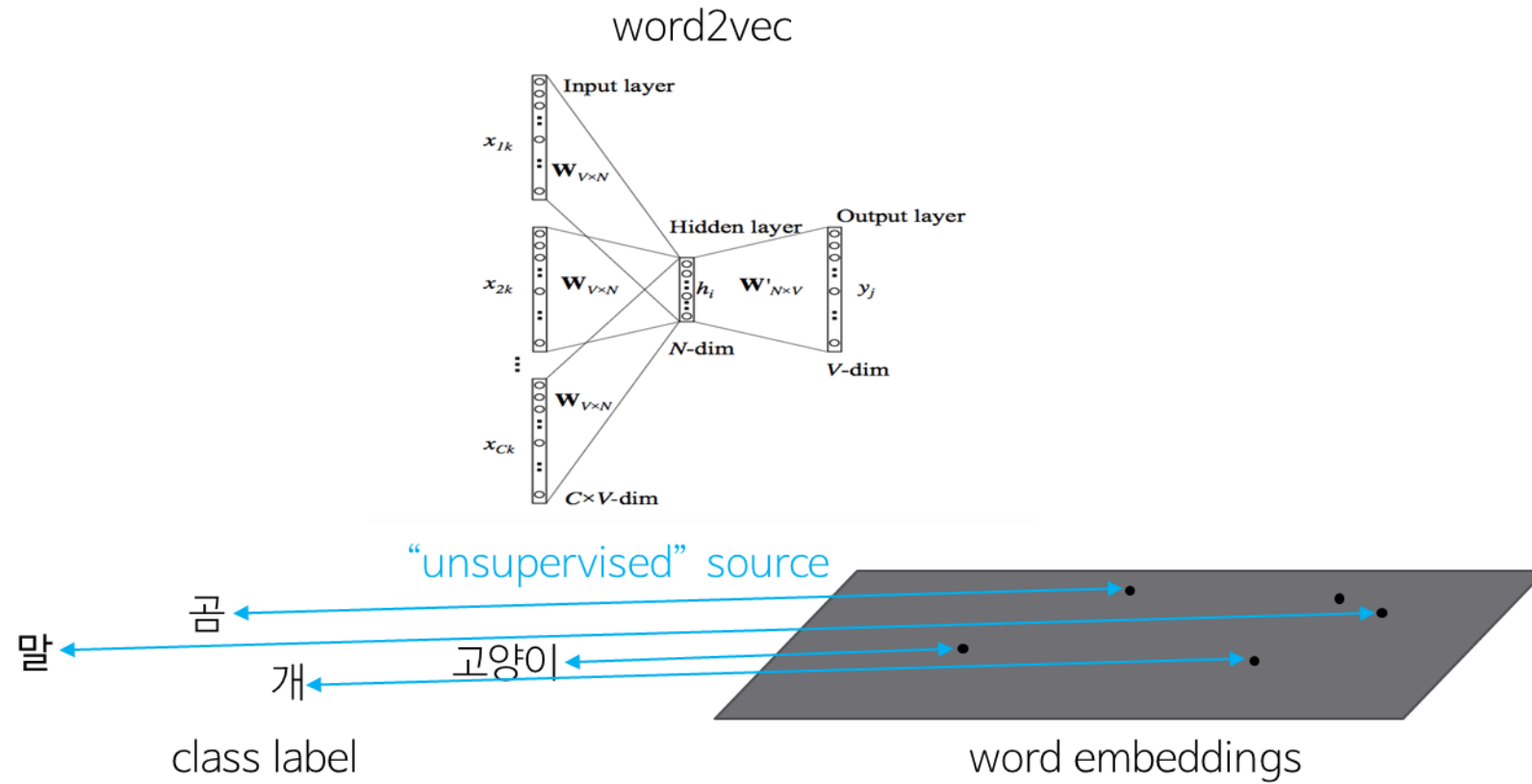
(2) Implicit knowledge representation 을 이용하는 방법

seen, unseen 전부 포함된 워드임베딩

→ 두 카테고리 간의 관련성이 높으면 높을 수록,
두 카테고리의 워드임베딩 벡터 간 유사도가 더 높음.

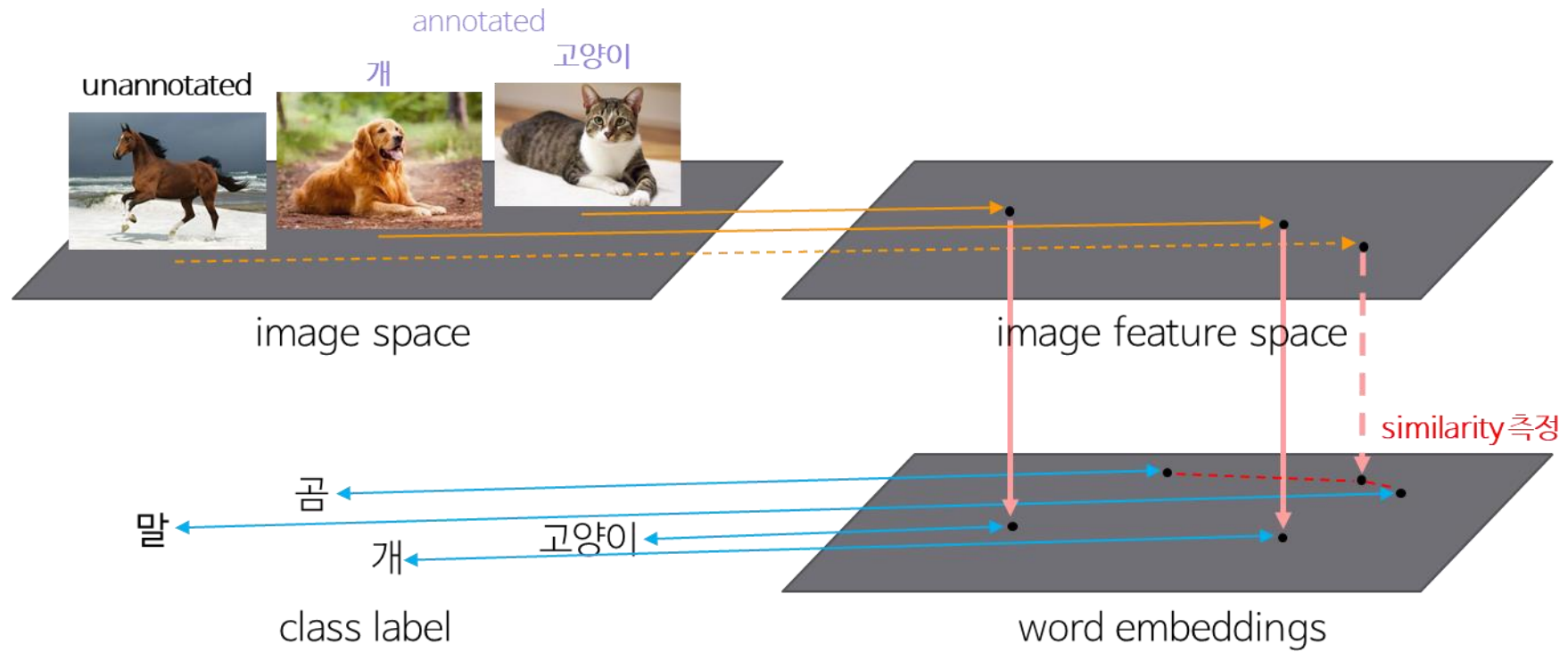
Related work

(2) Implicit knowledge representation



Related work

(2) Implicit knowledge representation



Approach

Approach

GCN

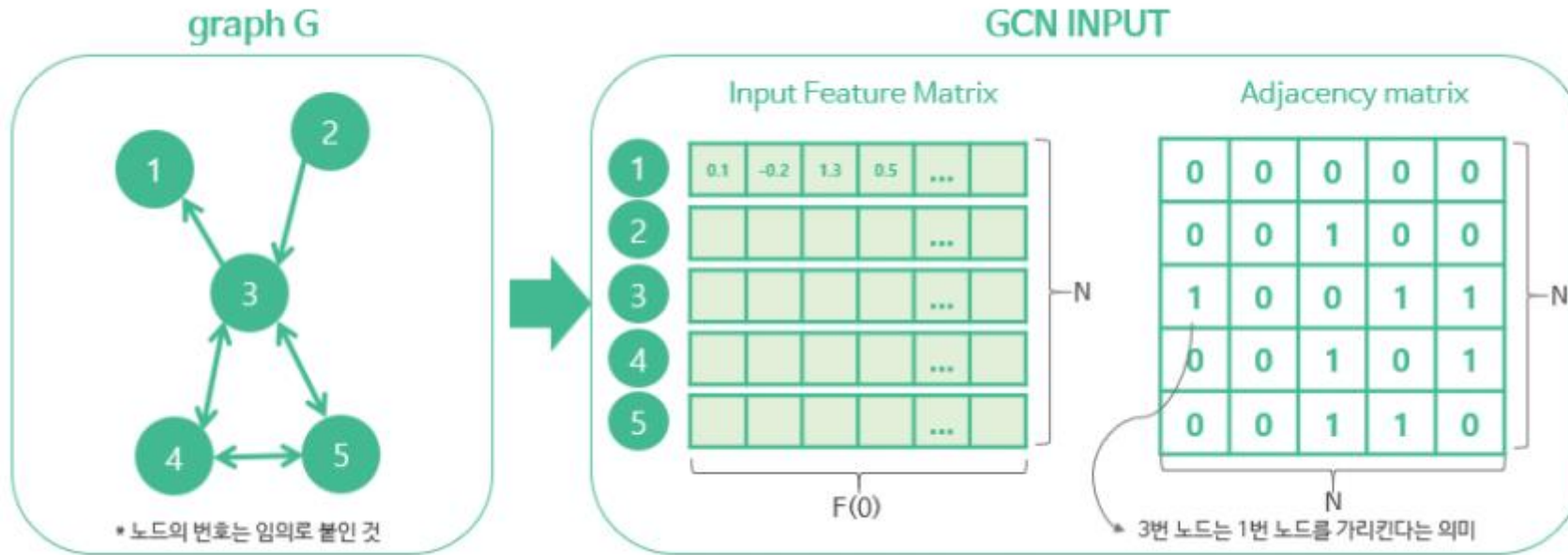
Graph Convolutional Network

Approach

GCN

$$H^i = f(H^{i-1}, A)$$

where $H^0 = X, f = \text{propagation}$



노드 피쳐

H^i

인접행렬

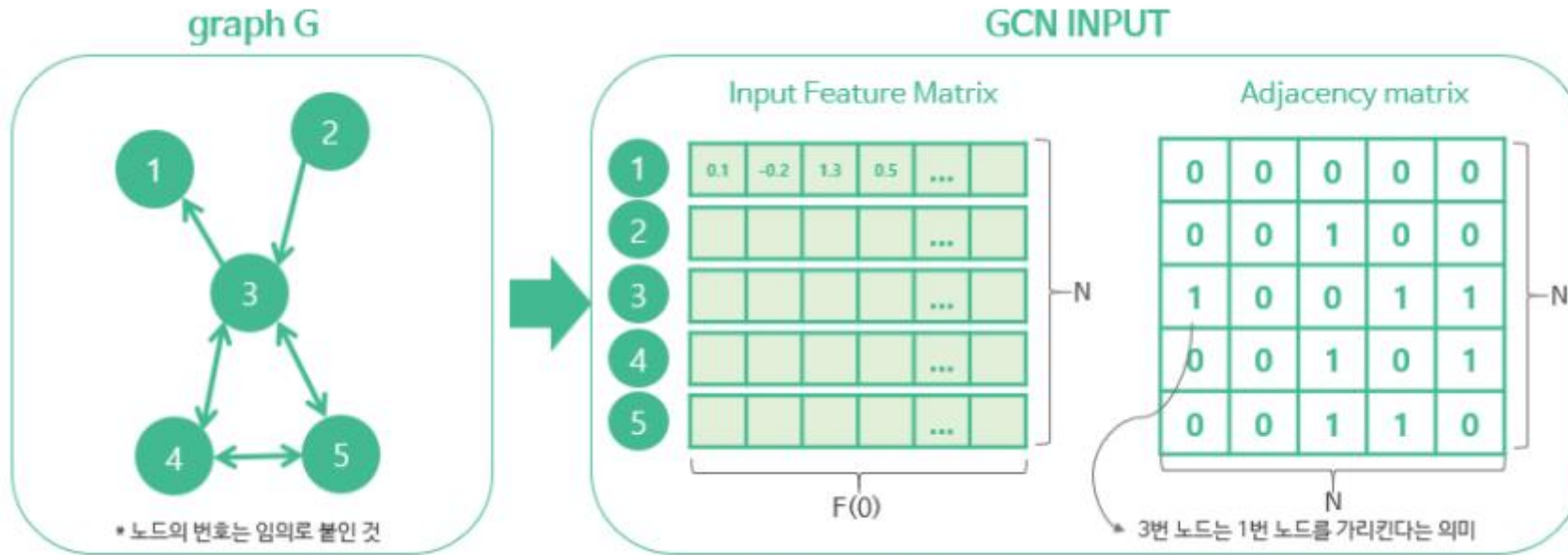
A

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non-linear activation function



$$AH^i$$

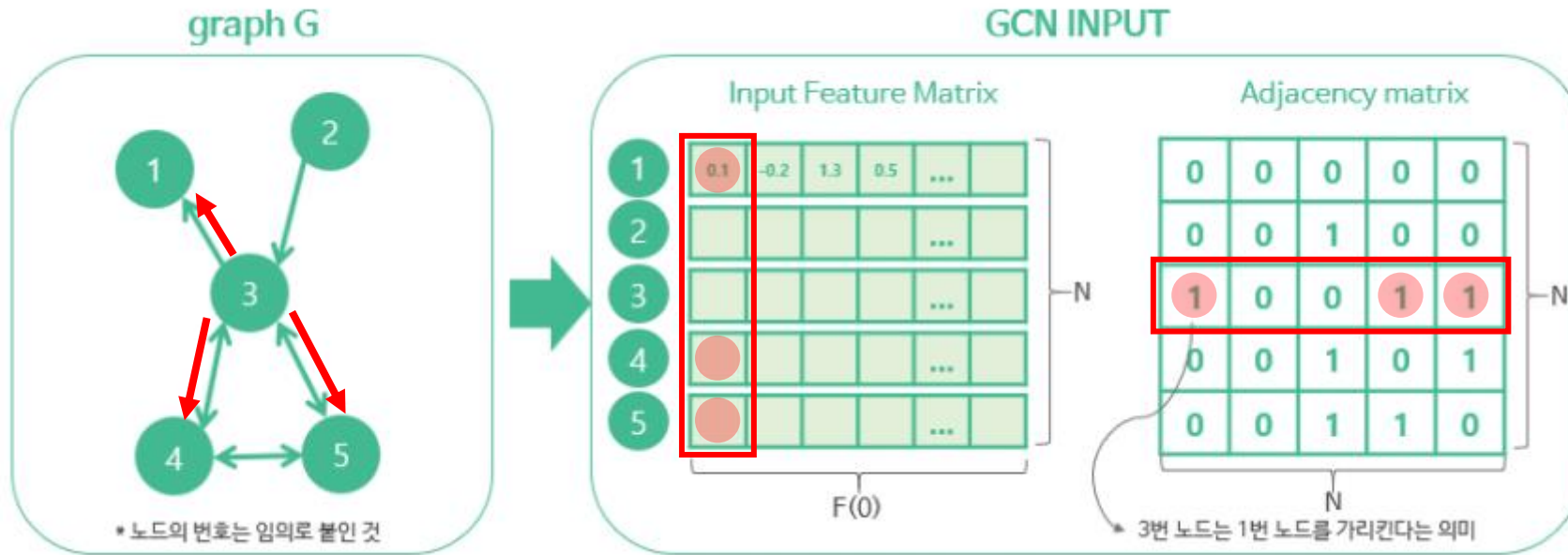
이 행렬 곱의 의미

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non – linear activation function



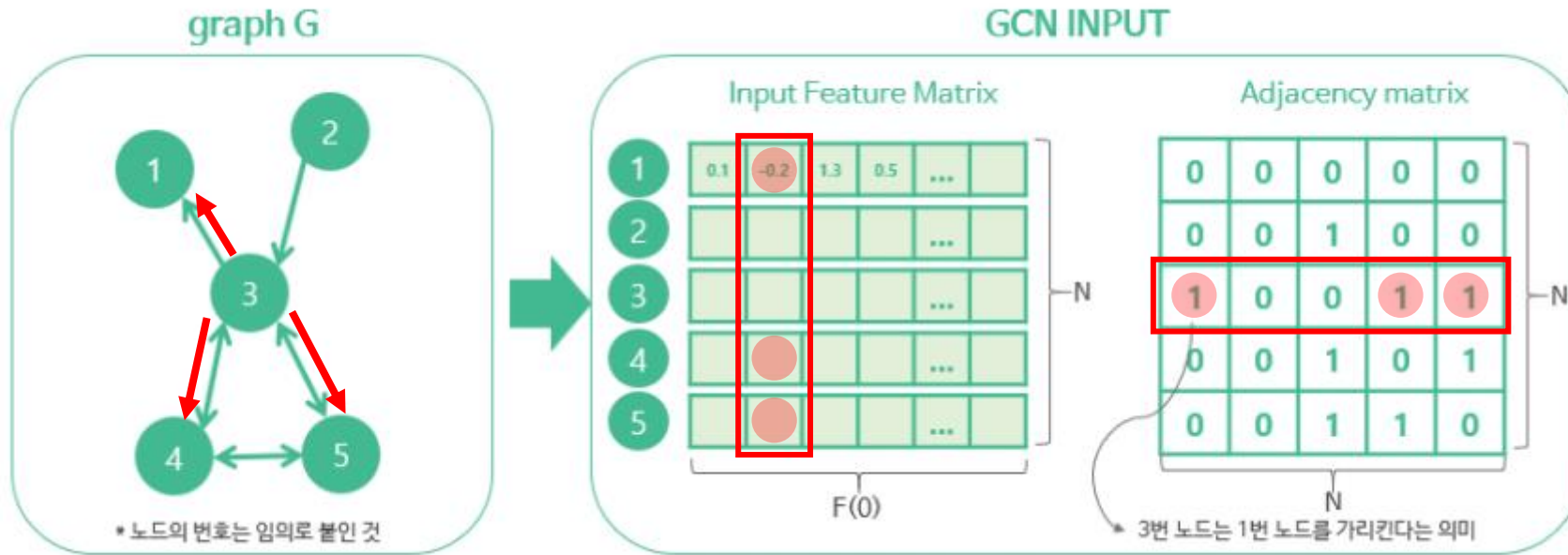
3번 노드를 기준으로 계산을 해보면,

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non – linear activation function



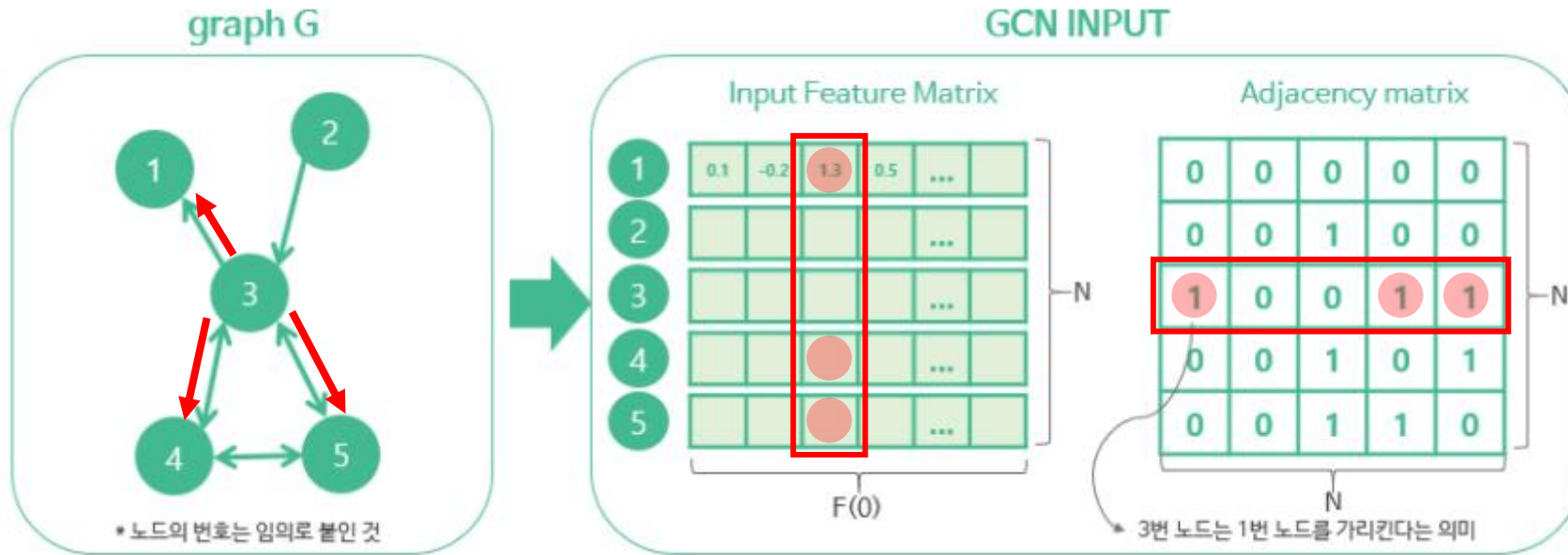
3번 노드를 기준으로 계산을 해보면,

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non-linear activation function



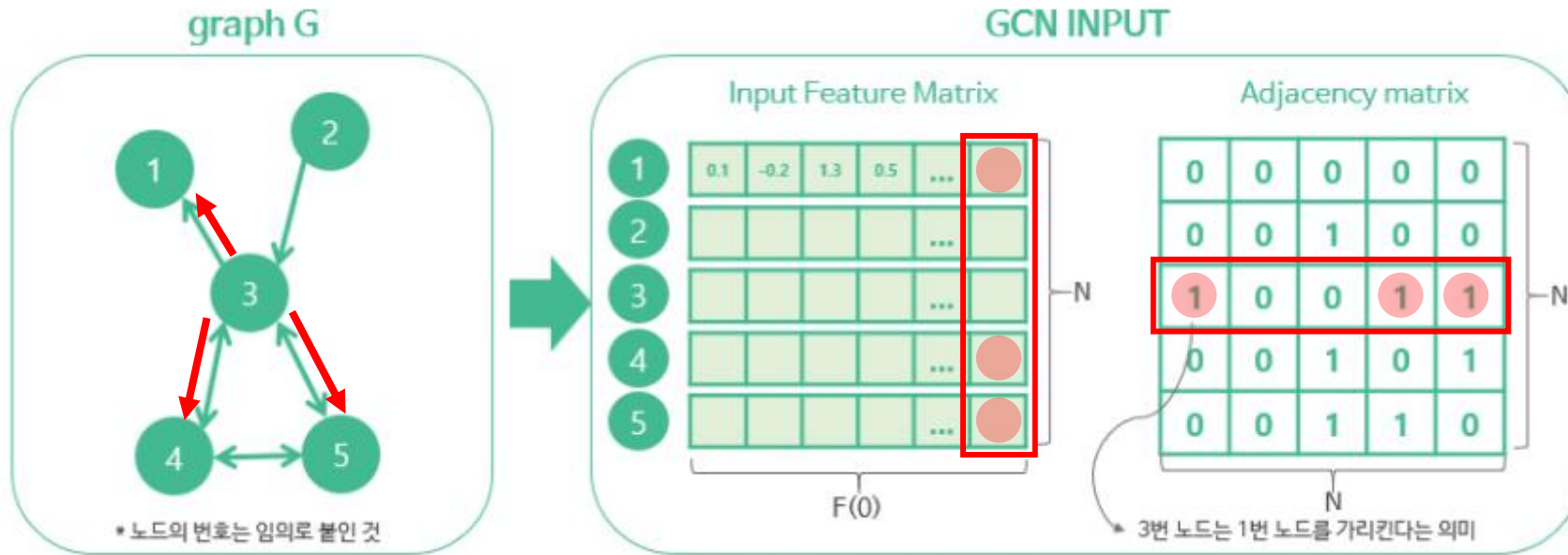
3번 노드를 기준으로 계산을 해보면,

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non-linear activation function



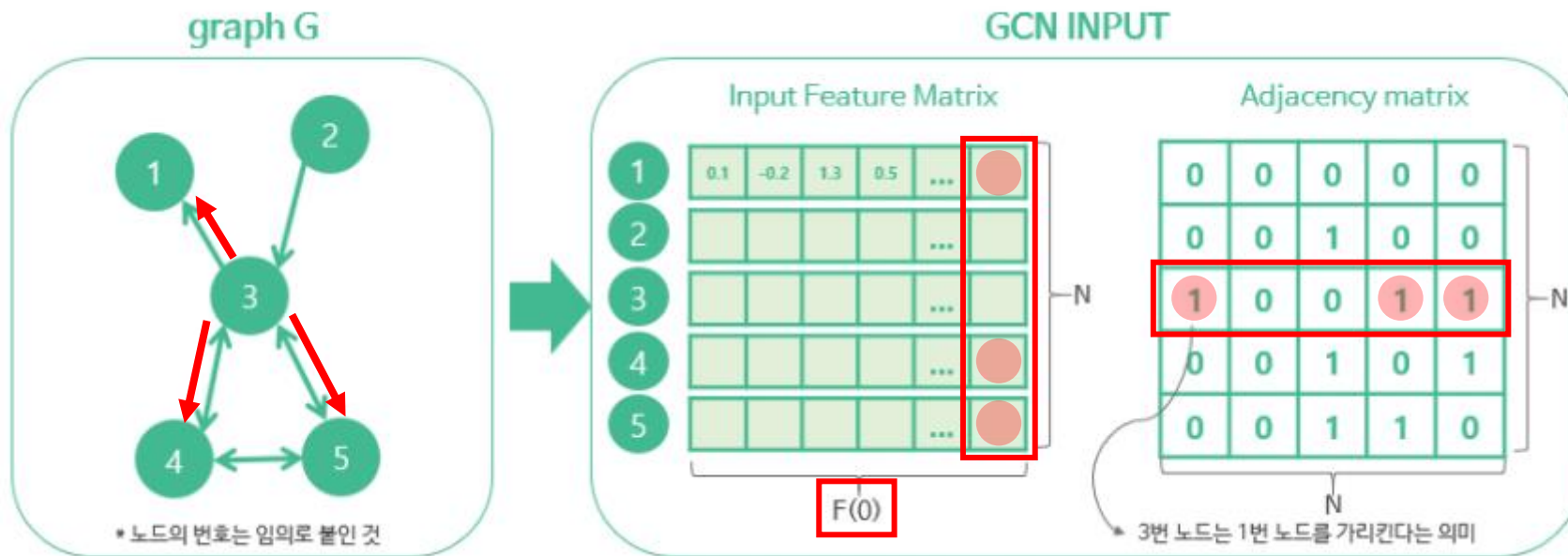
3번 노드를 기준으로 계산을 해보면,

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non-linear activation function



계산 결과 : F(0) 차원의 벡터

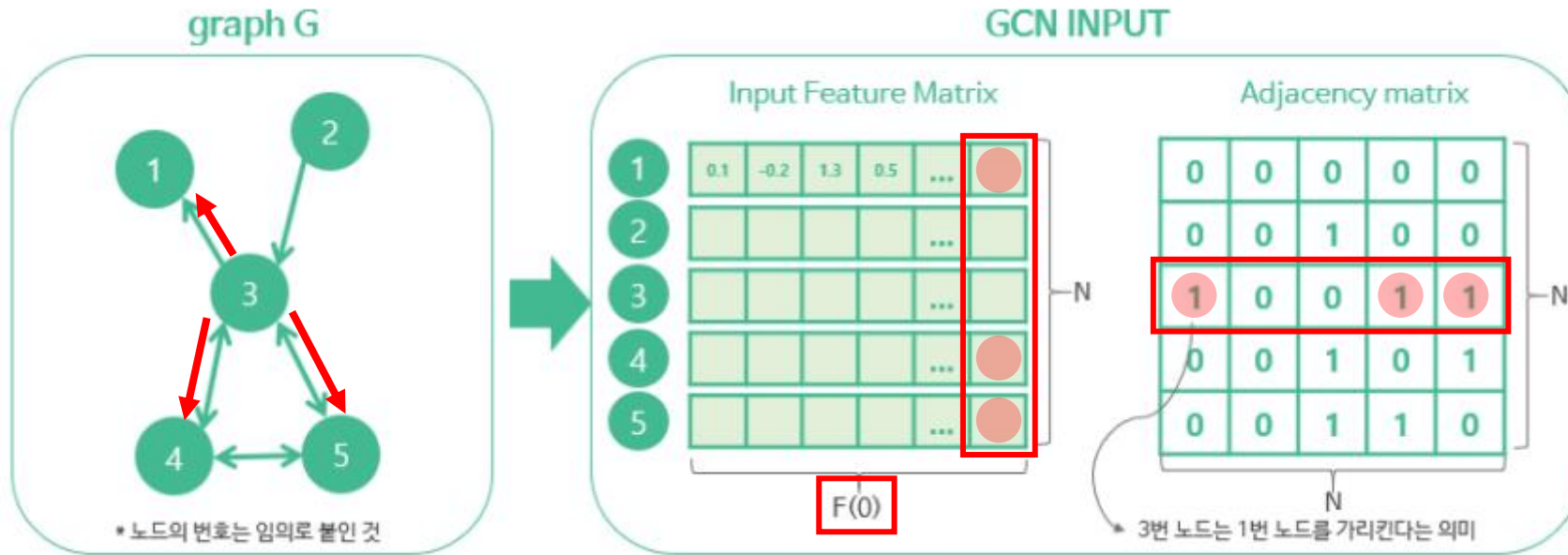
그 벡터의 의미 : 3번 노드와 연결된 노드인 1번, 4번, 5번 노드 피쳐의 합

Approach

GCN

$$f(H^i, A) = \sigma(AH^iW^i)$$

where σ is a non – linear activation function



$$AH^i$$

모든 노드에 대하여 계산하면,

F(0) 차원 벡터가 N개 있는 형태 = $N \times F(0)$ 매트릭스

Approach

GCN

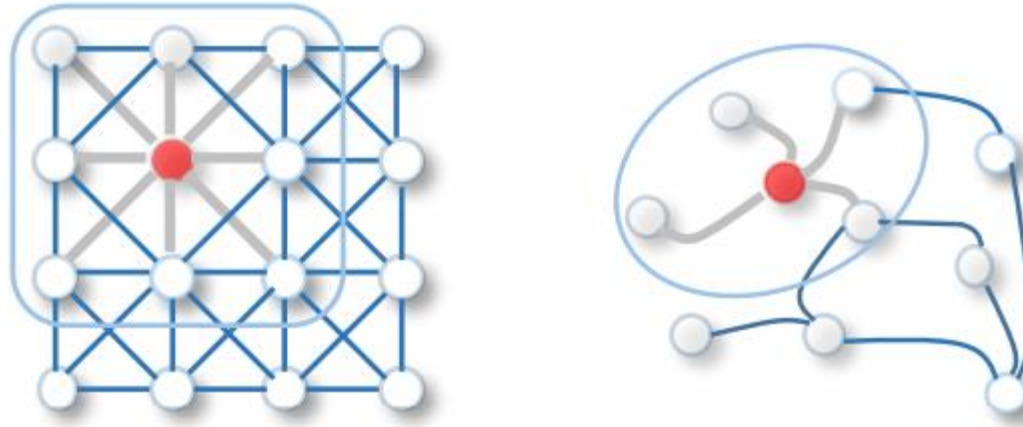
문제점

1. 자기 자신의 feature는 합해지지 않음
 - 위 예시에서 1번 노드의 경우
 - 인접 행렬의 대각 성분은 전부 1로 채우자!
2. 연결이 아주 많은 노드는 feature 의 값이 너무 커진다.
 - 위 예시에서 3번 노드의 경우
 - 인접 행렬의 차원(degree) 행렬 D 의 역행렬을 곱해주어 정규화(normalize)!

Approach

GCN

이미지 도메인에 주로 적용되는 2D 컨볼루션과 그래프 컨볼루션



모두 지역적인 정보를 취합하여 featur를 좀더 정교하게 추출하는 연산!

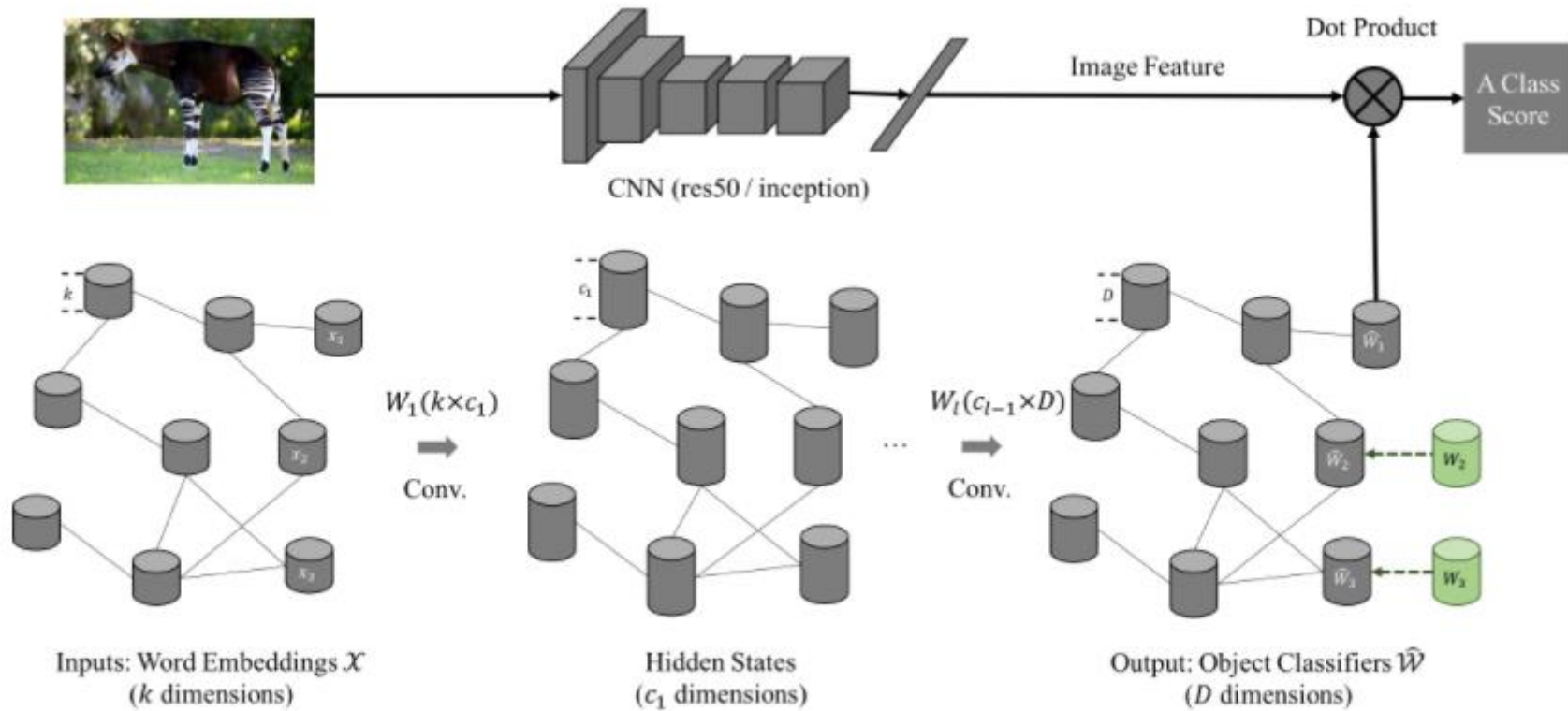
Approach

GCNZ

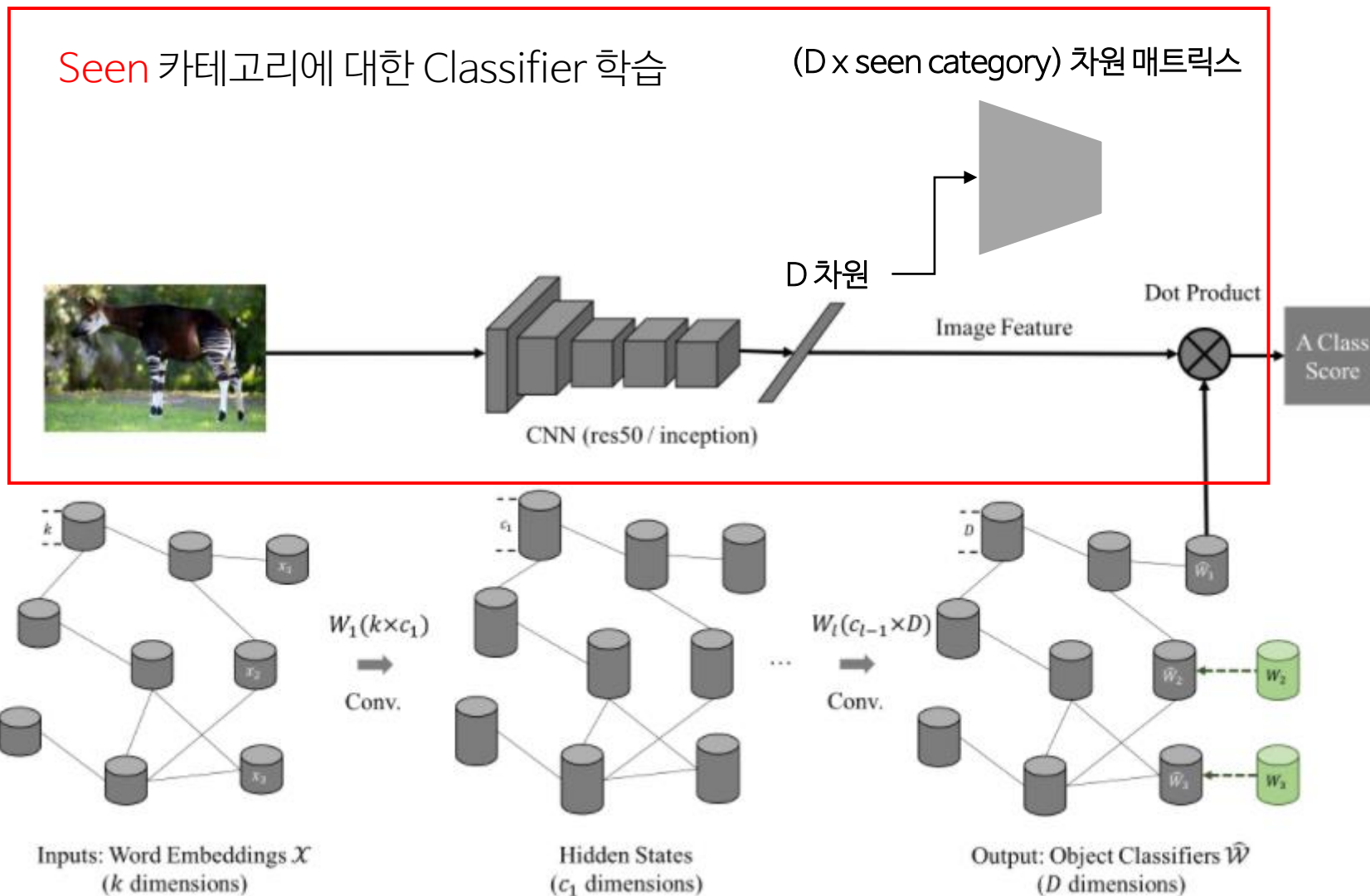
GCN for Zero-shot learning

Approach

GCNZ



Approach GCNZ



학습이 모두 완료되면,
분류를 해주는 매트릭스의 각 column 벡터는 각 카테고리에 대한 score를 계산해줌

Approach GCNZ

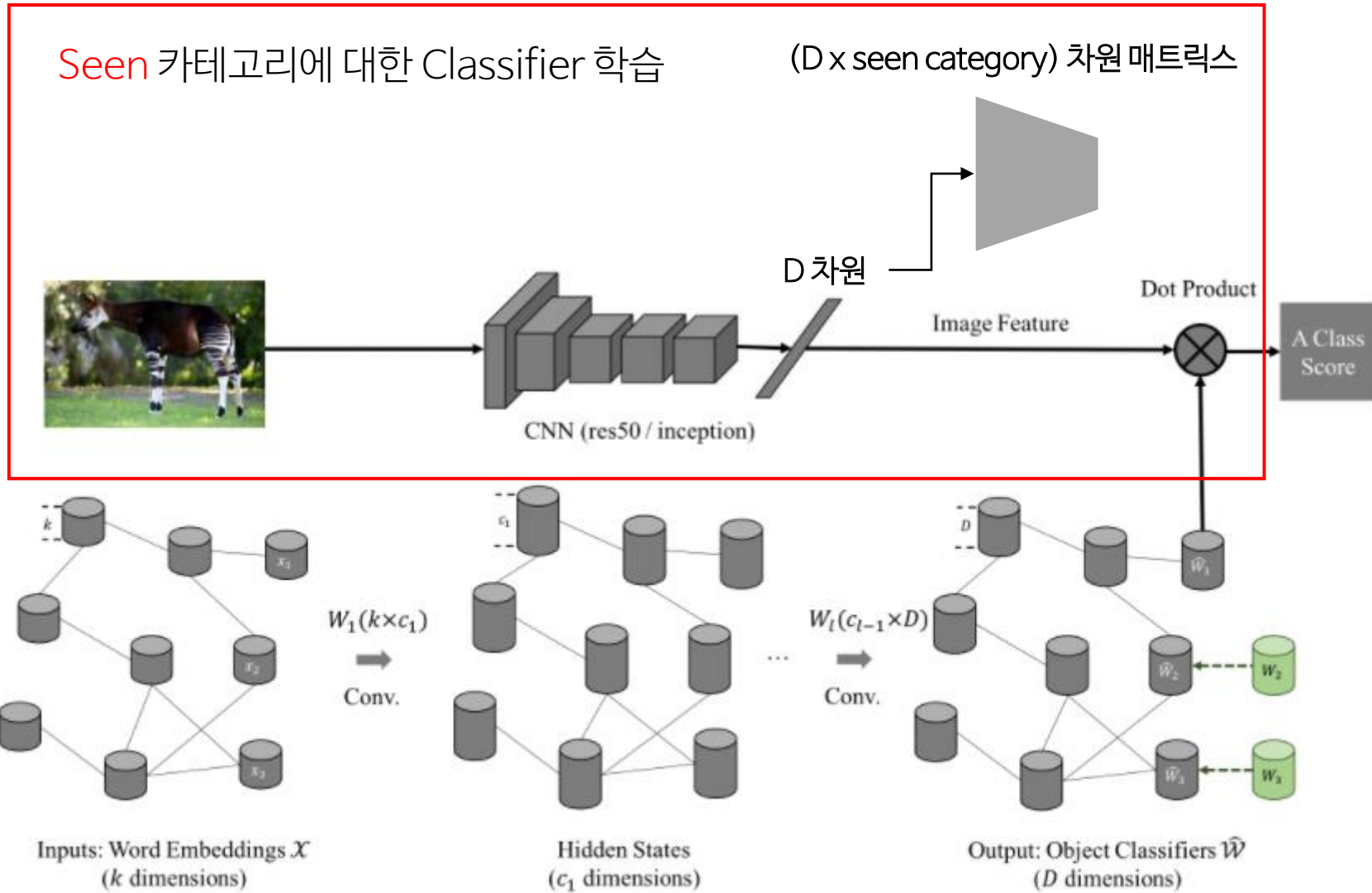
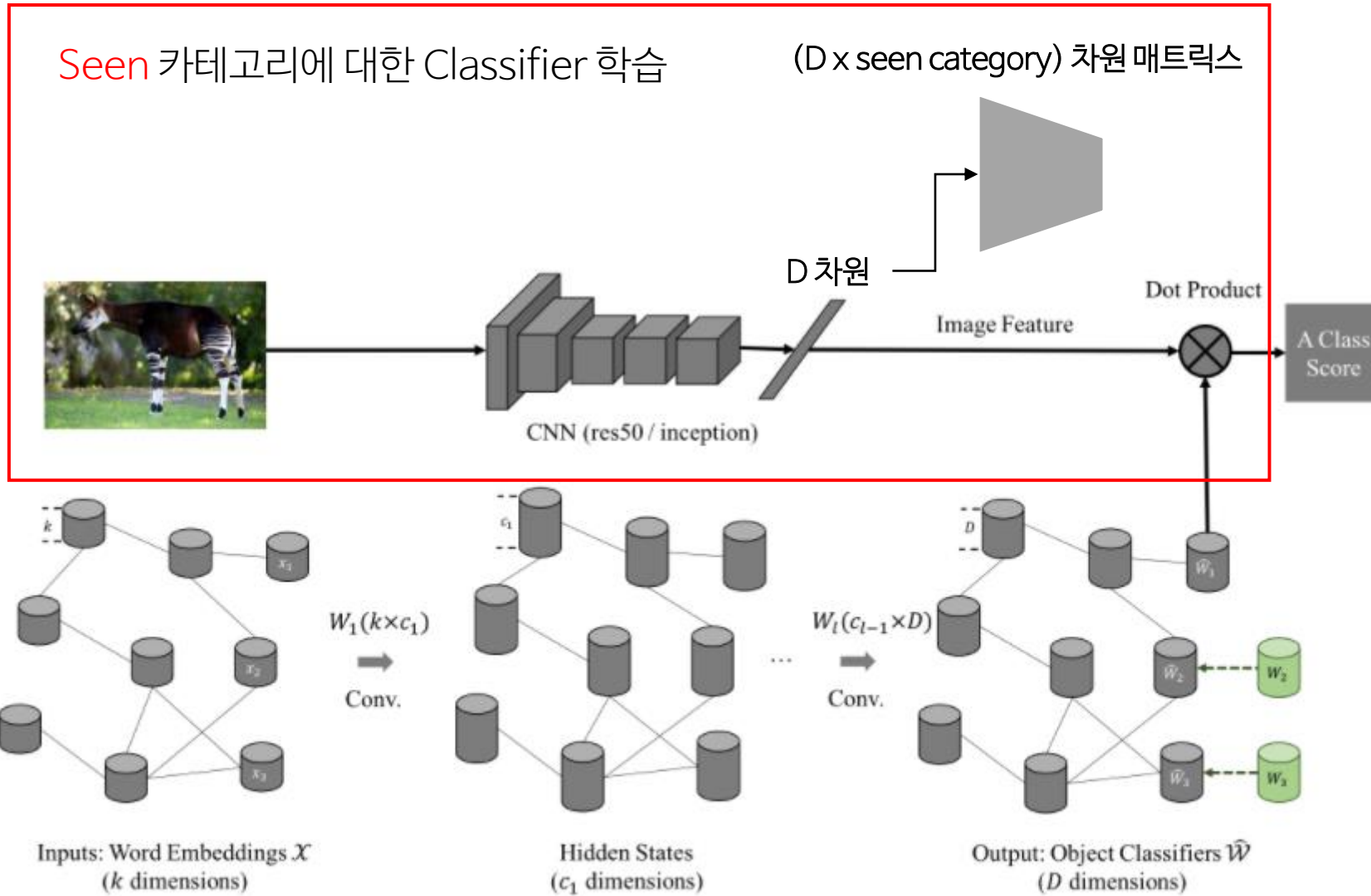


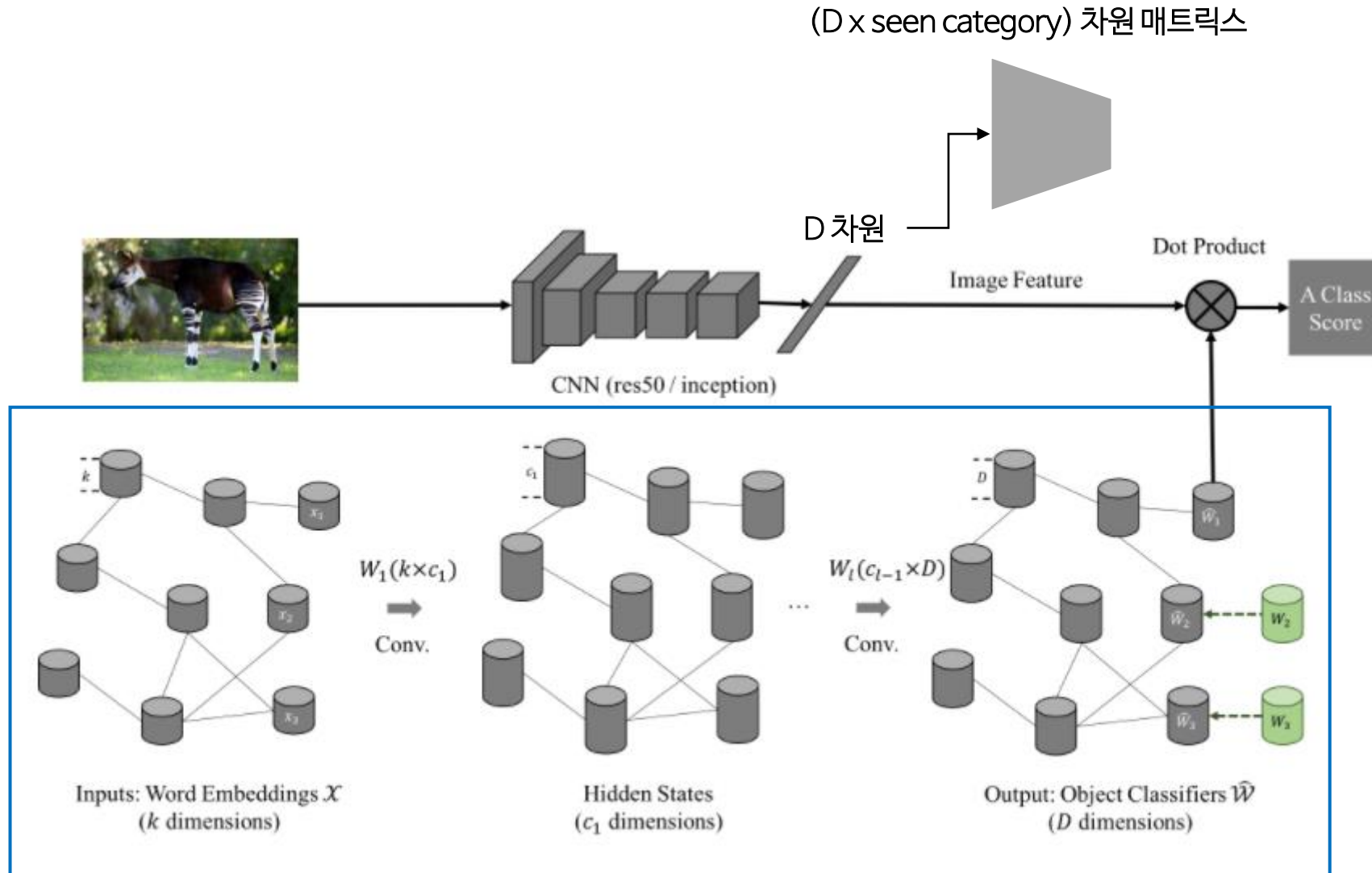
Image feature와 학습된 매트릭스의 각 column 벡터를 곱하면,
각 seen 카테고리일 score를 계산해줌

Approach GCNZ



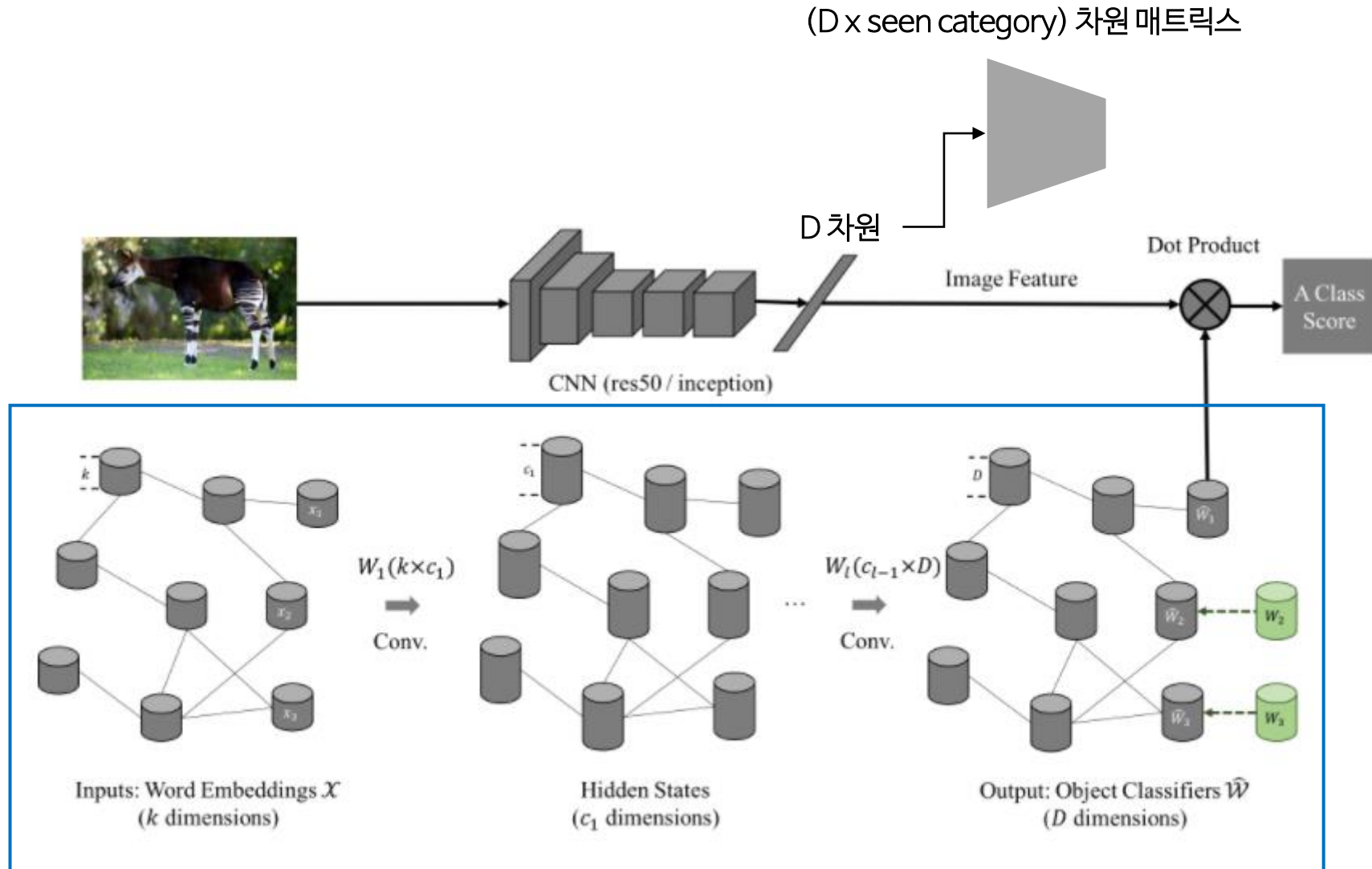
각 seen 카테고리에 해당하는 column 벡터를
각 seen 카테고리의 ground truth classifier 라고 부르자

Approach GCNZ



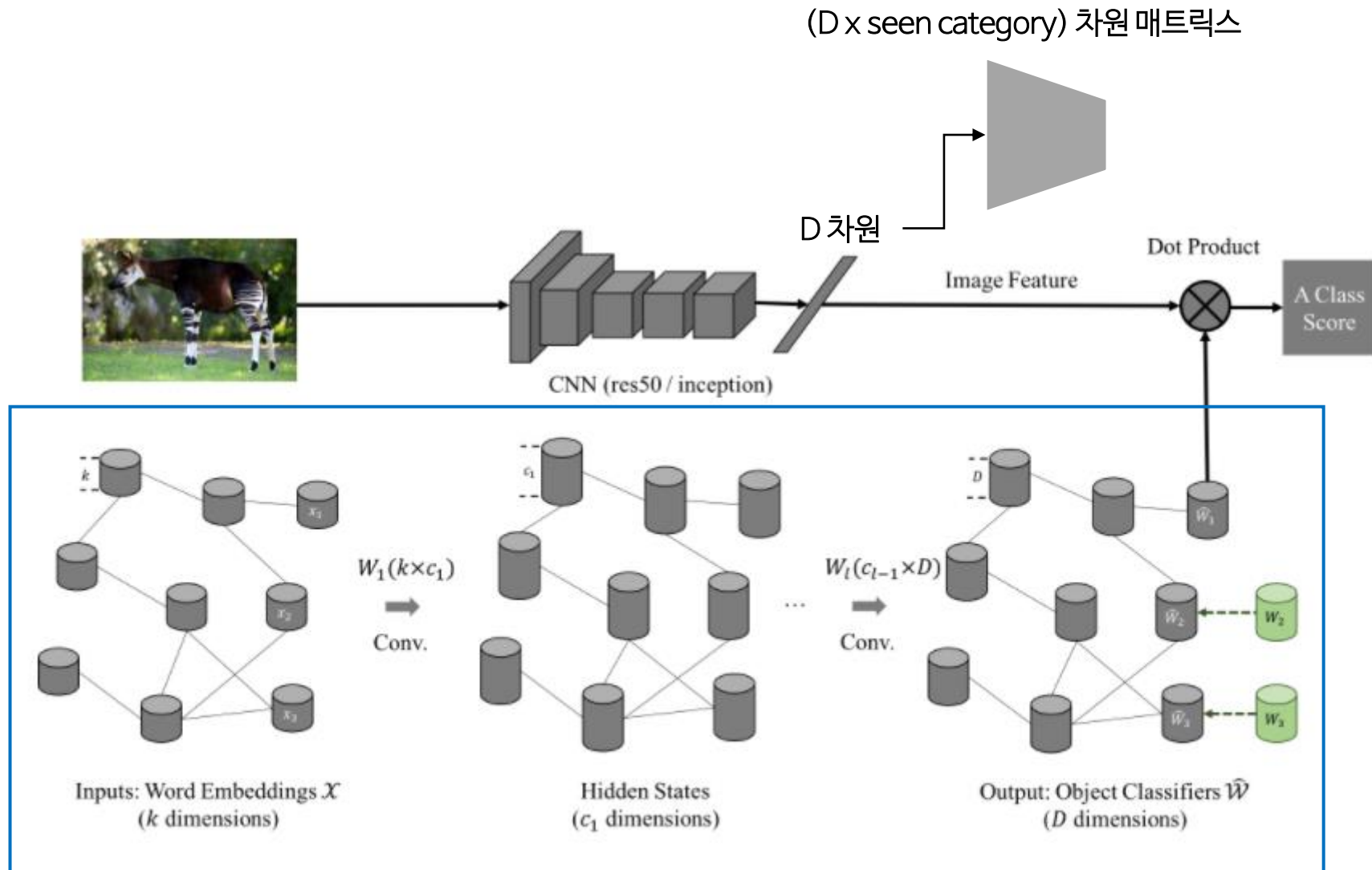
Seen & Unseen 모두를 포함하는 그래프
(WordNet 등을 사용)

Approach GCNZ



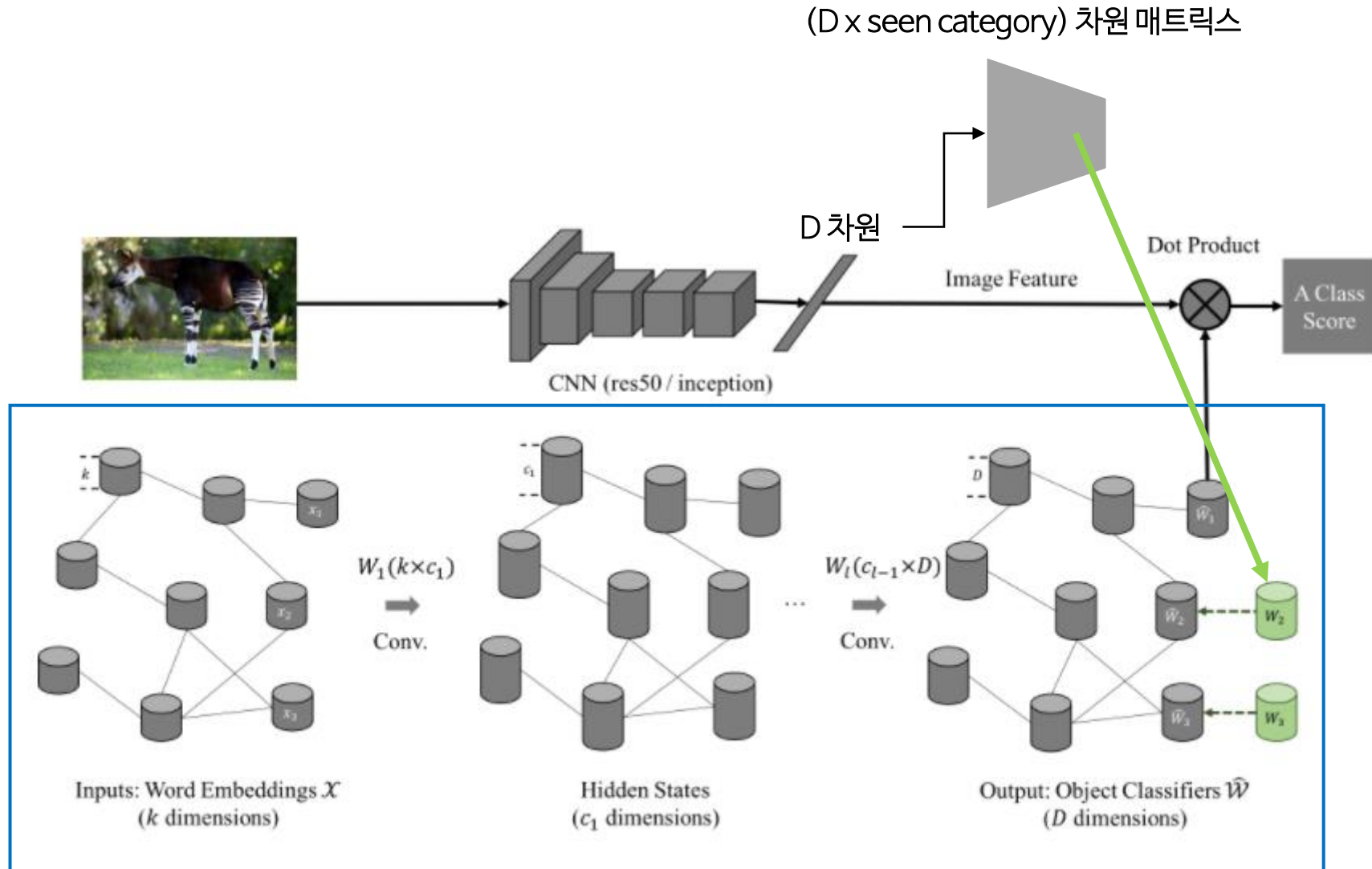
각 노드의 초기 feature 는 각 카테고리의 워드임베딩

Approach GCNZ



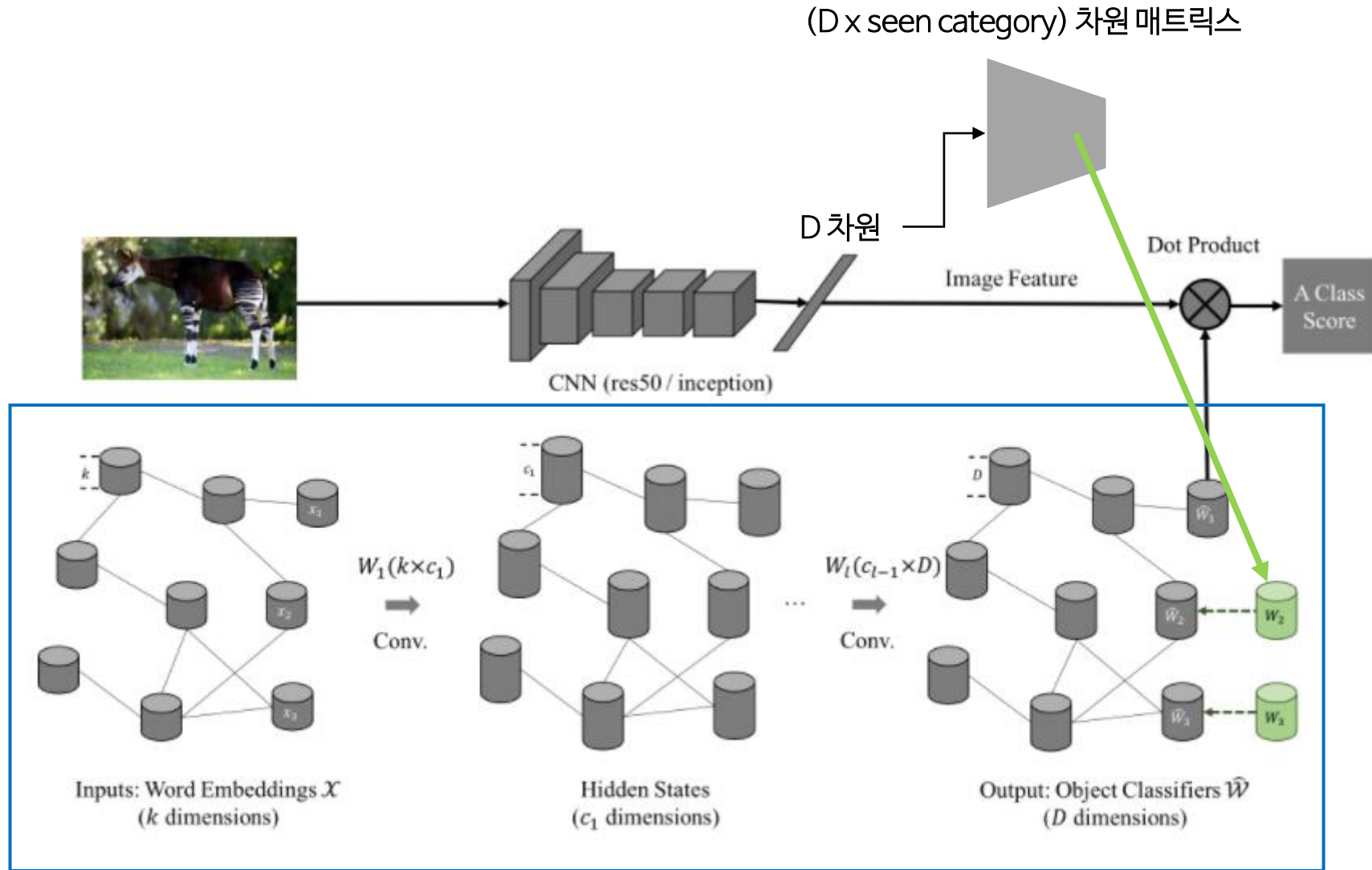
GCN 연산 수행

Approach GCNZ



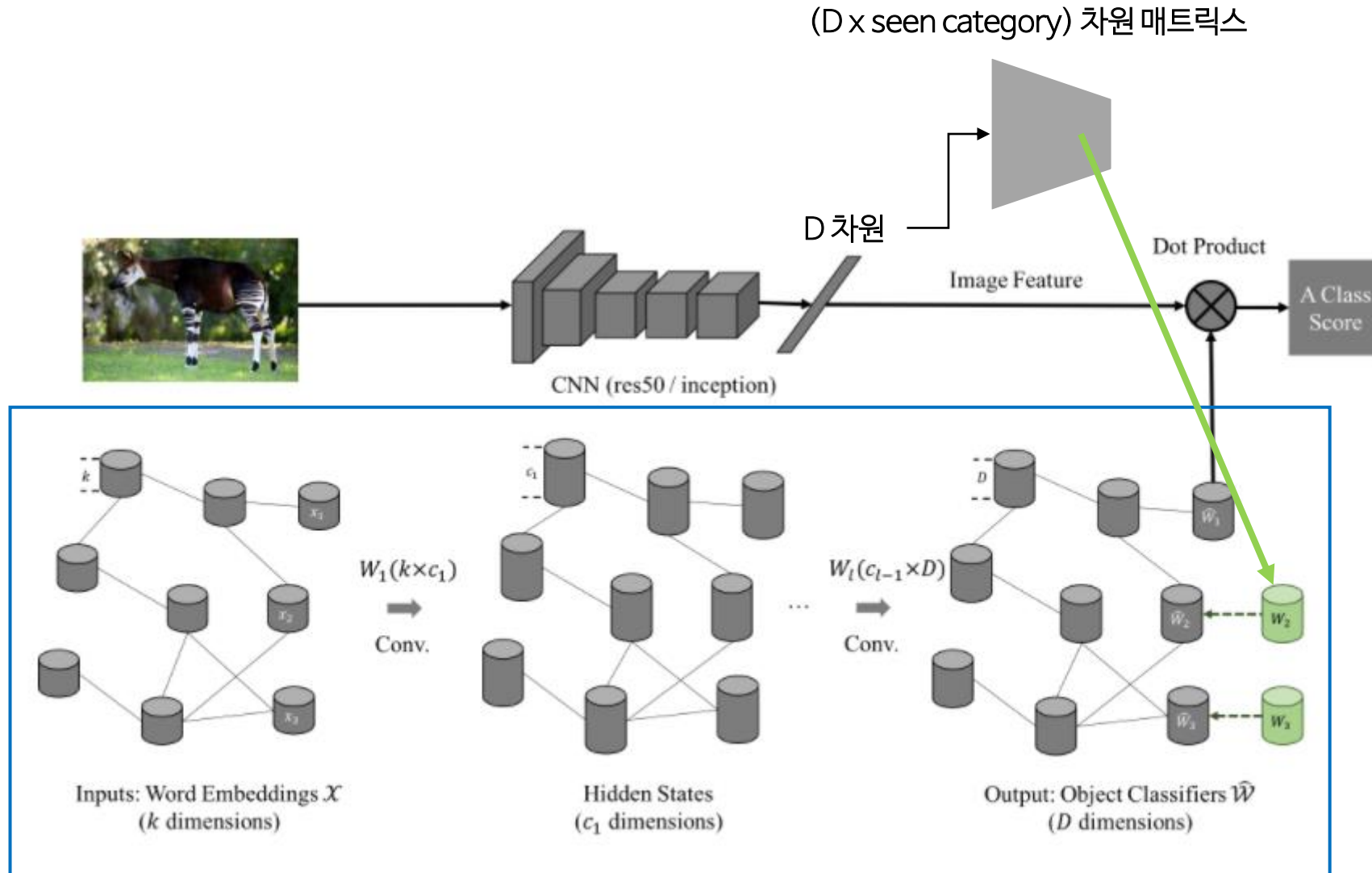
GCN 연산의 결과
 \approx seen 카테고리의 score 계산해주는 벡터

Approach GCNZ



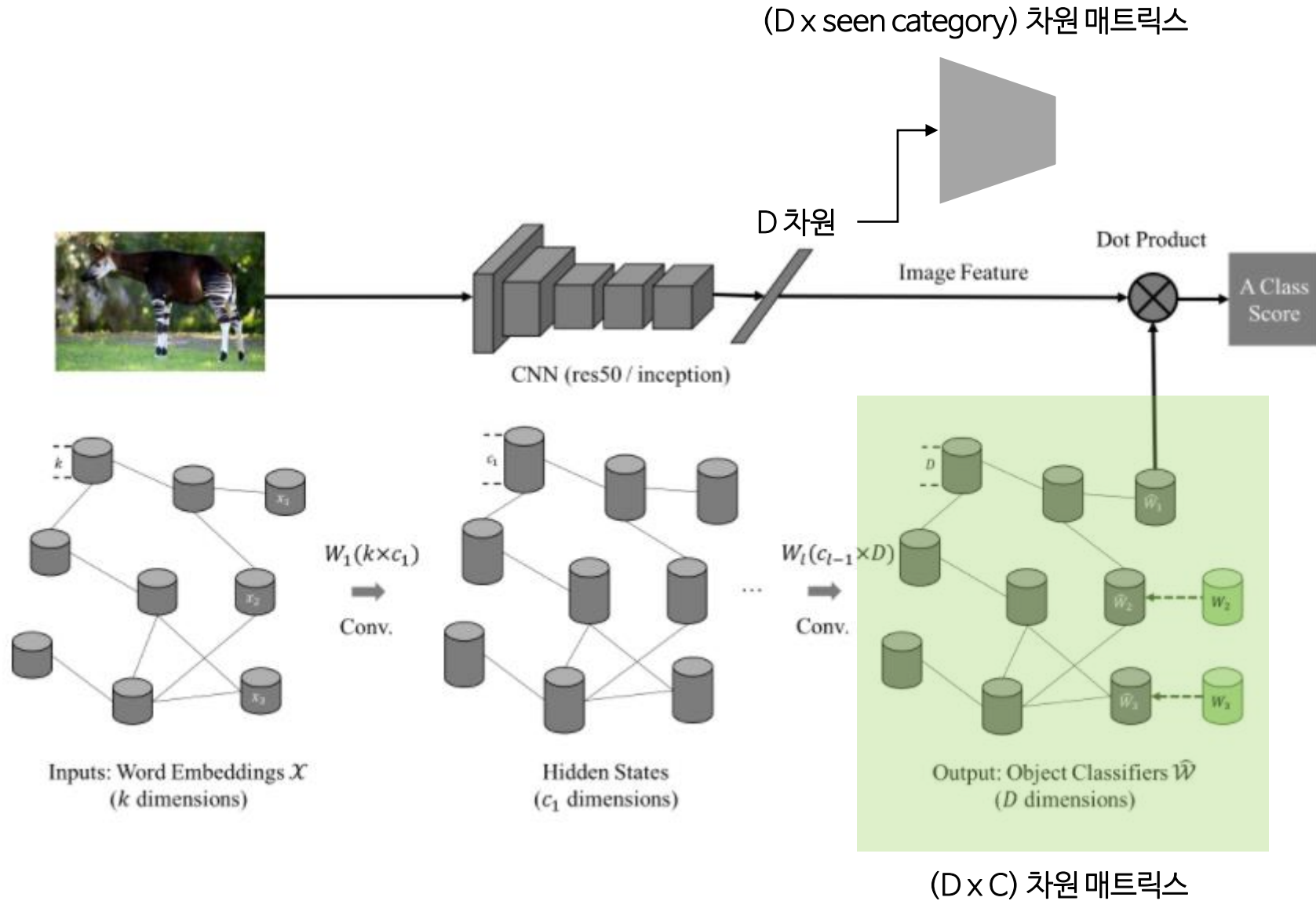
GCN 연산의 결과
 \approx seen 카테고리의 ground truth (GT) classifier

Approach GCNZ



GCN 연산의 결과와 GT classifier와의
MSE를 loss 함수로 사용하여 GCN 모델 학습

Approach GCNZ



C는 seen unseen 모두를 포함하는 클래스 개수

Experiment

Experiment

Classic zero-shot

ImageNet 데이터셋에 대한 실험 결과

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.

분류 대상이 **unseen** 만 있는 경우

Experiment

Generalized zero-shot

ImageNet 데이터셋에 대한 실험 결과

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).

분류 대상이 seen + unseen 인 경우

Conclusion

Conclusion

- 1) 이 논문에서는 Zero-shot recognition 태스크에 GCN 기법을 적용하는 방법론을 제시함
- 2) 기존 state-of-the-art 성능을 월등히 뛰어넘는 성능을 보임

Q & A

Thank you