Zero-Shot Recognition

Progressive Ensemble Networks for Zero-Shot Recognition

Ye, Meng, and Yuhong Guo. "Progressive ensemble networks for zero-shot recognition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

CONTENTS

CONTENTS

01 What is Zero-shot learning?

What is Zero-shot learning?

- 기계 학습에서 훈련 중 학습되지 않은 클래스를 테스트 때 예측하는 것

What is Zero-shot recognition?

- 이미지 classification에서, 학습되지 않은 클래스에 대한 이미지를 테스트 때 분류하는 것

01 What is Zero-shot learning?

Training (Seen class)

\+

Side-Information

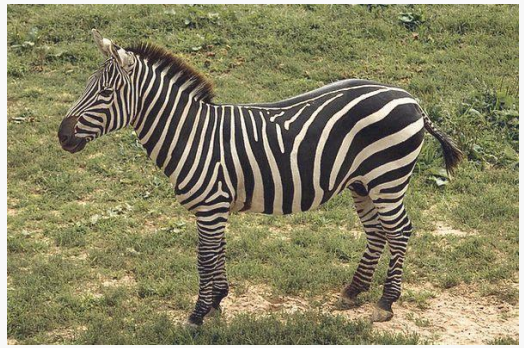
- **호랑이**
  현존하는 가장 큰 고양이 종
  주황색-갈색 털에 어두운 세로 줄무늬

=

Classification

호랑이

Test (Unseen class)



Side-Information

- 호랑이
- 말
- **얼룩말 : 검고 흰 얼룩무늬가 있는 말**
- 고양이
- ⋮

Classification

얼룩말

Why do we need Zero-shot recognition?

세상의 모든 물체를 분류하기 위해서는, 몇 개의 class가 필요할까?

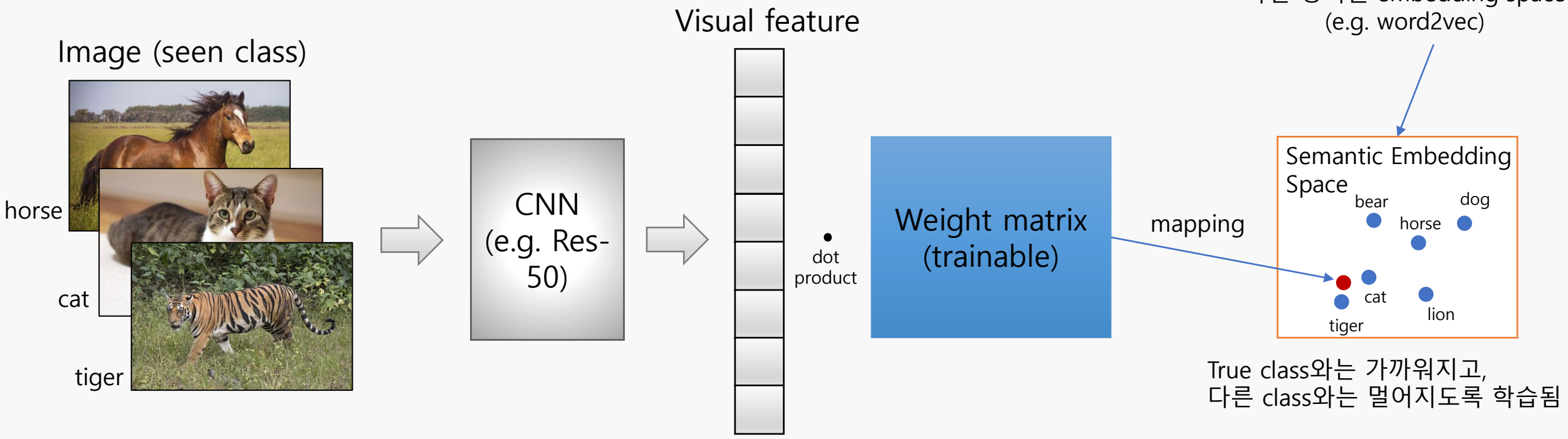
- ImageNet : 14,197,122 images, 21,841 categories
- Open Images : 59,000,000 images, nearly 20,000 categories

모든 클래스에 대한 이미지를 사람이 annotation 하는 것은 불가능함

→ 처음 보는 물체를 식별할 수 있는 능력이 필요

Simple architecture of Zero-shot recognition

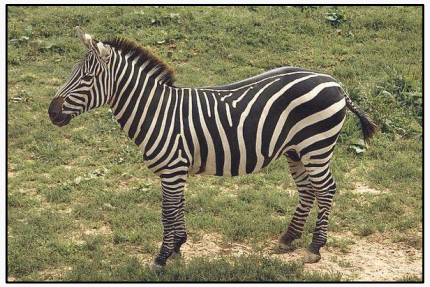
Training Time



Simple architecture of Zero-shot recognition

Test Time

Image (unseen class)



CNN (e.g. Res-50)

Visual feature

dot product

Weight matrix (trained)

mapping

Side-information을 통해 사전 정의된 embedding space (e.g. word2vec)

Semantic Embedding Space

bear dog horse cat tiger lion

mapping된 point와 가장 가까운 class를 찾는다

03 Different Zero-shot recognition settings

Progressive Ensemble Networks for Zero-Shot Recognition
: Generalized + Transductive setting

- Classical vs. Generalized setting

  - Classical

    unseen 데이터를 분류할 때, unseen 클래스에 확률만을 고려함

  - Generalized

    unseen 데이터를 분류할 때, seen + unseen 클래스 모두에 대한 확률을 고려함

    unseen 클래스 이미지가 입력되었을 때 모델이 seen 클래스로 더 잘 분류하는 경향이 있음

    → classification 성능 저하

03 Different Zero-shot recognition settings

Progressive Ensemble Networks for Zero-Shot Recognition
: Generalized + Transductive setting

- Inductive vs. Transductive setting

  - Inductive

    seen 클래스의 데이터로만 모델을 학습시킴

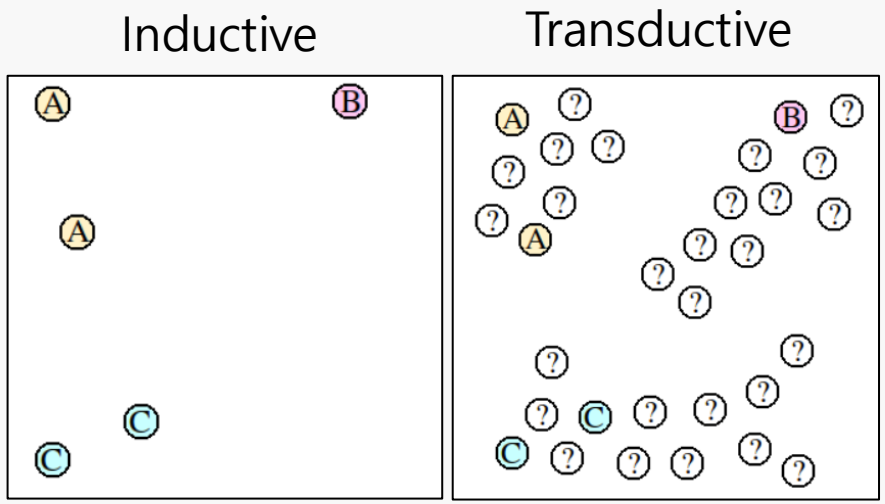
    → 잘 학습된 모델은 unseen 클래스 데이터도 잘 처리할거라 기대함

  - Transductive

    모델 학습 때 unseen 클래스의 데이터도 활용함

    (unseen 데이터의 true class는 주어지지 않음)

    → seen 데이터와 unseen 데이터의 관계를 추가로 학습함



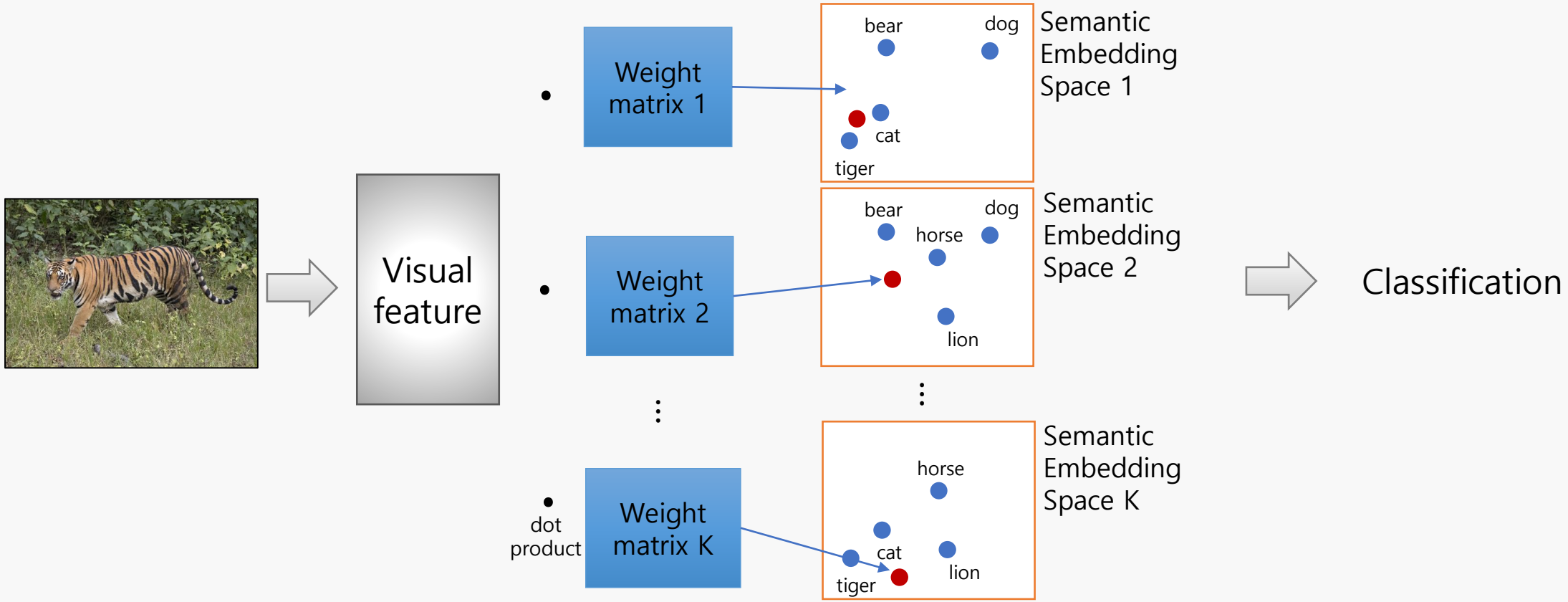
Progressive Ensemble Networks for Zero-Shot Recognition

- Training 과정에서 seen 클래스의 데이터와 unseen 클래스의 데이터를 모두 사용함 (=transductive)
  이 때 unseen 클래스의 데이터는 점진적으로(progressive) training 데이터에 통합됨
  → 모델이 training 데이터에 overfitting 되는 것을 방지함

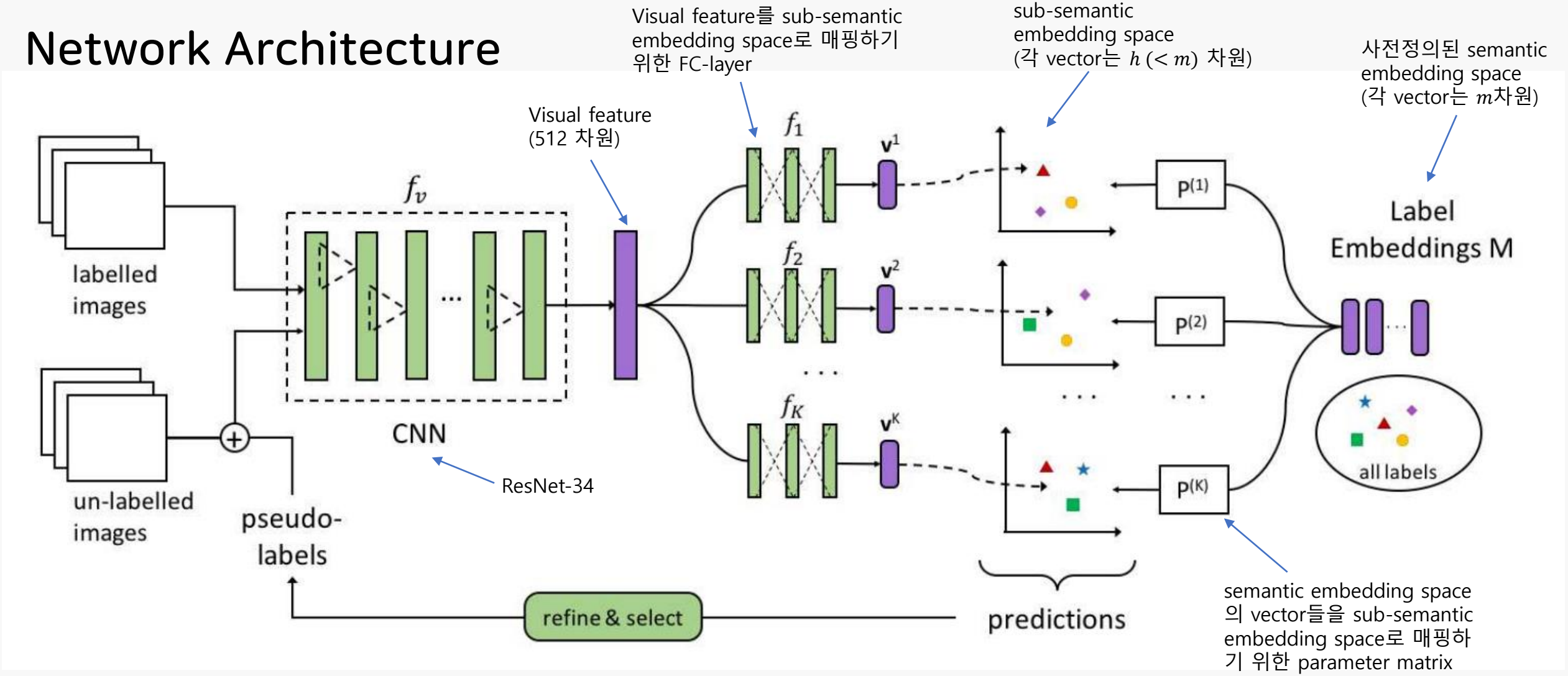
- class 예측에 여러 개의 classifier를 ensemble 방식으로 사용함
  → 모델의 robustness 증가

Ensemble 방식을 사용하는 이유는?

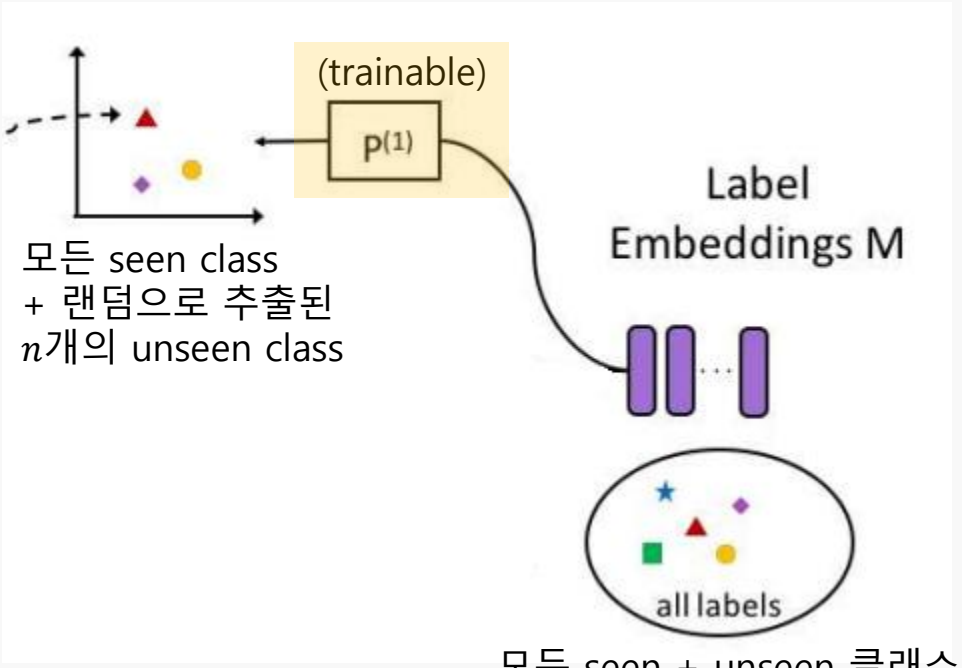
- Visual feature와 semantic embedding space 사이의 domain 차이가 큼
  → 하나의 mapping 만으로는 두 영역의 연관성을 충분히 학습할 수 없다



Network Architecture

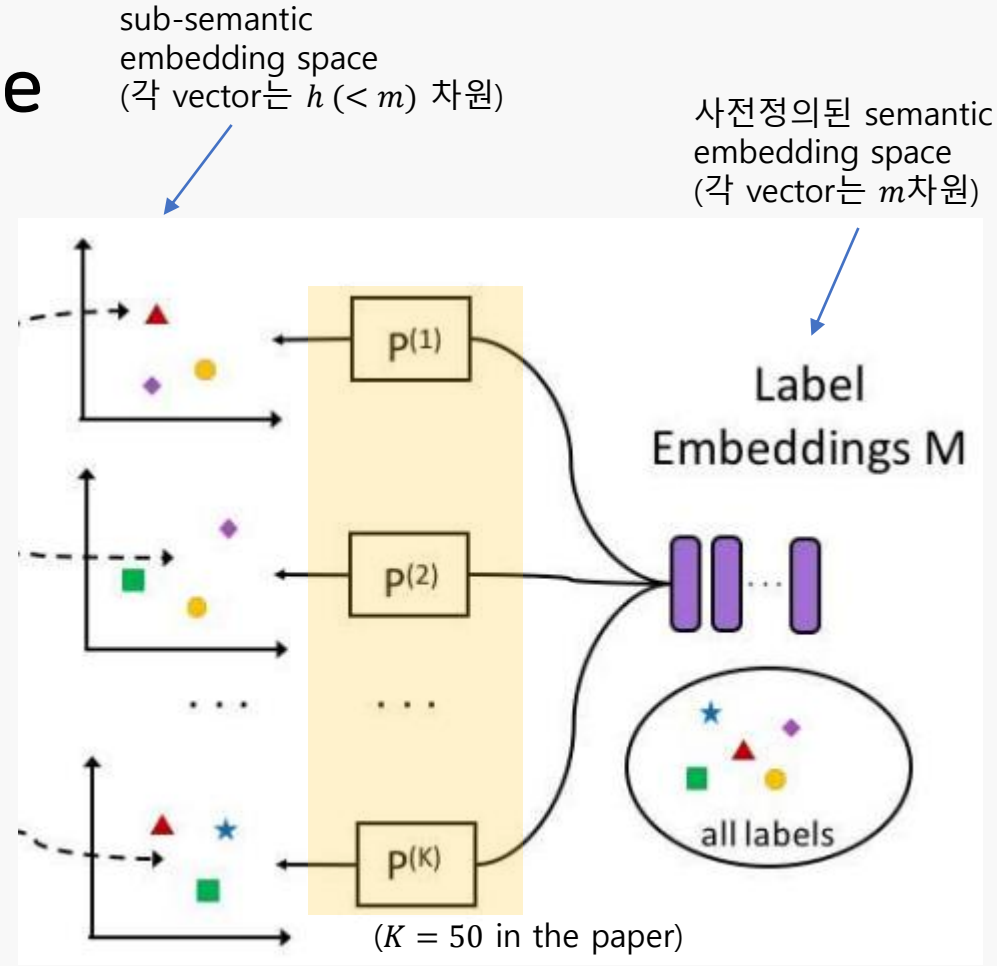


Training Sub-semantic embedding space



모든 seen class
+ 랜덤으로 추출된
n개의 unseen class

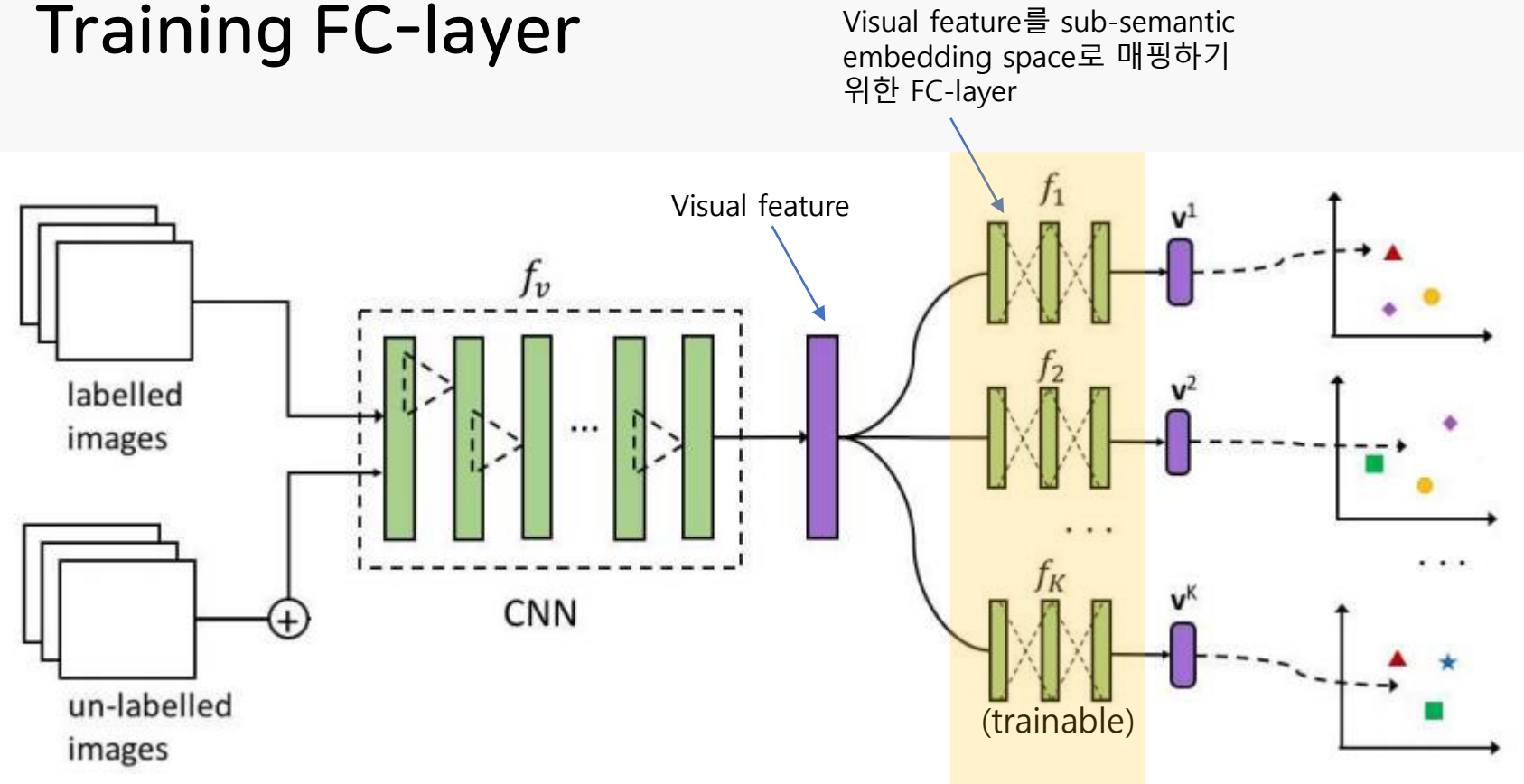
모든 seen + unseen 클래스



($K = 50$ in the paper)

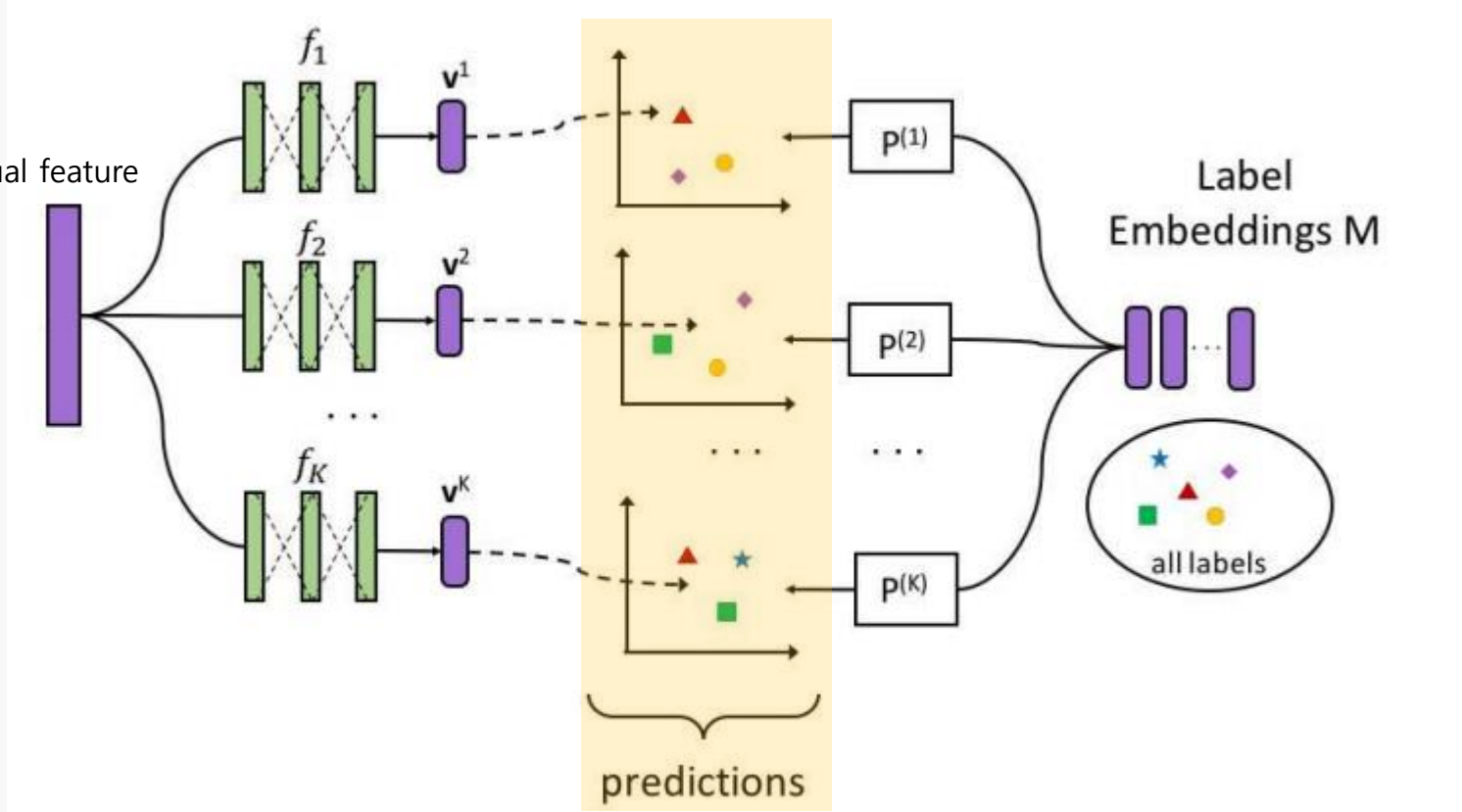
사전 정의된 semantic embedding space에서의 seen class와 unseen class 간의 유사도가 클 수록, sub-semantic embedding space 에서의 해당 클래스 간의 유사도가 커지도록 훈련됨

Training FC-layer



계산된 semantic vector ($\mathbf{v}^k$)와 sub-semantic embedding space 내의 true class vector의 유사도(내적)는 커지도록, 나머지 vector와의 유사도는 작아지도록 훈련됨

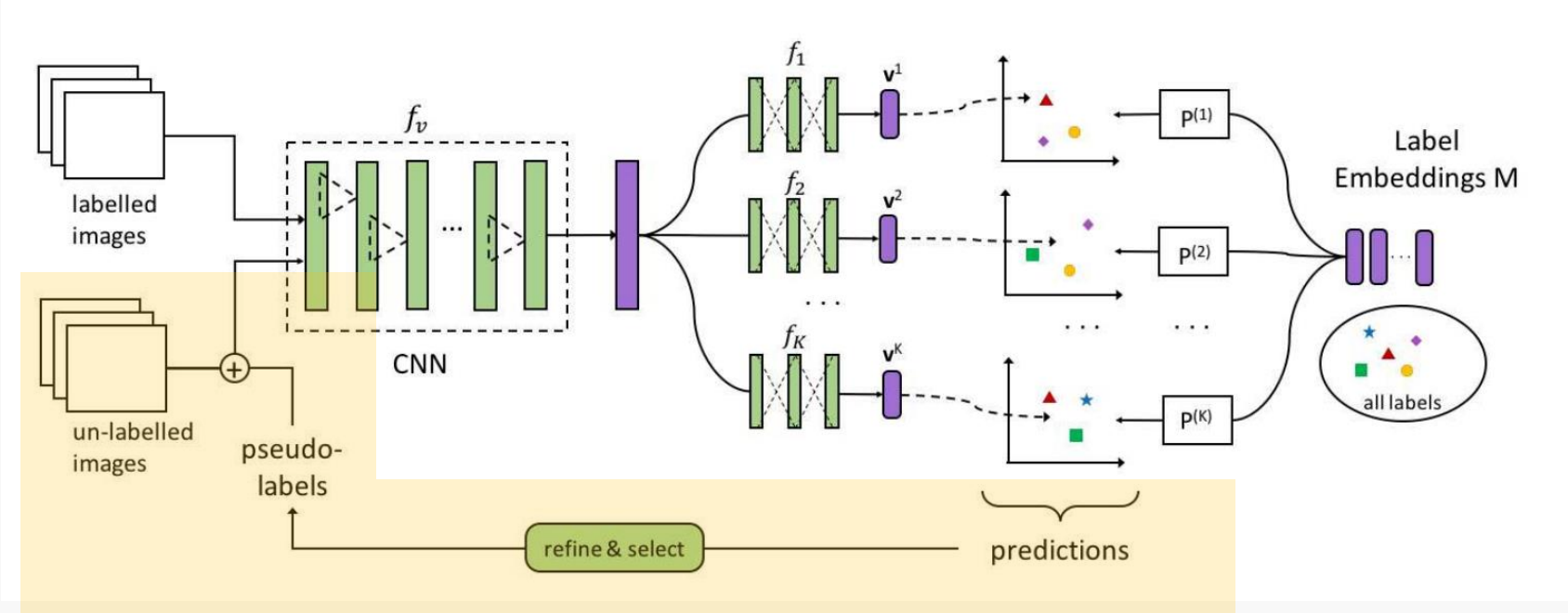
Prediction



K 개 sub-semantic embedding space 각각에서 $\mathbf{v}^k$ 와 가장 가까운 class를 선택한다.

→ 가장 많이 선택된 class를 최종 예측 class로 선택한다.

Using unseen(unlabeled) data during training



unseen 데이터를 training에 사용하기 위해, unseen 데이터에 pseudo-class를 추가하여 training 데이터에 추가한다.
증강된 training 데이터를 사용하여 네트워크를 다시 최적화한다.

Using unseen(unlabeled) data during training

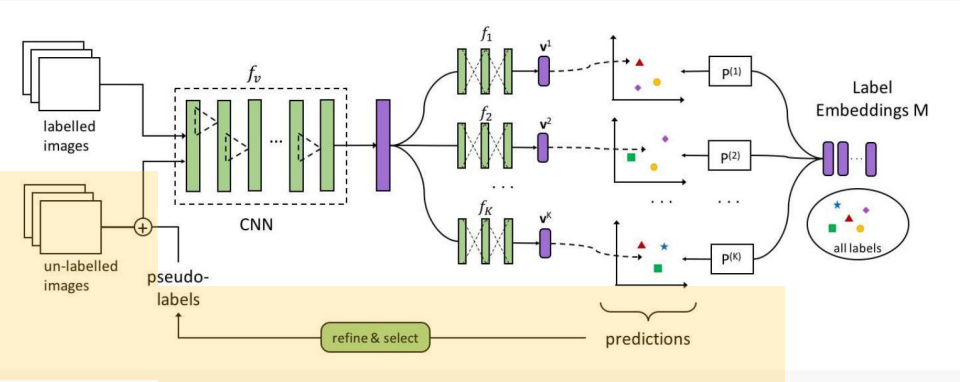
Unseen class

Unseen data	c_1	c_2	c_3	⋯
x_1	**0.4**	**0.2**	0.1	
x_2	**0.3**	0.1	**0.5**	
x_3	0.2	0.1	0.1	
x_4	0.1	**0.5**	**0.2**	
⋮				⋱

(K개 sub-semantic embedding space 모두에서)

Score(x_i, c_j) =

$$\frac{x_i\text{에 의해 예측된 } class\text{가 } c_j\text{인 수}}{c_j \text{ 가 포함된 sub semantic embedding space 수}}$$



1. 각 unseen class마다 score를 기준으로 top N개의 unseen 데이터를 뽑는다.
2. 뽑힌 데이터 쌍 (x_i, c_j)을 training 데이터에 추가한다. (c_j는 x_i에 대한 pseudo-class)
3. 증강된 training 데이터로 네트워크를 학습시킨다.
4. 추가된 데이터를 제거하고, 1번부터 다시 반복한다.

THANK YOU –

경청해주셔서 **감사합니다.**