

zero shot classification

0224

zero shot learning

여러 언어로 sentence space를 만들면
다른 언어라도 같은 의미를 가진 문장들이 한 곳에 모여있다



이미 번역된 결과를 학습하지 않고서도 서로 다른 나라 언어를 번역할 수 있다
ex) 영어-한국어 & 영어-일본어 -> 한국어-일본어

모델 아키텍처를 변경할 필요 없이
인공 토큰을 입력한 문장을 붙여 필요한 언어로 번역할 수 있다.

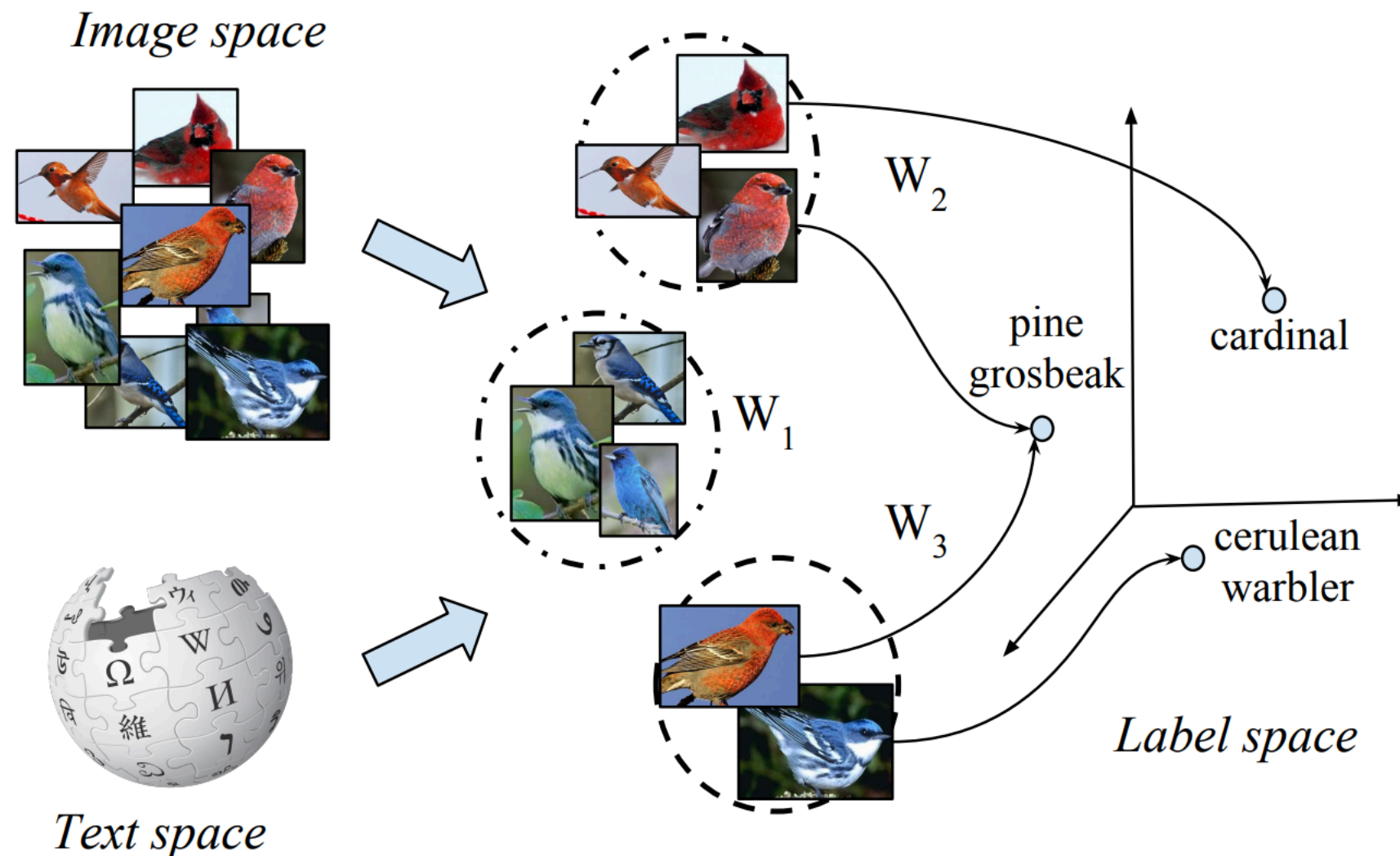
abstract ~ 나는 이런 문제를 풀거야 ~

- 쌍 선형 모델 제안
- 주어진 이미지에 대해 잘못된 클래스에 라벨링했을 때 불이익을 주는 ranking 모델 제시

introduction

~ 사실 이 문제는 이런 동기에서 연구가 시작된건데 ~

- 말뚱치로부터 추출해낸 벡터를 이용해 탐구
- 비슷한 클래스는 가까운 곳으로 끌어당기고 다른 것은 멀게 만든다
- 여기에선, 선형 모델을 수집하면서 이미지-클래스 쌍을 선택할 수 있게 함



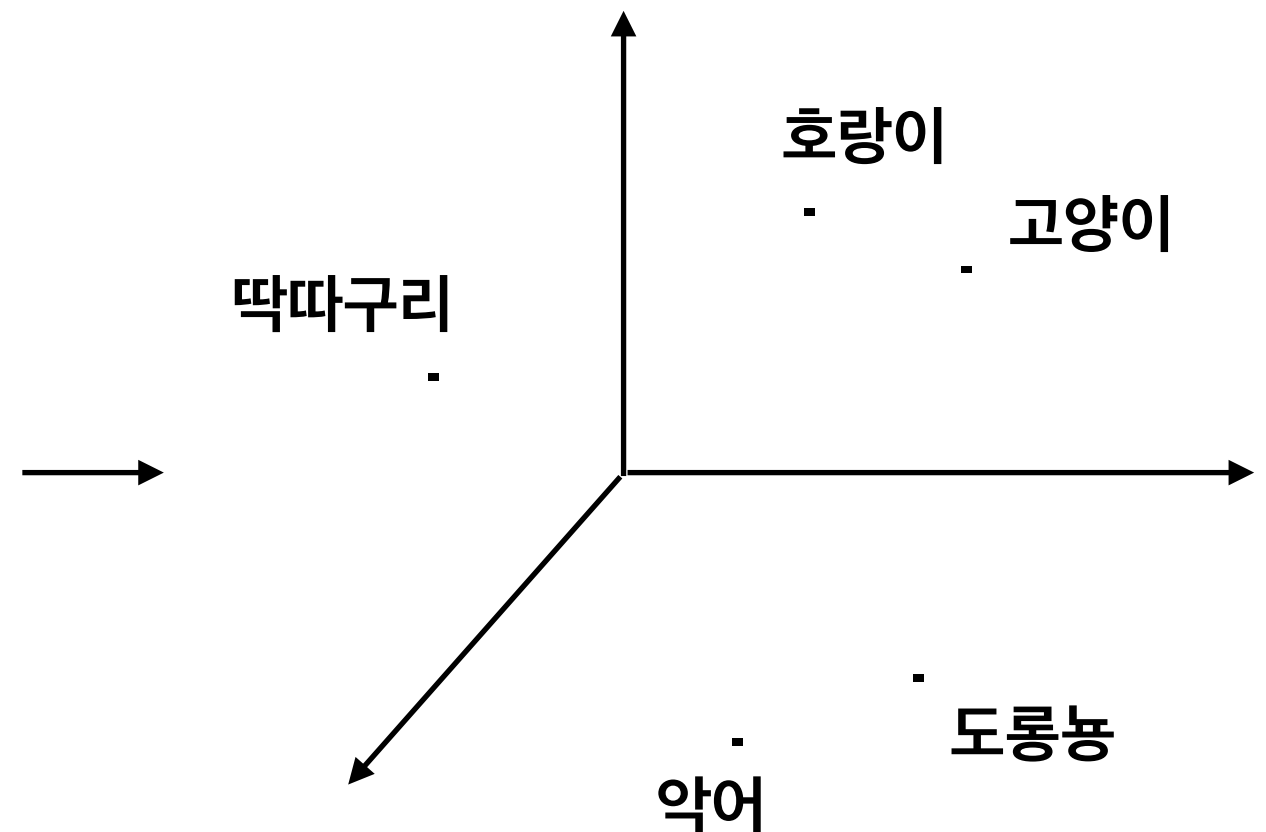
related works

~ 관련해서 이런 접근이 있었어 ~

- zero-shot learning에서는 2차 정보를 말뭉치로 사용 (word2vec, glove)
- visual 특징들을 캡처

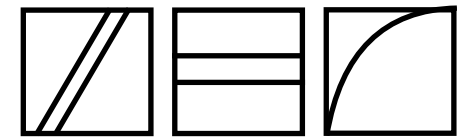
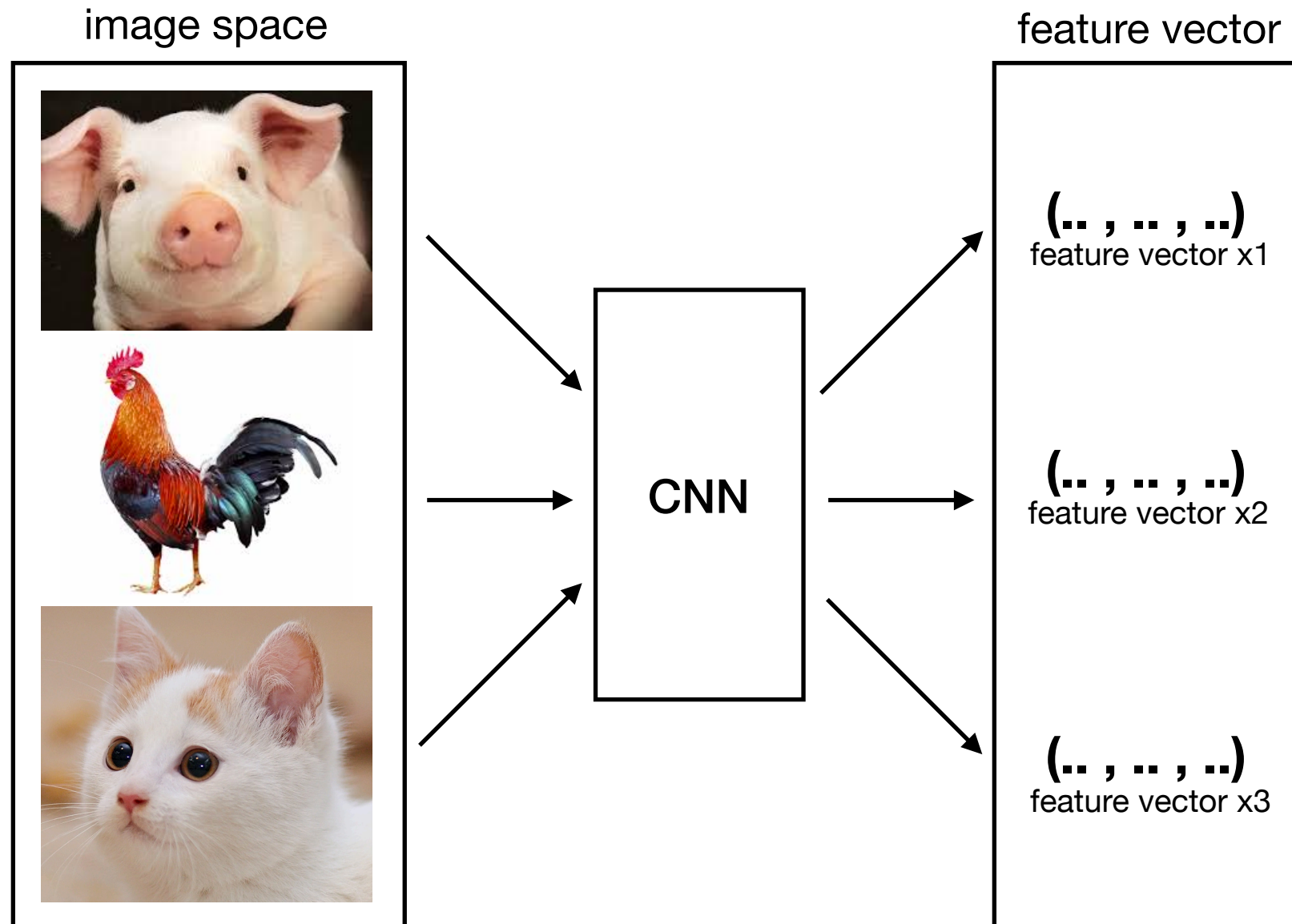
Cats are similar in anatomy to the other felid species, with a strong **flexible body**, quick reflexes, **sharp teeth** and **retractable claws** adapted to killing small prey.

The **tiger** (*Panthera tigris*) is the **largest** cat species, most recognizable for its pattern of **dark vertical stripes** on **reddish-orange fur** with a lighter underside.



related works

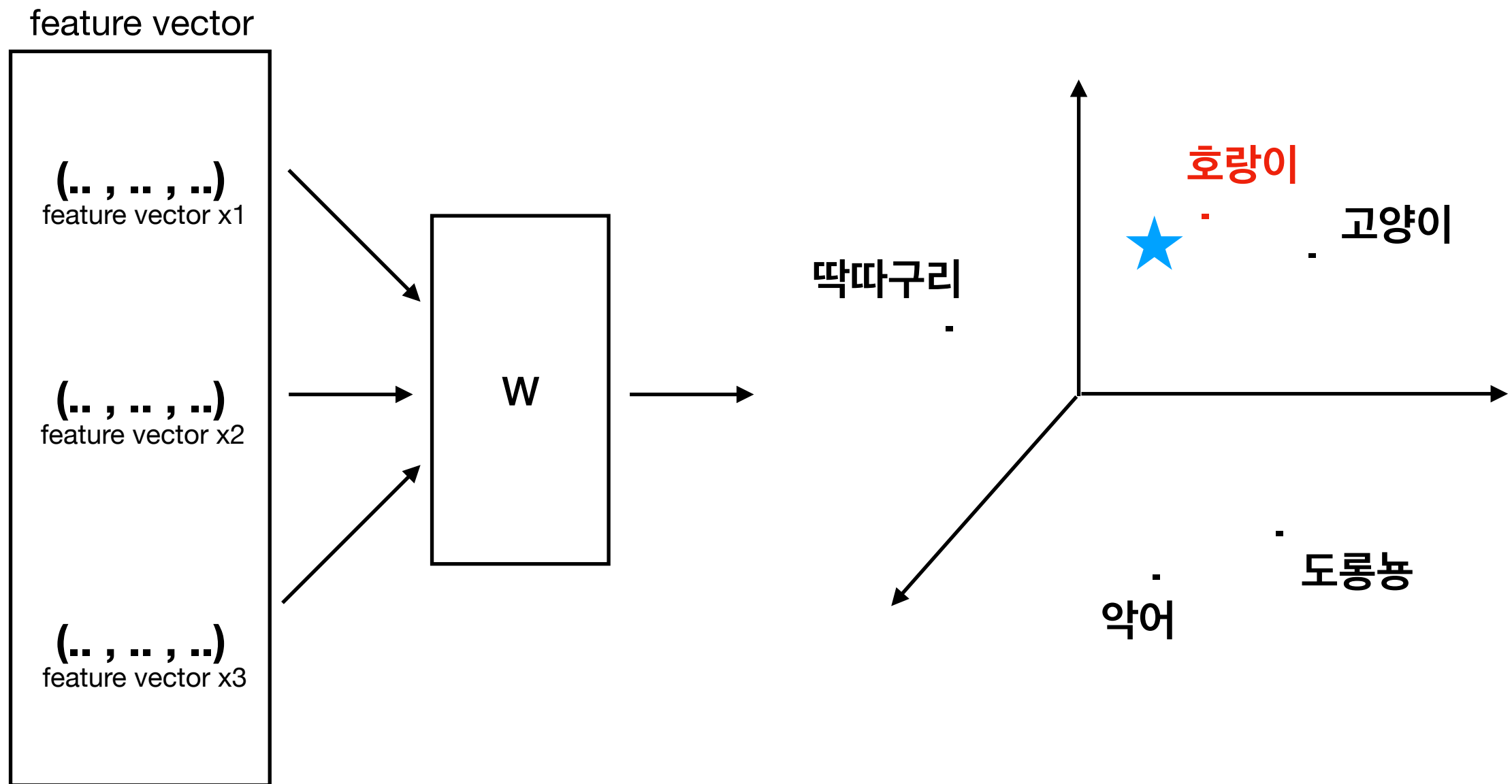
~ 관련해서 이런 접근이 있었어 ~



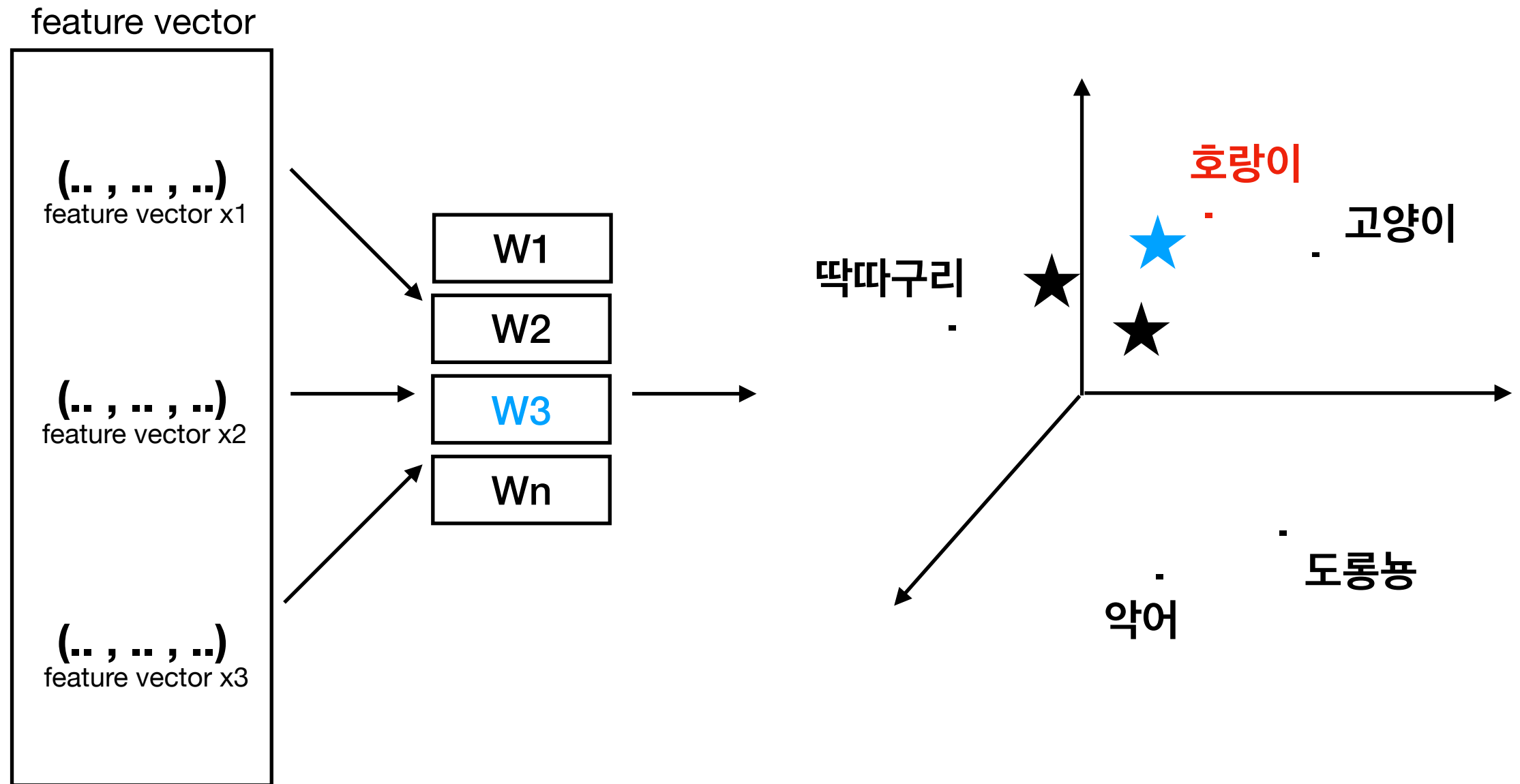
: \mathbf{X}

$\mathcal{Y} \subset \mathbb{R}^{d_y}\}$ where \mathbf{x} is the image embedding defined in an image feature space \mathcal{X} , e.g. CNN features [19], and \mathbf{y} is the

method ~ 나는 이런 방식으로 접근할건데 ~



method



method

$$F(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq K} \mathbf{x}^\top W_i \mathbf{y}. \quad (4)$$

K : number of W

y
target y

Our main goal is to learn a set of compatibility spaces that minimizes the following empirical risk,

$$\frac{1}{N} \sum_{n=1}^{|\mathcal{T}|} L(\mathbf{x}_n, \mathbf{y}_n), \quad (5)$$

where $L : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ is the loss function defined for a particular example $(\mathbf{x}_n, \mathbf{y}_n)$ as

$$L(\mathbf{x}_n, \mathbf{y}_n) = \sum_{\mathbf{y} \in \mathcal{Y}} \max\{0, \Delta(\mathbf{y}_n, \mathbf{y}) + F(\mathbf{x}_n, \mathbf{y}) - F(\mathbf{x}_n, \mathbf{y}_n)\} \quad (6)$$

where $\Delta(\mathbf{y}, \mathbf{y}_n) = 1$ if $\mathbf{y} \neq \mathbf{y}_n$ and 0 otherwise. This

method

Algorithm 1 SGD optimization for LatEm

epochs $\mathcal{T} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \in \mathbb{R}^{d_x}, \mathbf{y} \in \mathbb{R}^{d_y}\}$

- 1: **for all** $t = 1$ to T **do**
- 2: **for all** $n = 1$ to $|\mathcal{T}|$ **do**
- 3: Draw $(\mathbf{x}_n, \mathbf{y}_n) \in \mathcal{T}$ and $\mathbf{y} \in \mathcal{Y} \setminus \{\mathbf{y}_n\}$!= y_n
- 4: **if** $F(\mathbf{x}_n, \mathbf{y}) + 1 > F(\mathbf{x}_n, \mathbf{y}_n)$ **then**
- 5: $i^* \leftarrow \arg \max_{1 \leq k \leq K} \mathbf{x}_n^\top W_k \mathbf{y}$: 예측
- 6: $j^* \leftarrow \arg \max_{1 \leq k \leq K} \mathbf{x}_n^\top W_k \mathbf{y}_n$: target
- 7: **if** $i^* = j^*$ **then**
- 8: $W_{i^*}^{t+1} \leftarrow W_{i^*}^t - \eta_t \mathbf{x}_n (\mathbf{y} - \mathbf{y}_n)^\top$ update matrices
- 9: **end if**
- 10: **if** $i^* \neq j^*$ **then**
- 11: $W_{i^*}^{t+1} \leftarrow W_{i^*}^t - \eta_t \mathbf{x}_n \mathbf{y}^\top$
- 12: $W_{j^*}^{t+1} \leftarrow W_{j^*}^t + \eta_t \mathbf{x}_n \mathbf{y}_n^\top$ update using SGD
- 13: **end if**
- 14: **end if**
- 15: **end for**
- 16: **end for**

experiment ~ 실험도 해봤어 ~

	Total		train+val		test	
	imgs	cls	imgs	cls	imgs	cls
CUB	11786	200	8855	150	2931	50
AWA	30473	50	24293	40	6180	10
Dogs	19499	113	14681	85	4818	28

비교비교
SJE vs LatEm

	CUB		AWA		Dogs	
	SJE	LatEm	SJE	LatEm	SJE	LatEm
att	50.1	45.5	66.7	71.9	N/A	N/A
w2v	28.4	31.8	51.2	61.1	19.6	22.6
glo	24.2	32.5	58.8	62.9	17.8	20.9
hie	20.6	24.2	51.2	57.5	24.3	25.2

discussion

~ 이런것도 알아 났지만 한계점은 이래 ~

:)

conclusion ~ 3줄 요약 ~

- text corpora를 이용해 만든 w2v와 image space를 한 공간에 두기로 함
- 비슷하면 끌어당기고 다르면 밀어내는 성질을 이용 (ex. 구글 실험)
- 모델 수 K 정하기 : W 늘린 후 5회 알고리즘 돌리고 결과 5%미만인 W 제거