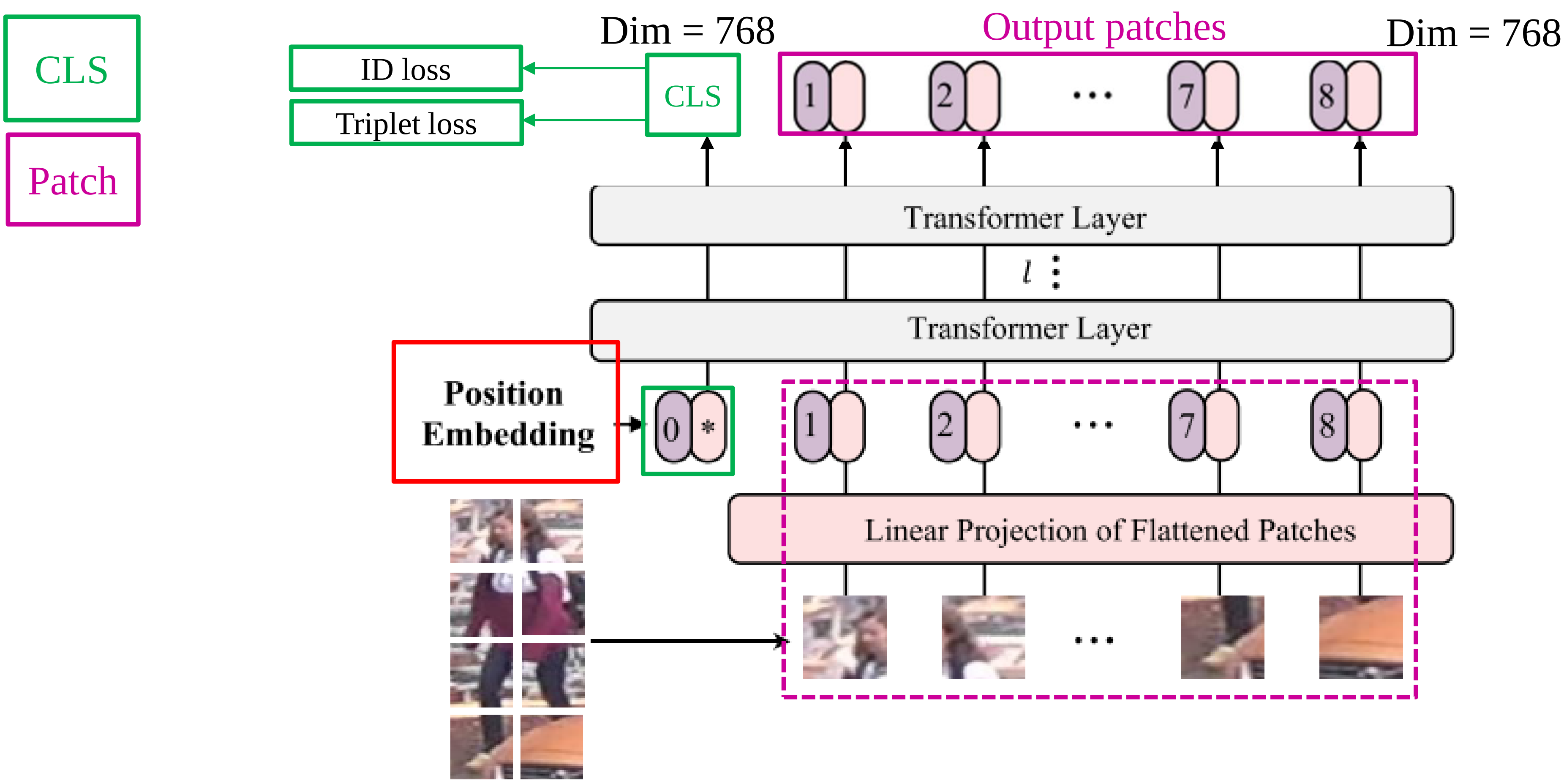
This week

- *Person Re-identification using ViT*
  - *Relation between position encoding & person re-identification*



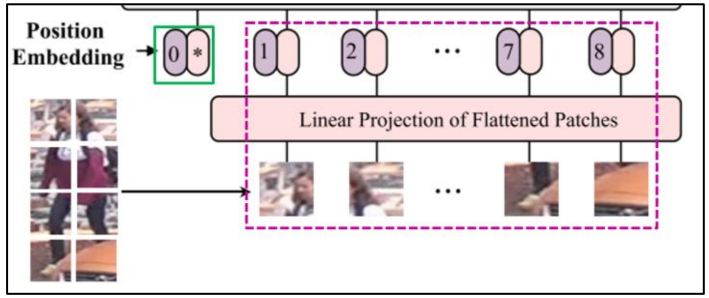
김성수

This week

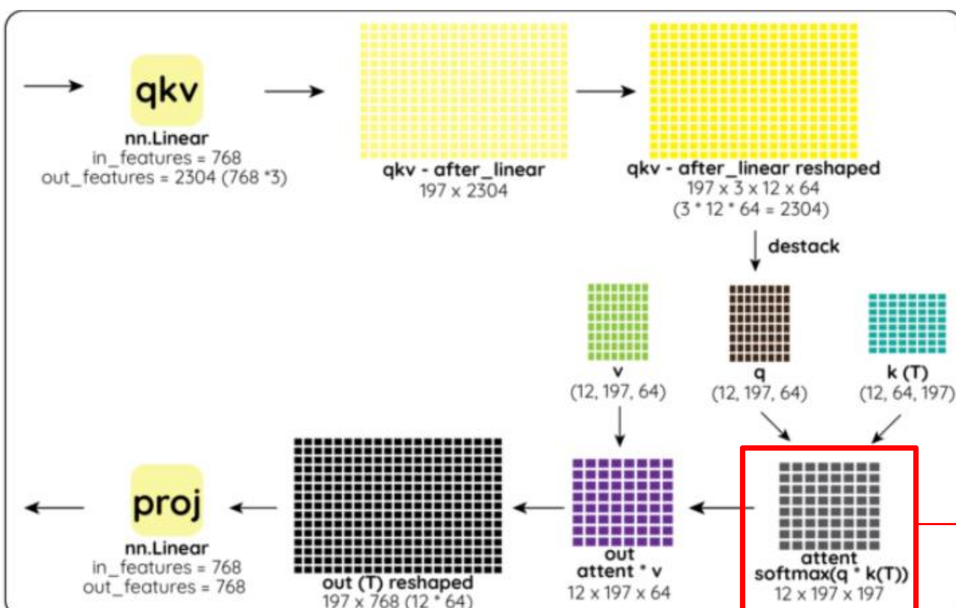
[4]P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," *arXiv preprint arXiv:1803.02155, 2018.*

Positional encoding(PE)

- APE : Patch가 encoder에 들어가기 직전



- RPE[4] : Self-attention mechanism 내부



$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{\boxed{QK^T}}{\sqrt{d_k}}\right)V$$

Learnable Bias term

$$e_{ij} = \frac{x_i W^Q (x_j W^K + \boxed{a_{ij}^K})^T}{\sqrt{d_z}}$$

Proposed method

- *Relative position triplet loss*
  - *Idea : "같은 ID의 image끼리는 주요 patch들간 relative position bias가 비슷할 것이다."*

A	0,0	0,1	B
C	1,0	1,1	D

Absolute position

0	1	2	3	4	5	6	7	8
0.1	0.03	0.8	0.5	0.07	0.2	0.4	0.01	0.06

Relative position bias table

	A	B	C	D
A	4	3	1	0
B	5	4	2	1
C	7	6	4	3
D	8	7	5	4

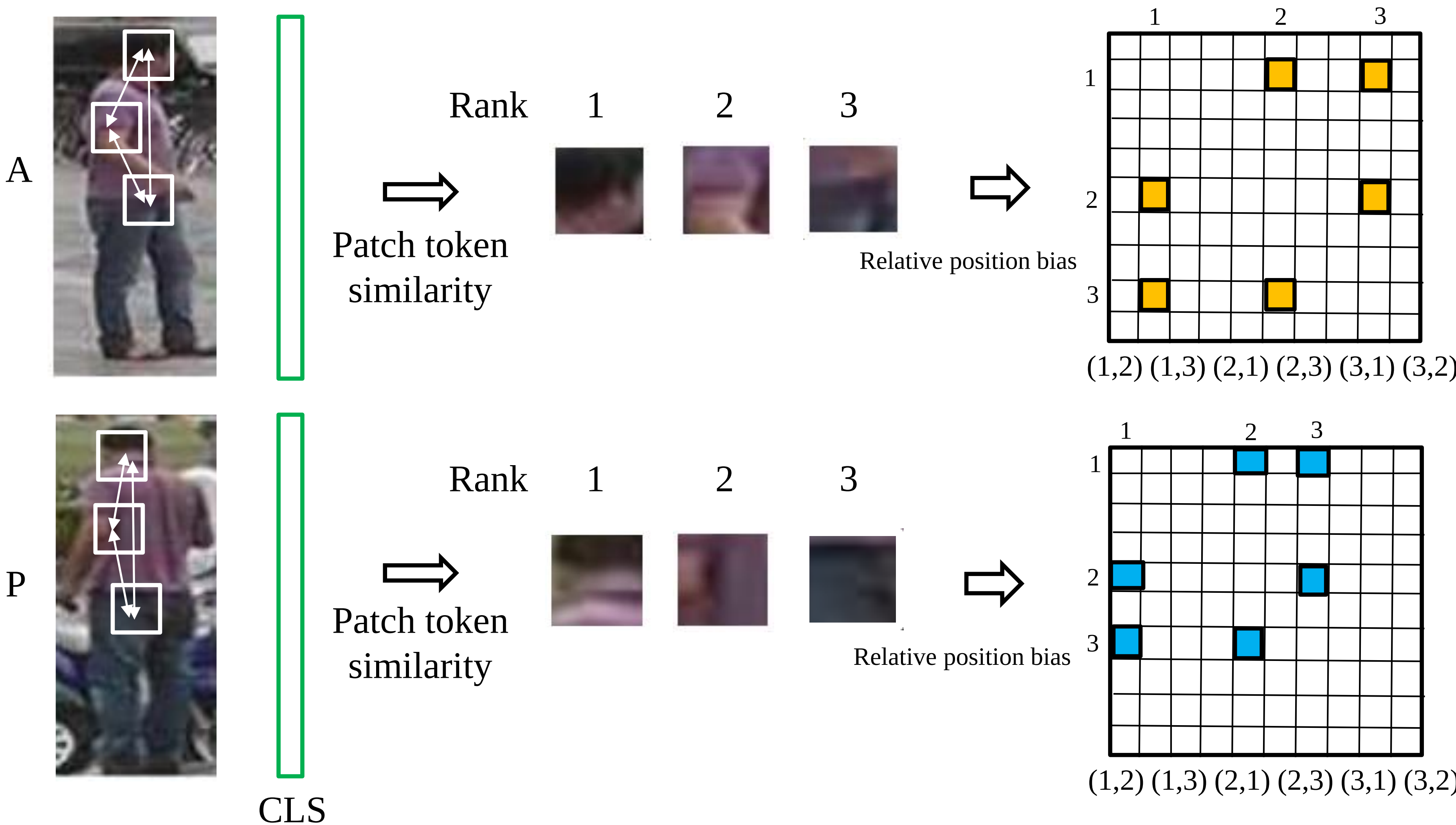
Relative position index

	A	B	C	D
A	0.07	0.5	0.03	0.1
B	0.2	0.07	0.8	0.03
C	0.01	0.4	0.07	0.5
D	0.06	0.01	0.2	0.07

Relative position bias

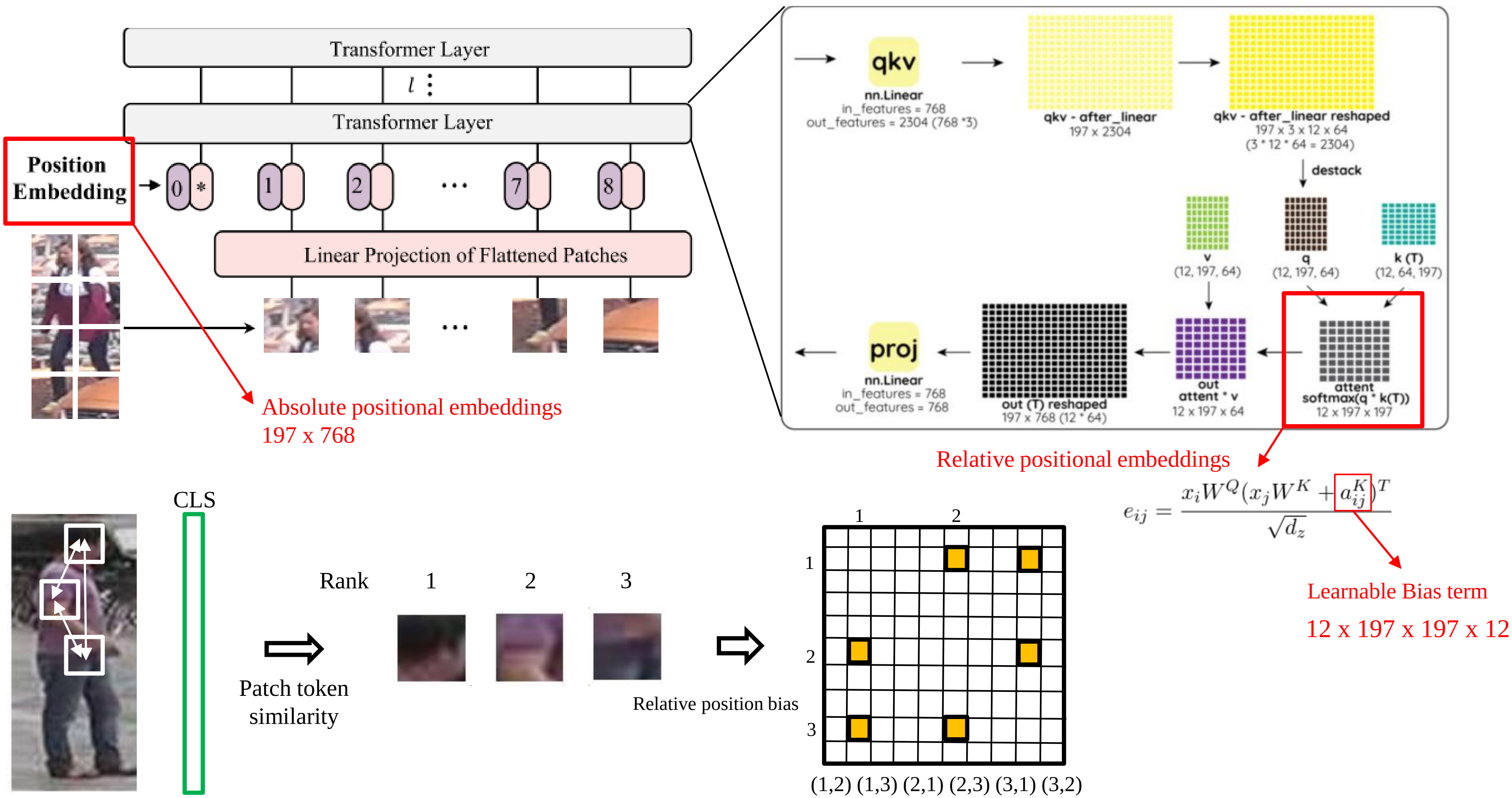
Proposed method

- *Relative position triplet loss*
  - *Idea : "같은 ID의 image끼리는 주요 patch들간 relative position bias가 비슷할 것이다."*



Proposed method

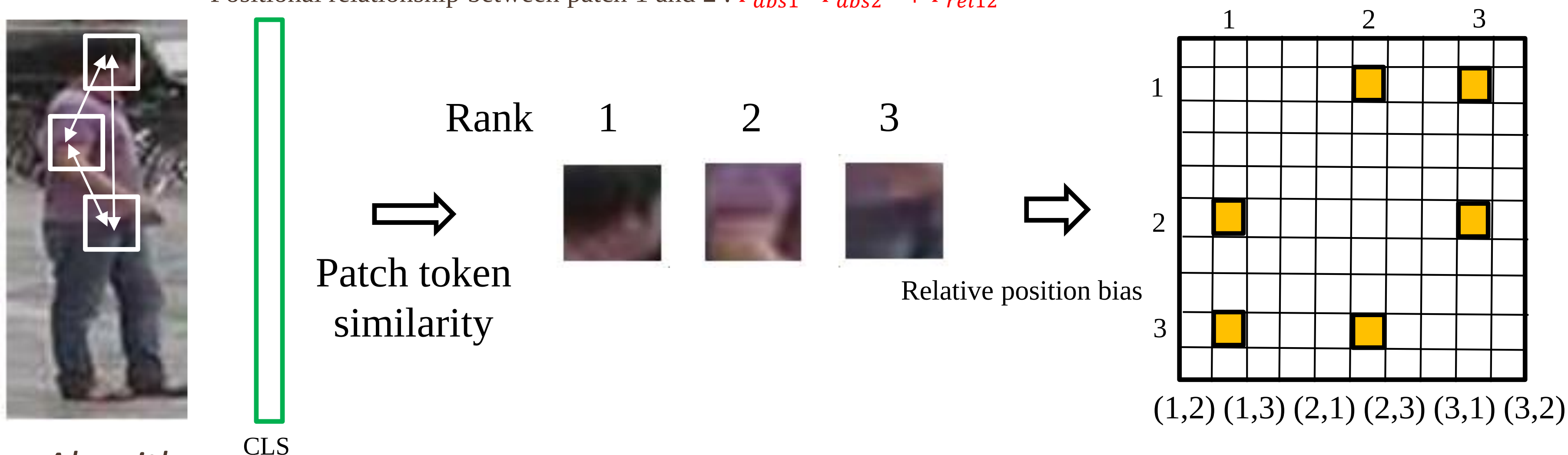
- *Relative position triplet loss*
  - *같은 relative position을 갖는 patch들이 영향을 최소화 하도록 absolute position 을 함께 사용*



김성수

Proposed method

- *Relative position triplet loss*
  - *같은 relative position을 갖는 patch들이 영향을 최소화 하도록 absolute position 을 함께 사용*
    - Positional relationship between patch 1 and 2 : $P_{abs1} \cdot P_{abs2}^{T} + P_{rel12}$



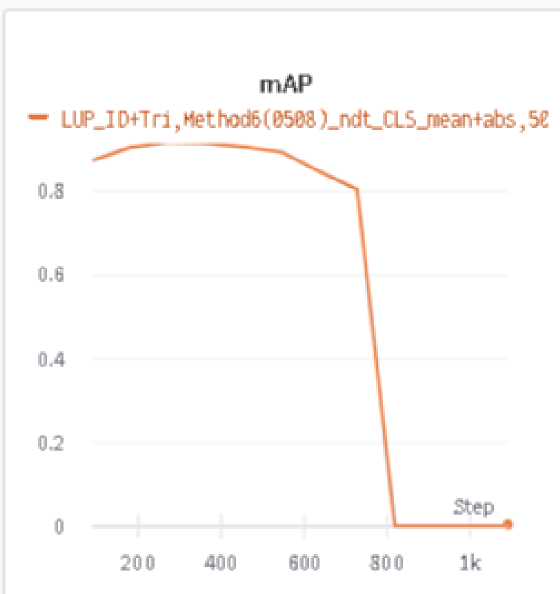
- *Algorithm*

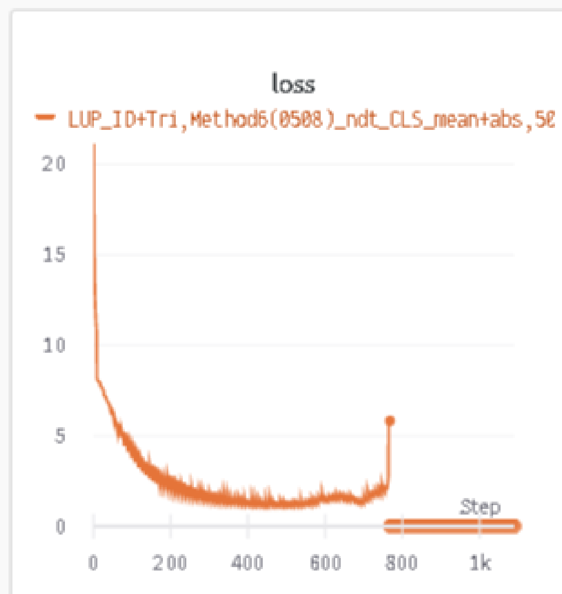
1. *Anchor, Positive, Negative sample 각각에 대해 CLS token과 similarity가 높은 N개의 patch 선별 (sorted)*
2. *N개의 patch에 대해 가능한 combination* ${}_{N}C_2$ 개의 *쌍 (a,b)에 대해 positional relation:* $P_{abs\_a} \cdot P_{abs\_b}^{T} + P_{rel\_ab}$*를 계산하여 원소의 개수가* ${}_{N}C_2$ *개인 patch distance vector (ex: [relation(1,2), relation(1,3), relation(2,3)]) 생성*
3. *Anchor, Positive, Negative의 positional relation vector에 대해 triplet loss를 적용*

김성수

Proposed method(05/09)

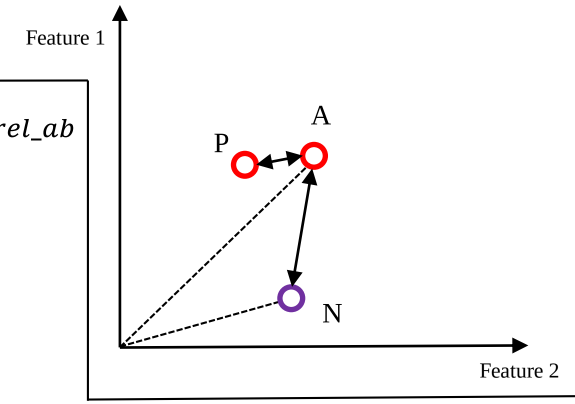
- *Relative position triplet loss*
  - *Algorithm*
  1. *Anchor, Positive, Negative sample 각각에 대해 CLS token과 similarity가 높은 N개의 patch 선별 (sorted)*
  2. *N개의 patch에 대해 가능한 combination ${}_N C_2$ 개의 쌍 (a,b)에 대해 positional relation: $P_{abs_a} \cdot P_{abs_b}{}^T + P_{rel_ab}$를 계산하여 원소의 개수가 ${}_N C_2$ 개인 patch distance vector (ex: [relation(1,2), relation(1,3), relation(2,3)]) 생성*
  3. *Anchor, Positive, Negative의 positional relation vector에 대해 triplet loss를 적용*
- *Experimental result*
  - *초반 학습이 기존 방식보다 꽤 앞서지만, 후반부 학습에서 loss가 수렴하지 못하는 현상이 매번 발생 (Learning rate 문제 X)*
  - *Negative와 Anchor의 positional relation vector간 거리를 크게 만드는 것에서 loss가 발산*





Proposed method(05/12)

- *Triplet loss : Sample의 feature간 Euclidean distance를 통한 학습방식*
- *Positional relation between patch a and b :* $R_{a,b} = P_{abs\_a} \cdot P_{abs\_b}{}^{T} + P_{rel\_ab}$



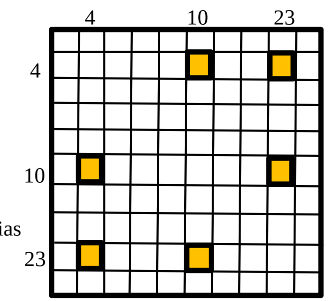
A

CLS

Patch Index 4 10 23

⇨ Patch token similarity

⇨ Relative position bias



Ex)

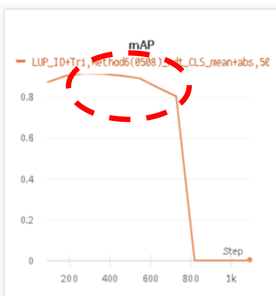
Combination : (4,10) , (4,23) , (10,4) , (10,23) , (23,4) , (23,10)

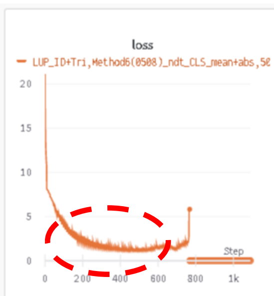
Similarity : $R_{4,10}$, $R_{4,23}$, $R_{10,4}$, $R_{10,23}$, $R_{23,4}$, $R_{23,10}$

A	$R_{4,10}$	$R_{4,23}$	$R_{10,4}$	$R_{10,23}$	$R_{23,4}$	$R_{23,10}$
P						
N						

Triplet loss

적용한 목적
→샘플들의 주요 patch 간 position관계의 유사성

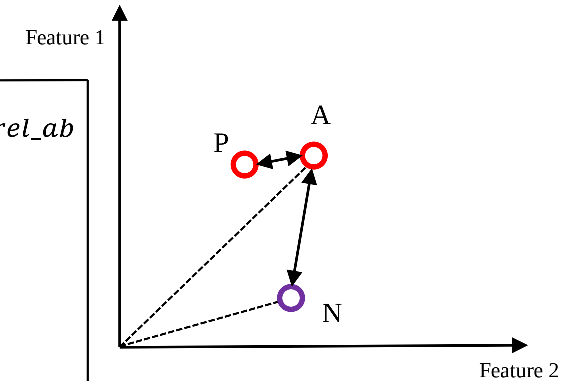




김성수

Proposed method(05/12)

- *Triplet loss : Sample의 feature간 Euclidean distance를 통한 학습방식*
- *Positional relation between patch a and b :* $R_{a,b} = P_{abs\_a} \cdot P_{abs\_b}{}^{T} + P_{rel\_ab}$



Combination : (4,10) , (4,23) , (10,4) , (10,23) , (23,4) , (23,10)
Similarity : $R_{4,10}$, $R_{4,23}$, $R_{10,4}$, $R_{10,23}$, $R_{23,4}$, $R_{23,10}$

A	$R_{4,10}$	$R_{4,23}$	$R_{10,4}$	$R_{10,23}$	$R_{23,4}$	$R_{23,10}$	list

- Feature vector ≠ list of similarity
  - Ex)

	색깔	크기	맛
사과	0.02	0.04	0.01
배	-0.05	0.3	0.2

Feature vector

	노란색과 유사도	초록색과 유사도	파란색과 유사도
사과	0.02	0.4	-0.7
배	0.7	0.3	-0.5

List of similarity

  - Embedding space ≠ Probability

김성수

Proposed method(05/12)

- Embedding space ≠ Probability
  - Ex) $R_{a,b} = P_{abs\_a} \cdot P_{abs\_b}{}^{T} + P_{rel\_ab}$

Combination : (4,10) , (4,23) , (10,4) , (10,23) , (23,4) , (23,10)
Similarity : $R_{4,10}$, $R_{4,23}$, $R_{10,4}$, $R_{10,23}$, $R_{23,4}$, $R_{23,10}$

A	$R_{4,10}$	$R_{4,23}$	$R_{10,4}$	$R_{10,23}$	$R_{23,4}$	$R_{23,10}$	list
P							
N							

⇨ Softmax ⇨ 샘플들의 주요 patch 간position관계를 probability distribution으로 modeling

- Kullback – Leibler divergence (KL divergence)

$$-\sum_{i=1}^{N} p_i \log q_i - \left(-\sum_{i=1}^{N} p_i \log p_i\right) = -\sum_{i=1}^{N} p_i \log\left(\frac{q_i}{p_i}\right)$$

Cross Entropy Entropy

  - Entropy : 어떤 확률 분포가 가지는 평균 정보량
  - Cross Entropy : 실제 확률 분포 P, 모델이 예측한 분포를 Q라고 할때 평균 정보량
  - Ex)
    - A,B,C,D 를 전송하는데 필요한 최소 비트수? 00,01,10,11($log_2 4$ = 2)
    - A ~ Z를 전송하는데 필요한 최소 비트수? 5bit ($log_2 26$)
    - A,B 가 90% 확률로 발생하고, 나머지가 10%의 확률로 발생한다면?
    - 첫번째 bit : A,B인지 아닌지 판단하는 bit
      - A,B라면 추가로 1bit만 전송
      - 아니라면 추가로 5bit($log_2 24$) 전송
      - 0.9*(1+1) + 0.1*(1+5) = 2.4 (최소 3bit)

Proposed method(05/12)

- Kullback – Leibler divergence (KL divergence)
  - Anchor와 Positive 의 position distribution list 간 KL divergence를 loss로 사용 (Negative 고려 X)

$$-\sum_{i=1}^{N} p_i \log q_i - \left(-\sum_{i=1}^{N} p_i \log p_i\right) = -\sum_{i=1}^{N} p_i \log\left(\frac{q_i}{p_i}\right)$$

Cross Entropy　　　Entropy

  - Nonsymmetric : $D_{KL}(P||Q) \neq D_{KL}(Q||P)$
  - *Idea : "같은 ID의 image끼리는 주요 patch들간 relative position bias가 비슷할 것이다."*
    - → D_{KL}(Anchor, Positive) = D_{KL}(Positive, Anchor)

- Jensen – Shannon divergence
  - KL divergence 가 symmetric한 성질을 가지고 있지 않기 때문에 거리 척도로 사용할 수 없어 symmetric한 성질을 가지도록 바꾸어준 식

$$JSD(P,Q) = \frac{1}{2} D(P||M) + \frac{1}{2} D(Q||M)$$

$$where \quad M = \frac{1}{2}(P+Q)$$

$$JSD(P,Q) = JSD(Q,P)$$