김성수

This week

- *TransReID[1]*
- *TransReID with Self-supervised pretraining[2]*
- *Self-supervised learning*
  - *DINO[3]*

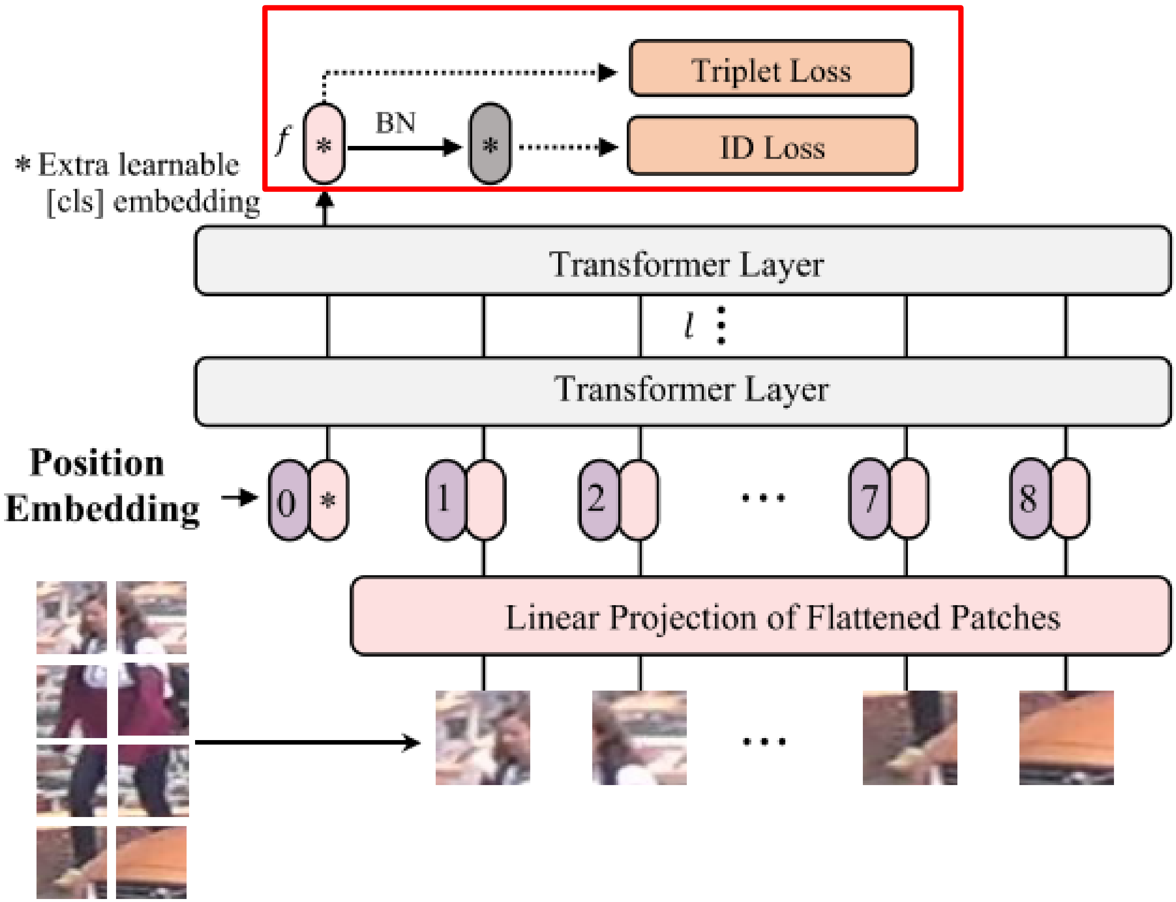
[1]S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15013-15022.*
[2]H. Luo *et al., "Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021*.
[3]M. Caron *et al., "Emerging properties in self-supervised vision transformers," in Proceedings of the IEEE/CVF International Conference on Computer Vision(ICCV), 2021, pp. 9650-9660.*

김성수

This week

[1]S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15013-15022.*

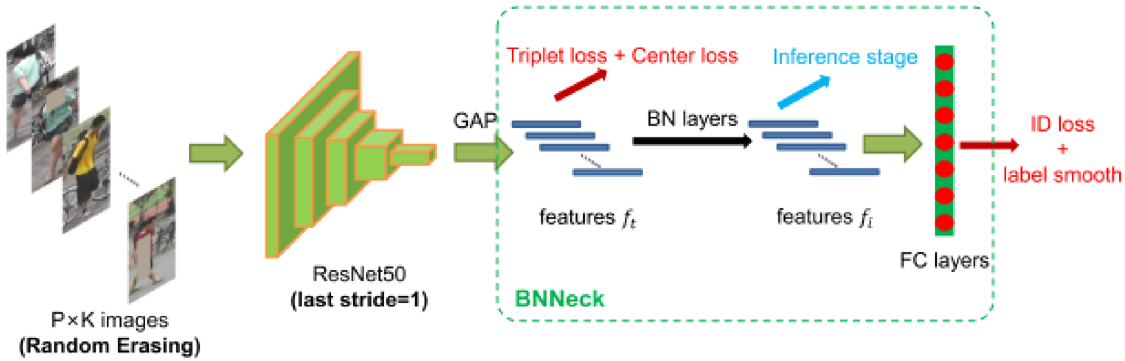
- *TransReID[1]*
  - *ViT 를 backbone으로 사용*
  - *[CLS] token 을 사용하여 feature를 비교*

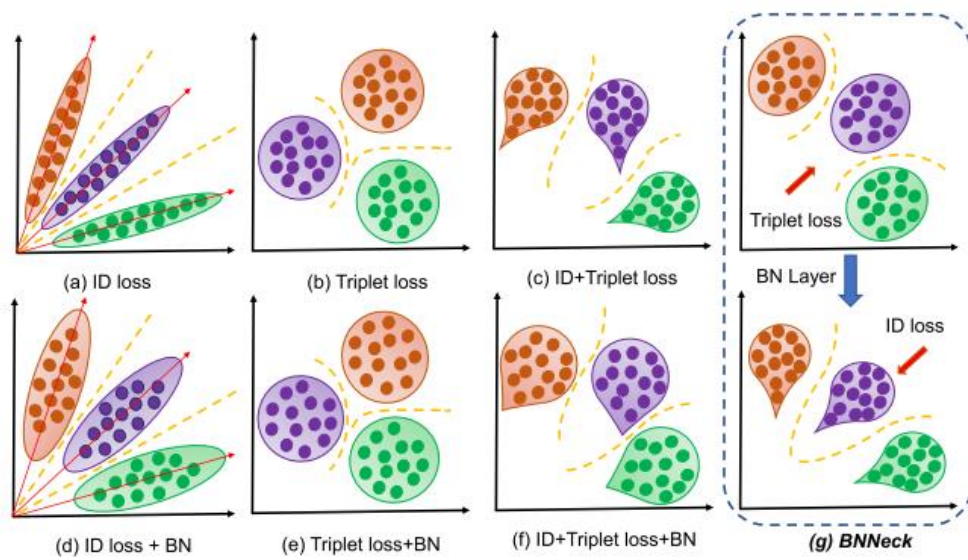


This week

[4]H. Luo *et al.*, *"A strong baseline and batch normalization neck for deep person re-identification," IEEE Transactions on Multimedia, vol. 22, no. 10, pp. 2597-2609, 2019.*

- *TransReID[1]*
  - *BNNeck : ID loss 와 Triplet loss의 embedding space 상에서 의 gap을 줄임*
    - *ID loss : Affine한 decision boundary를 만드는 것이 목적 (Cosine distance)*
    - *Triplet loss : Euclidean distance를 기반으로 한 clustering boundary를 만드는 것이 목적*

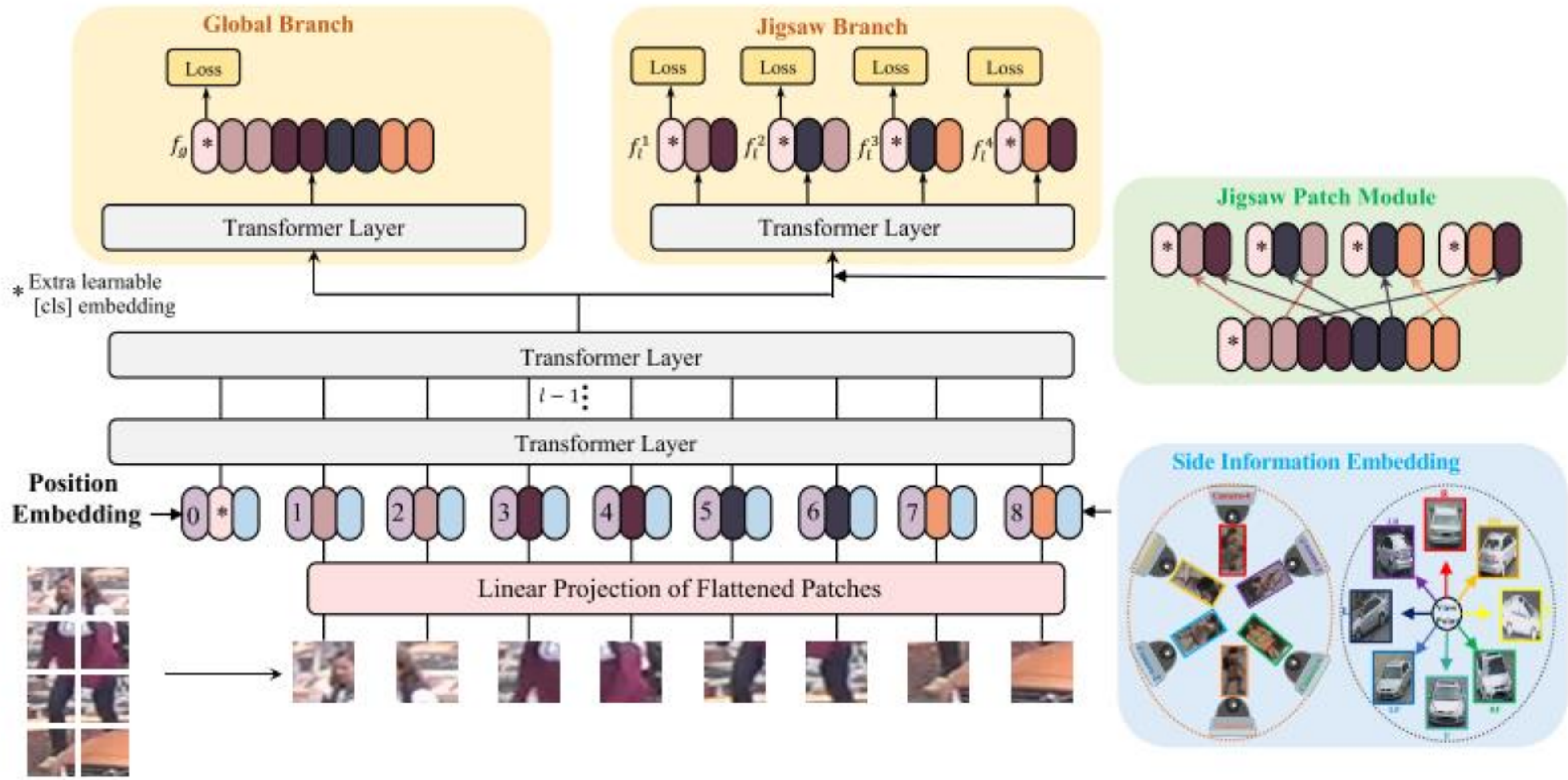




김성수

This week

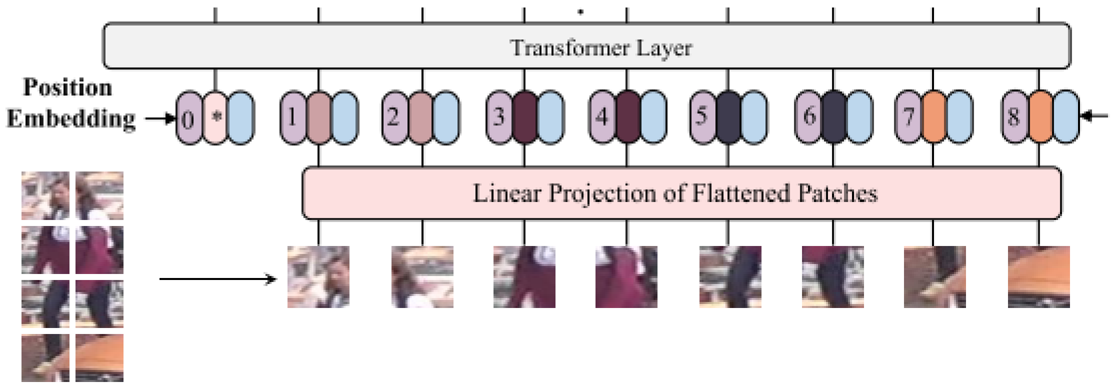
- *TransReID[1]*
  - *SIE (side information embedding), JPM(jigsaw patch module)*

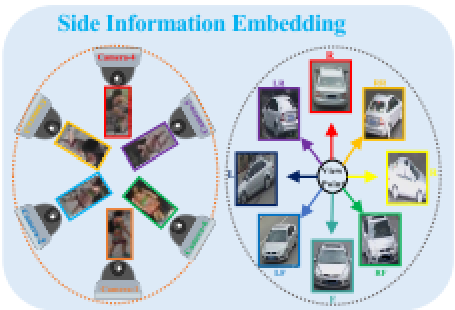


This week

[4]H. Luo *et al.*, "*A strong baseline and batch normalization neck for deep person re-identification,*" *IEEE Transactions on Multimedia, vol. 22, no. 10, pp. 2597-2609, 2019.*

- *TransReID[1]*
  - *SIE (side information embedding)*
    - *Camera, Viewpoint 특성을 model이 학습할 수 있도록 parameterize*

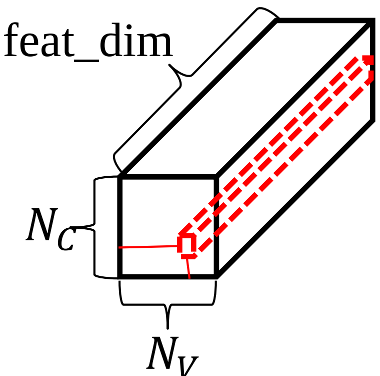




Dataset	Object	#ID	#image	#cam	#view
MSMT17	Person	4,101	126,441	15	-
Market-1501	Person	1,501	32,668	6	-
DukeMTMC-reID	Person	1,404	36,441	8	-
Occluded-Duke	Person	1,404	36,441	8	-
VeRi-776	Vehicle	776	49,357	20	8
VehicleID	Vehicle	26,328	221,567	-	2

Ex: 1500_c3s3_062728_03.jpg (Market1501)

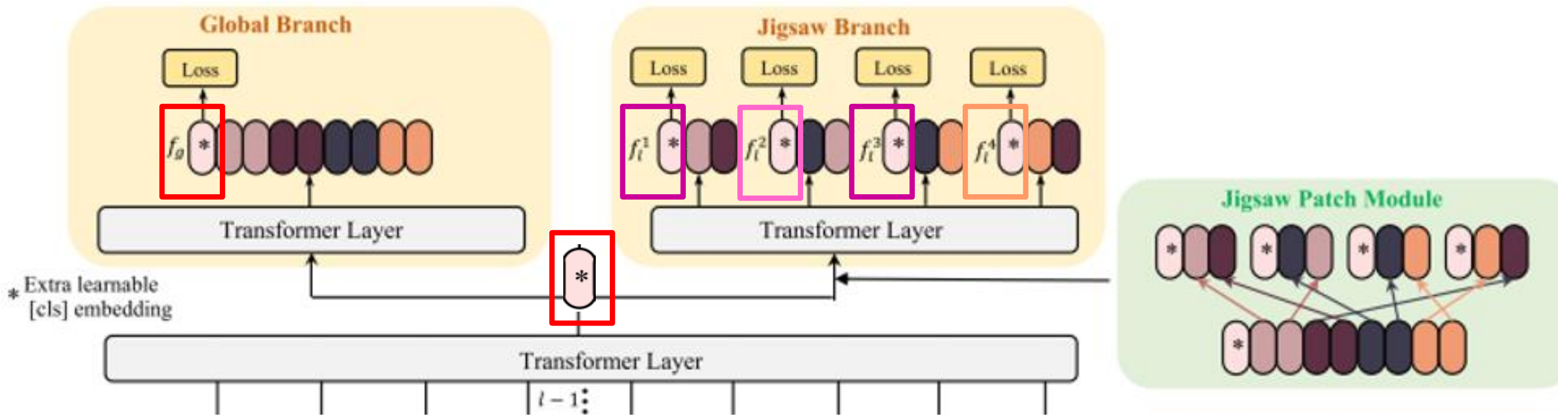
Camera 3번의 Side information embedding 이 patch embedding에 추가됨



김성수

This week

- TransReID[1]
  - JPM (JigSaw Patch module)
    - 일부 patch들의 정보에 집중한 [CLS]token을 통해 local feature를 modeling

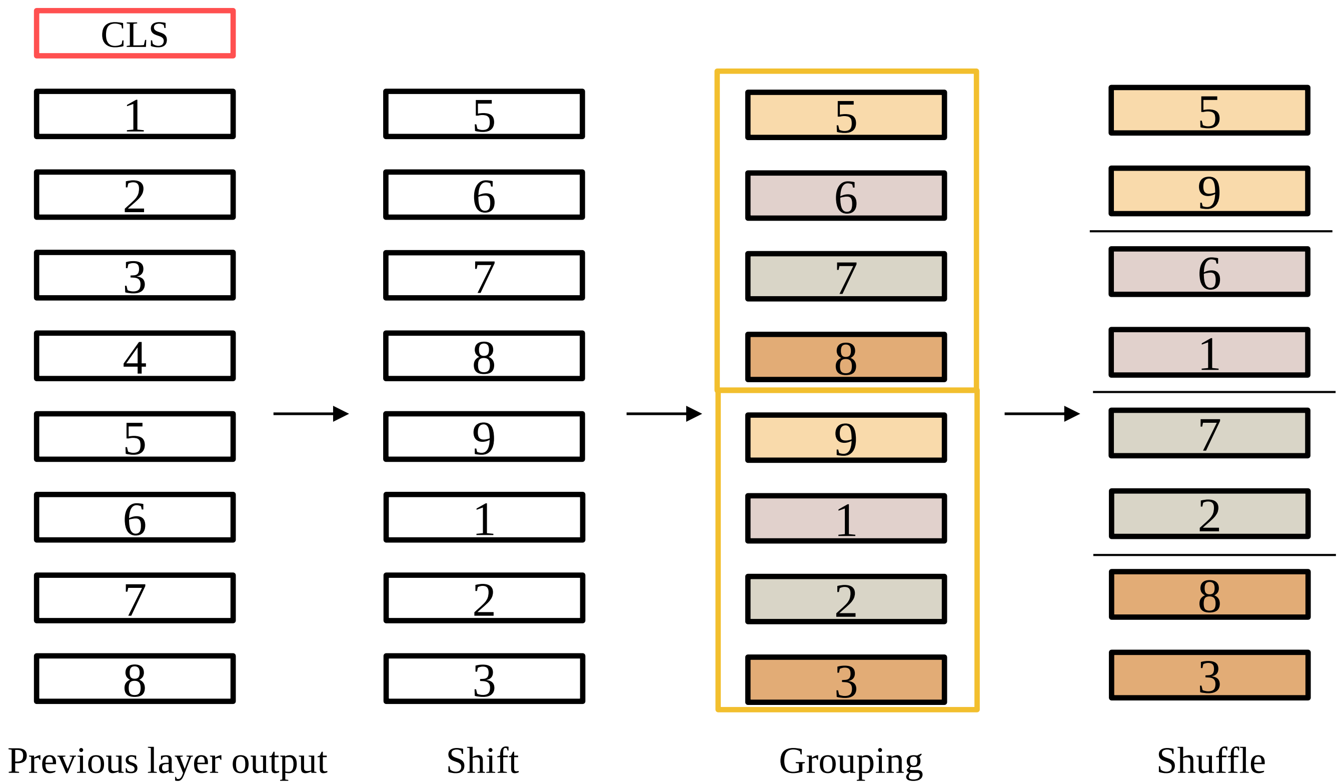


    - Transformer layer가 L개 라고 할때, L-1 layer까지는 동일하게 학습
    - 마지막 transformer layer에서, 별도의 trasnformer layer를 두어 local feature만으로 classification을 진행 (local feature의 개수 만큼 classification layer가 존재)
    - Loss function

$$\mathcal{L} = \mathcal{L}_{ID}(f_g) + \mathcal{L}_T(f_g) + \frac{1}{k}\sum_{j=1}^{k}(\mathcal{L}_{ID}(f_l^j) + \mathcal{L}_T(f_l^j))$$

This week

- *TransReID[1]*
  - *JPM (JigSaw Patch module)*
    - *일부 patch들의 정보에 집중한 [CLS]token을 통해 local feature를 modeling*

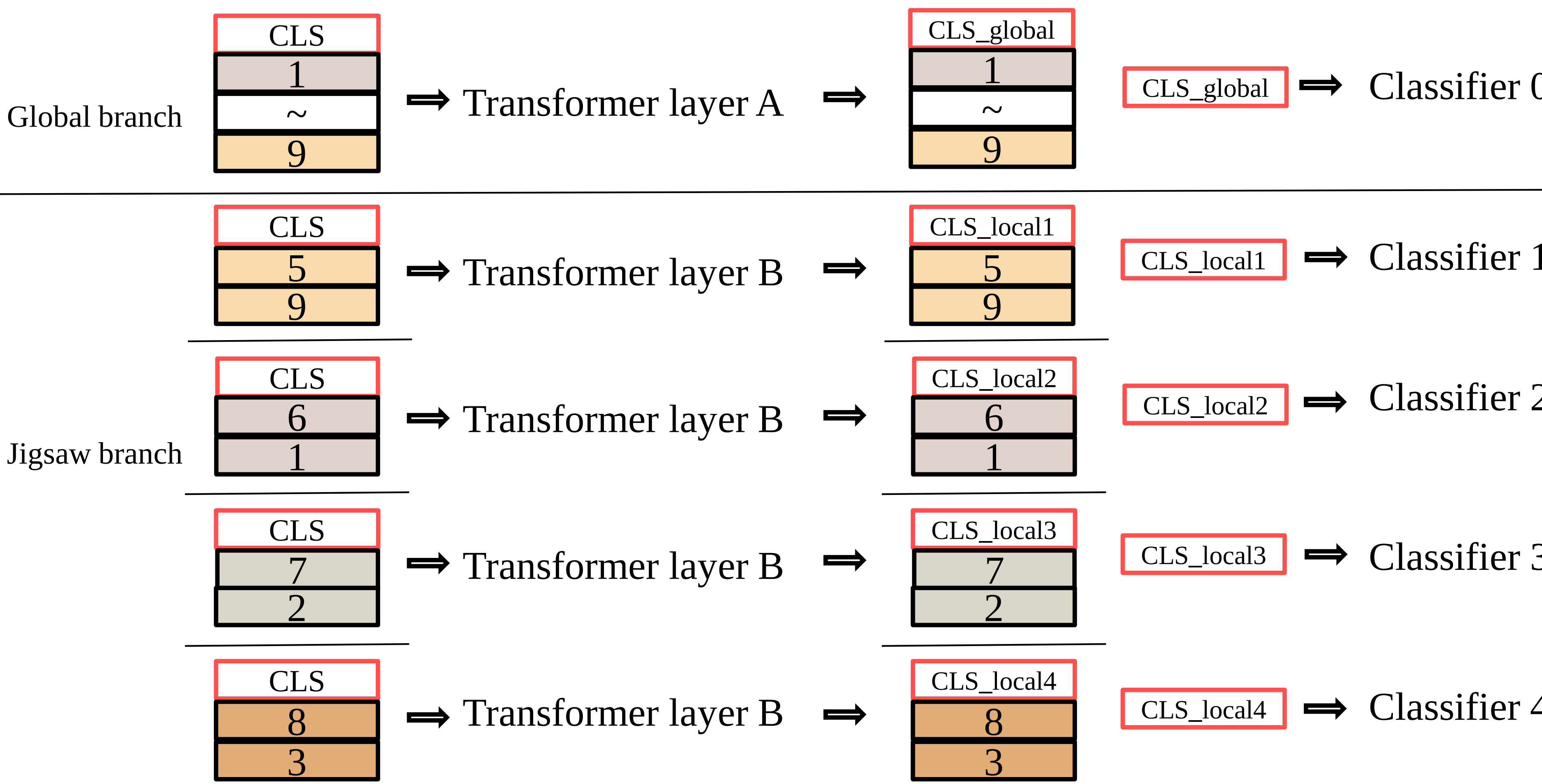


김성수

This week

- *TransReID[1]*
  - *JPM (JigSaw Patch module)*

$$\mathcal{L} = \mathcal{L}_{ID}(f_g) + \mathcal{L}_T(f_g) + \frac{1}{k}\sum_{j=1}^{k}\left(\mathcal{L}_{ID}(f_l^j) + \mathcal{L}_T(f_l^j)\right)$$



This week

- *TransReID[1]*
  - *JPM (JigSaw Patch module)*
    - *Ex) 16 * 8 patches, shift 5, group 4*

김성수

This week

- *TransReID[1]*
  - *Result*

Dataset	Object	#ID	#image	#cam	#view
MSMT17	Person	4,101	126,441	15	-
Market-1501	Person	1,501	32,668	6	-
DukeMTMC-reID	Person	1,404	36,441	8	-
Occluded-Duke	Person	1,404	36,441	8	-
VeRi-776	Vehicle	776	49,357	20	8
VehicleID	Vehicle	26,328	221,567	-	2

Method	JPM	SIE	MSMT17		VeRi-776	
			mAP	R1	mAP	R1
Baseline	✗	✗	61.0	81.8	78.2	96.5
	✓	✗	63.6	82.5	79.2	96.8
	✗	✓	62.4	81.9	79.6	96.9
TransReID	✓	✓	**64.9**	**83.3**	**80.6**	**96.9**

    - *Market-1501 test result*

JPM	SIE	mAP	Rank 1	Rank 5	Rank 10
X	X	87.1%	94.2%	98.2%	99.1%
O	X	87.4%	95.0%	98.2%	99.1%
X	O	87.9%	95.2%	98.6%	99.2%
O	O	88.2%	94.9%	98.5%	99.1%

This week

[2]H. Luo *et al.*, *"Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021*.

- *TransReID with Self-supervised pretraining[2]*
  - *ReID backbone은 imagenet을 통해 학습된 pretrained weight를 사용*
  - *ImageNet dataset과 ReID dataset은 domain gap이 심함*
    - *추가로 ImageNet을 통해 train된 weight는 category-level supervision에 초점이 맞추어져 있기 때문에 fine-grained identity information을 얻어야 하는 ReID task에 필요한 weight와는 성질이 다름*
  - *LUPerson dataset (Large-scale Unlabeled Person Re-ID dataset)*
    - *4M images , 200K ID, 46K scenes*
  - *LUPerson dataset을 사용한 self-supervised learning을 통해, transformer-based Re-ID model의 pretrained weight를 얻고자 하는것이 목적*
    - *Re-ID task의 pre-trained weight로 사용되기 가장 적합한 weight를 만들수 있는 SSL 방식은 무엇인가? -> DINO [3]*
    - *LUPerson dataset과 target dataset(fine-tuning) 간의 gap을 어떻게 줄일 수 있는가? (Target dataset에 더 잘 맞는 pretrained weight를 만들기 위해)*

→ *Conditional Transfer Learning*

This week

[2]H. Luo *et al.*, *"Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021*.

- *TransReID with Self-supervised pretraining[2]*
  - *Re-ID task의 pre-trained weight로 사용되기 가장 적합한 weight를 만들수 있는 SSL 방식은 무엇인가? -> DINO [3]*
    - *Supervised fine-tuning result*

Models	Pre-training Methods	Pre-training Data	Market mAP	Market R1	MSMT17 mAP	MSMT17 R1
R50	Supervised	IMG	86.7	94.8	52.2	76.0
	MoCoV2	LUP	88.2	94.8	53.3	76.0
	MoCoV3	LUP	87.3	95.1	52.9	76.8
	DINO	LUP	86.5	94.4	51.9	75.8
ViT-S/16	Supervised	IMG	85.0	93.8	53.5	75.2
	MoCoV2	IMG	63.6	72.1	19.6	36.1
	MoCoV3	IMG	81.7	92.1	46.6	70.3
	MoBY	IMG	83.3	92.2	49.1	71.5
	DINO	IMG	84.6	93.1	54.8	76.7
ViT-S/16	MoCoV2	LUP	72.1	87.6	27.8	47.4
	MoCoV3	LUP	82.2	92.1	47.4	70.3
	MoBY	LUP	84.0	92.9	50.0	73.2
	DINO	LUP	**90.3**	**95.4**	**64.2**	**83.4**
	DINO	LUP*	89.6	95.1	62.3	82.6

김성수

This week

[2]H. Luo *et al.*, *"Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021.*

- *TransReID with Self-supervised pretraining[2]*
  - *LUPerson dataset과 target dataset(fine-tuning) 간의 gap을 어떻게 줄일 수 있는가? (Target dataset에 더 잘 맞는 pretrained weight를 만들기 위해) → Conditional Transfer Learning*
    - *LUPerson dataset은 unlabeled dataset 이기 때문에 data의 선별이 필요(Ex: Low resolution, cropped image)*
    - *Target dataset과 특징이 비슷한 data를 선별한 후(first stage), 이들 만을 이용하여 model을 pre-training 시키는 (second stage) 방식을 사용*

Algorithm 1 Our Proposed Conditional Filtering

procedure FILTER($\mathcal{D}_s, \mathcal{D}_t$)
- $\theta_s \leftarrow TRAIN(\mathcal{D}_s)$ ▷ Source Proxy Model
- $\theta_t \leftarrow TRAIN(\theta_s, \mathcal{D}_t)$ ▷ Target Proxy Model
- **for** $i \leftarrow 1$ to N **do** ▷ $x_s^i \in \mathcal{D}_s$
  - $c_s^i \leftarrow CFS(x_s^i)$ ▷ Compute CFS
- $c_s \leftarrow SORT(c_s^1, c_s^2, ..., c_s^N)$ ▷ Get Score Set
- $\mathcal{D}_s' \leftarrow TOP(\mathcal{D}_s, N', c_s)$ ▷ Filter Source Dataset
- **return** $\mathcal{D}_s'$ ▷ Return the Filtered Subset

CFS : catastrophic forgetting score

$$c_s^i = \frac{\langle \theta_s(x_s^i), \theta_t(x_s^i) \rangle}{||\theta_s(x_s^i)||\ ||\theta_t(x_s^i)||}$$

김성수

This week

[2]H. Luo *et al., "Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021*.

TransReID with Self-supervised pretraining[2]

- *LUPerson dataset과 target dataset(fine-tuning) 간의 gap을 어떻게 줄일 수 있는가? (Target dataset에 더 잘 맞는 pretrained weight를 만들기 위해) → Conditional Transfer Learning*
  - *LUPerson dataset은 unlabeled dataset 이기 때문에 data의 선별이 필요(Ex: Low resolution, cropped image)*
  - *Target dataset과 특징이 비슷한 data를 선별한 후(first stage), 이들 만을 이용하여 model을 pre-training 시키는 (second stage) 방식을 사용*

(a) Sampled data with high CFS

(b) Filtered data with low CFS

(c) Data in the target domain

Source : LUPerson
Target : Market1501
CFS : catastrophic forgetting score

$$c_s^i = \frac{\langle \theta_s(x_s^i), \theta_t(x_s^i) \rangle}{||\theta_s(x_s^i)|| \, ||\theta_t(x_s^i)||}$$

김성수

This week

[2]H. Luo *et al.*, *"Self-Supervised Pre-Training for Transformer-Based Person Re-Identification," arXiv preprint arXiv:2111.12084, 2021*.

- *TransReID with Self-supervised pretraining[2]*
  - *Result*

	JPM	SIE	mAP	Rank 1	Rank 5	Rank 10
TransReID (IMG) 120 Epoch	X	X	87.1%	94.2%	98.2%	99.1%
	O	X	87.4%	95.0%	98.2%	99.1%
	X	O	87.9%	95.2%	98.6%	99.2%
	O	O	88.2%	94.9%	98.5%	99.1%
TransReID -SSL(LUP) 30 Epoch	X	X	90.8%	95.2%	98.2%	99.0%
TransReID -SSL(LUP) 120 Epoch	X	X	93.1%	96.6%	98.9%	99.2%