Boosting Contrastive Self-Supervised Learning with False Negative Cancellation

[WACV, 2022]

2023. 05. 04.

김성수

Data Mining and Quality Analytics

Data Mining Quality Analytics

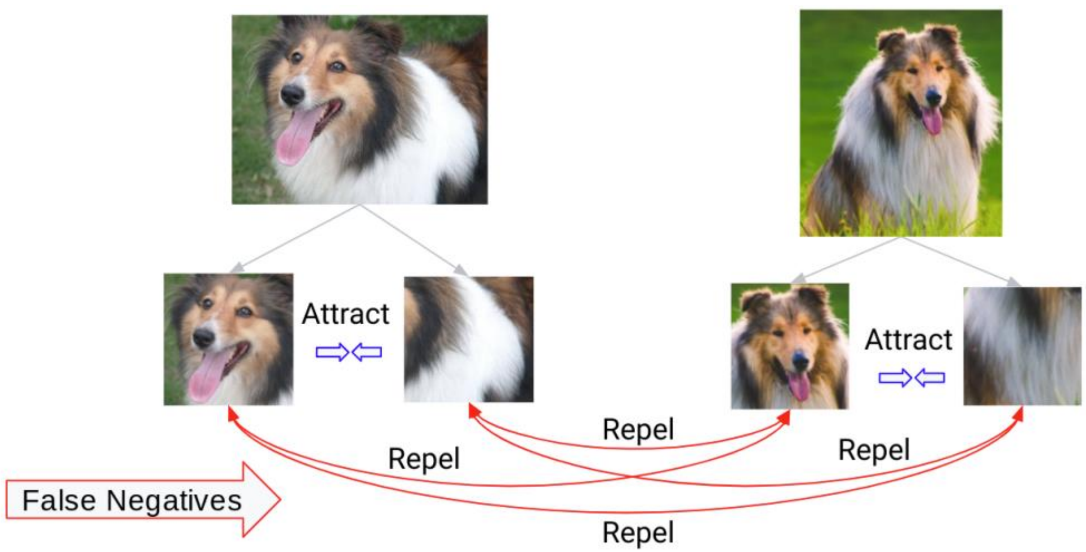
고려대학교 KOREA UNIVERSITY

연구배경

- Task명

❖ Task: Contrastive Learning

- Contrastive Learning: 유사한 데이터는 가깝게, 상이한 데이터는 멀어지도록 학습
- False Negative Cancellation: Contrastive Learning의 False Negative를 식별하고, 학습에 활용하는 방법 제안

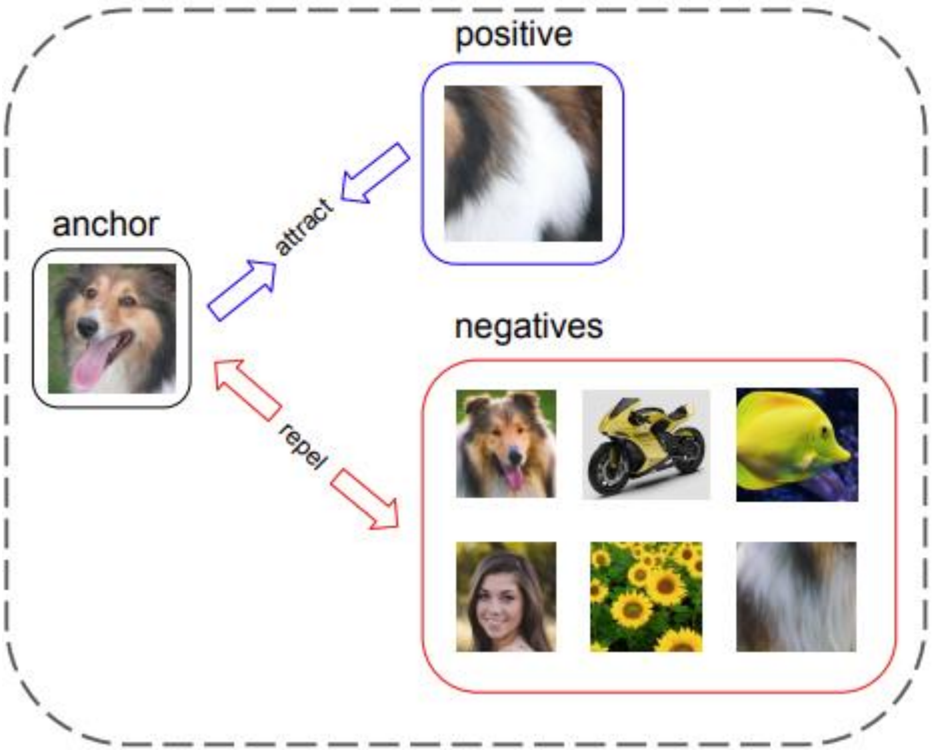


연구배경

- 선행연구의 한계

❖ Limitation of Previous Research

- Contrastive Learning 시, Negative Sample들이 무엇인지에 상관 없이 모두 멀어지도록 학습
  - ✓ 이러한 Negative Sample들 중 Anchor와 유사한 Sample들은 학습 방해 및 수렴속도 저하
- 최근 연구들은 Hard Negative Mining에만 집중하고, False Negative에 대한 연구는 저조



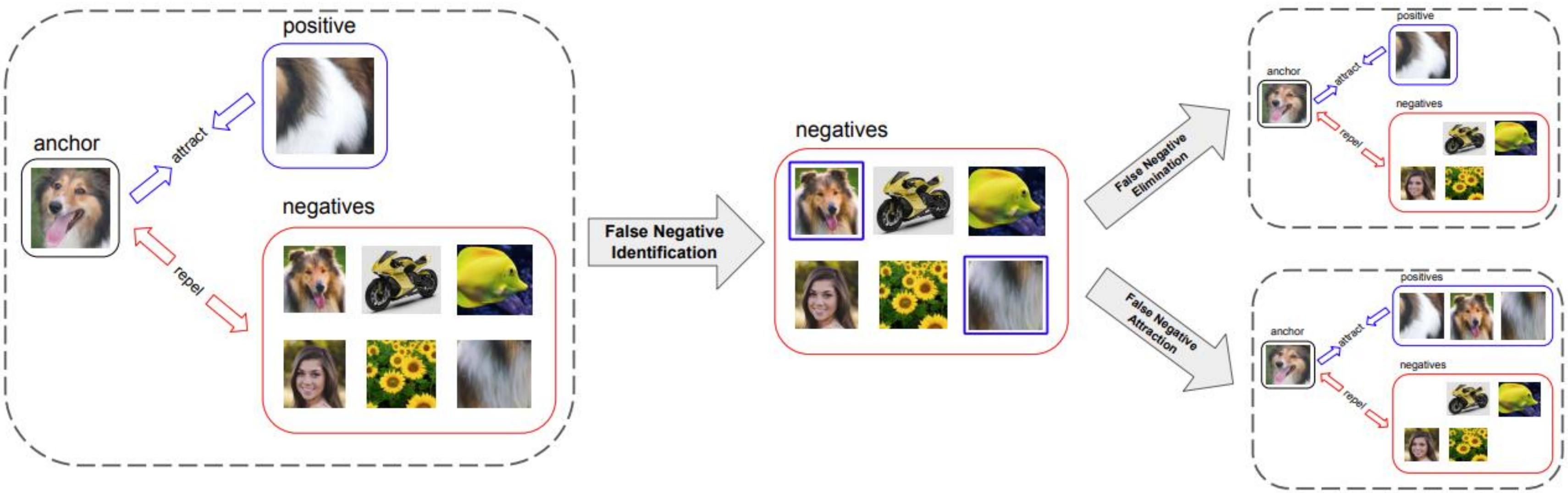
Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

연구배경

- 선행연구의 한계를 극복과정 (Overview of Research)

❖ Overcome the Limitation

- Step1. False Negative Identification
  - ✓ Embedding Vector의 유사도를 활용
- Step2. False Negative Elimination OR False Negative Attraction

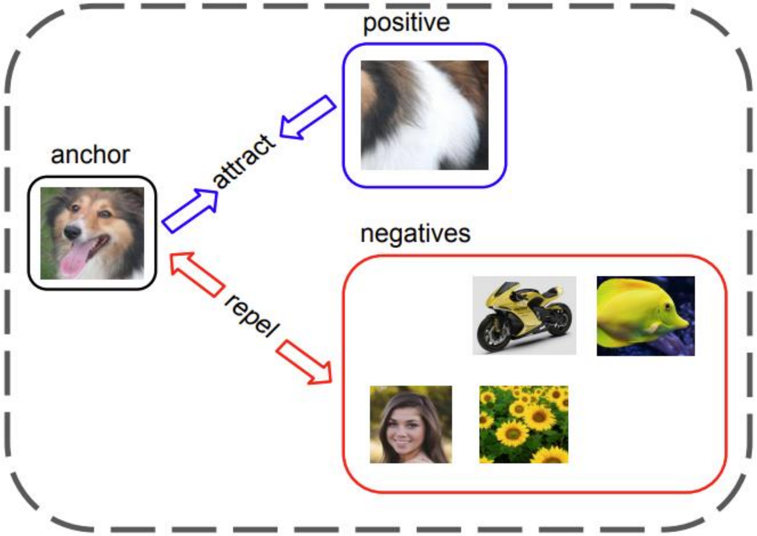


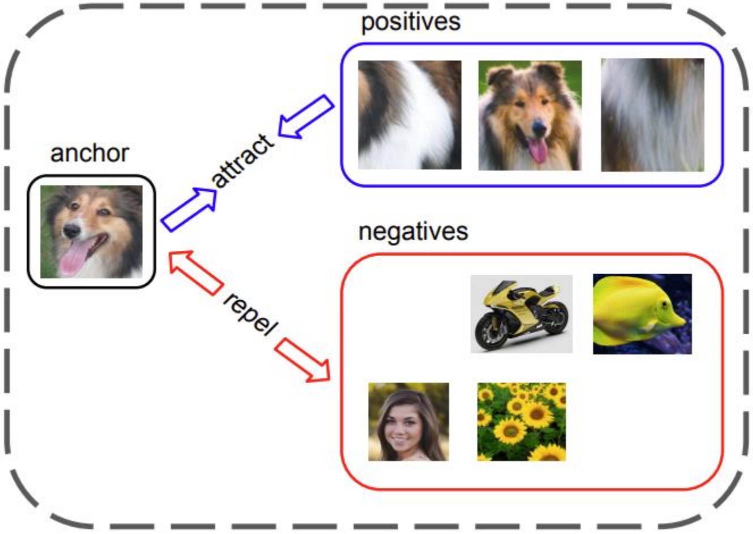
연구배경

- Contribution

❖ Contribution

- Contrastive Learning에서 False Negative를 식별하는 방법론을 최초로 제안
- False Negative를 Contrastive Learning에서 활용하는 방안 제안
  - ✓ Elimination, Attraction



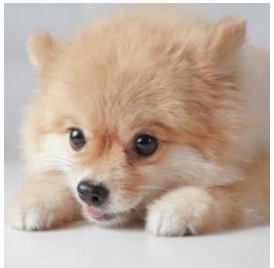


방법론

- Ideation

❖ 방법론 아이디어

- False Negative를 찾는 것은 Label 없이는 매우 어려움
  - ✓ Contrastive Learning은 비지도학습이기에, 레이블을 갖지 않기 때문
- Representation Vector간 유사도로 False Negative를 판별할 수는 없을까?


Label ? ? ?

Embedding Vector



방법론

- 세부내용

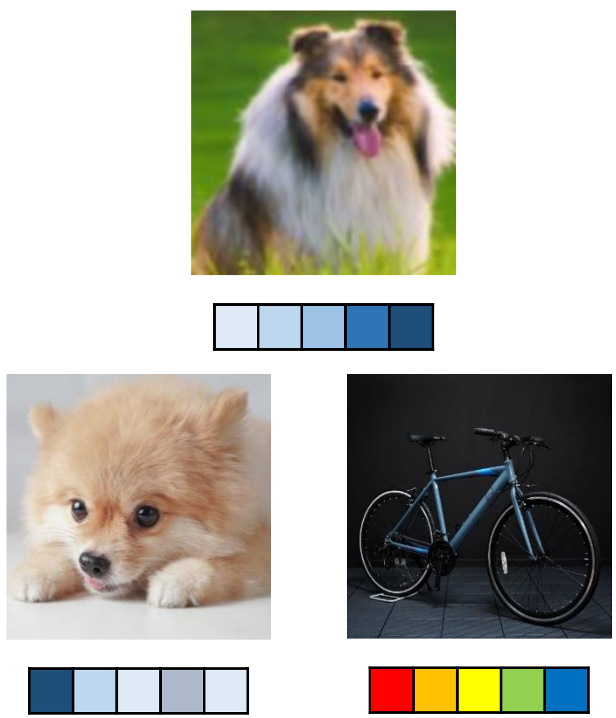
❖ Details① False Negative Identification

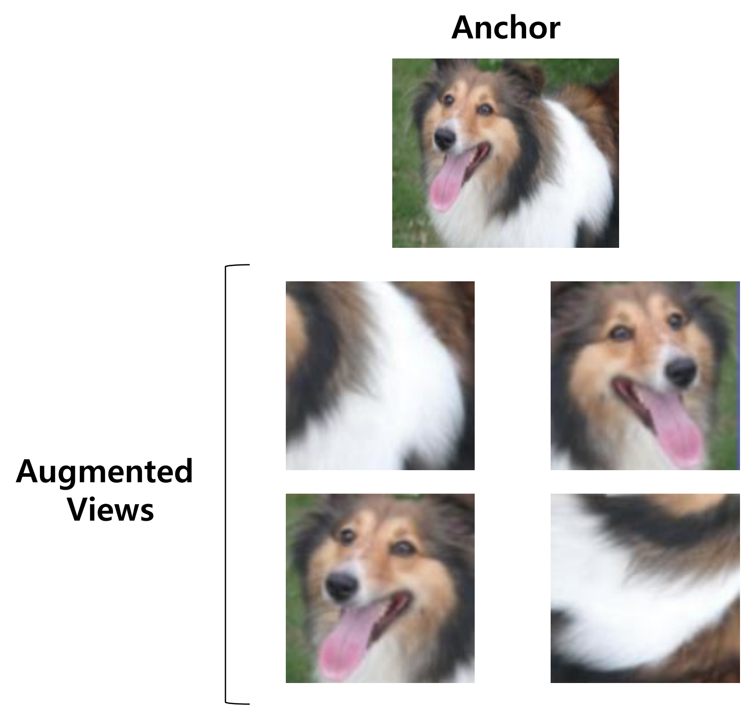
- 가정

  ✓ False Negative의 Embedding Vector는 Anchor의 Embedding Vector와 유사성을 가질 것이다.

  ✓ False Negative는 같은 의미를 포함한 이미지지만, 데이터 증강 시 유사함이 깨질 수 있다.

  * 데이터 증강이 특정 View만 포함할 수 있기 때문



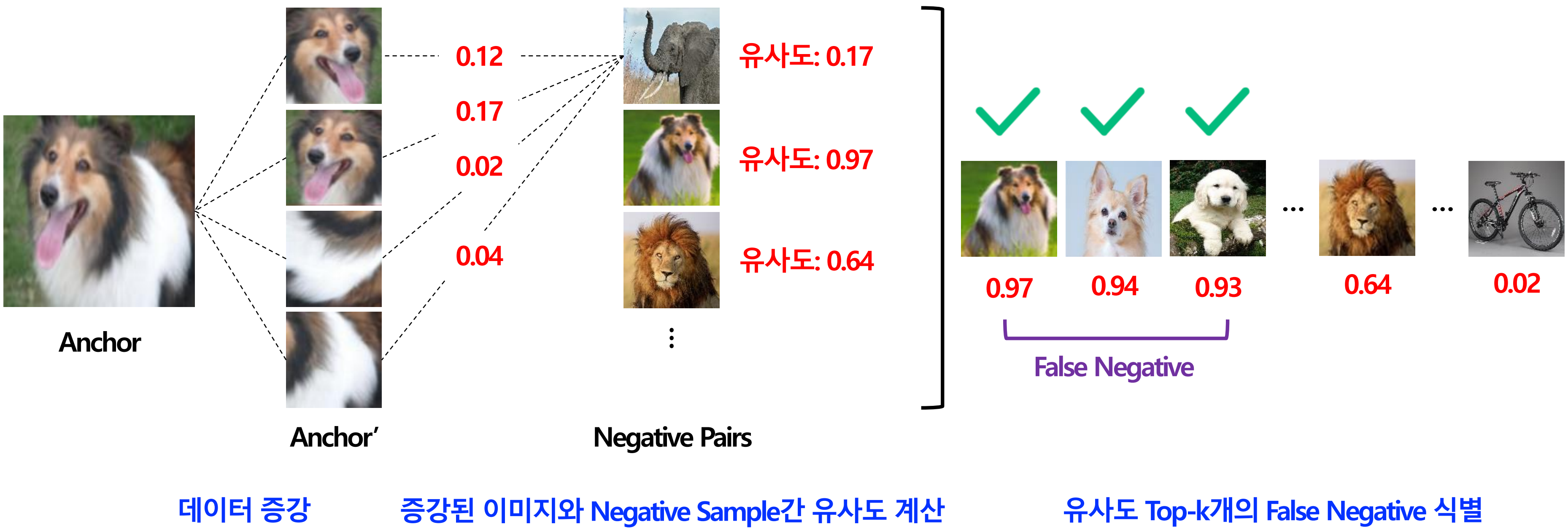


방법론

- 세부내용

❖ Details① False Negative Identification

- Step1. Anchor를 여러 번(N번) 증강한 이미지와 하나의 Negative Sample간 유사도 연산
- Step2. N개 유사도 중 가장 높은 유사도(또는 N개 평균)를 해당 Negative Sample의 Score로 산정
- Step3. Top-k개(또는 특정 Threshold 초과)의 유사도를 갖는 Sample들을 False Negative로 판단



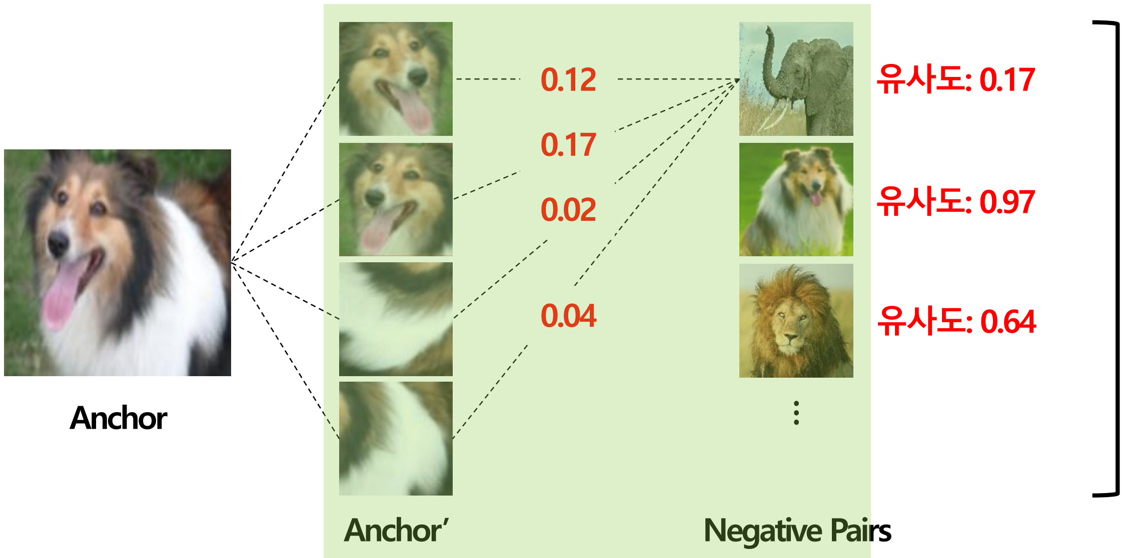
Data Mining Quality Analytics
고려대학교 KOREA UNIVERSITY

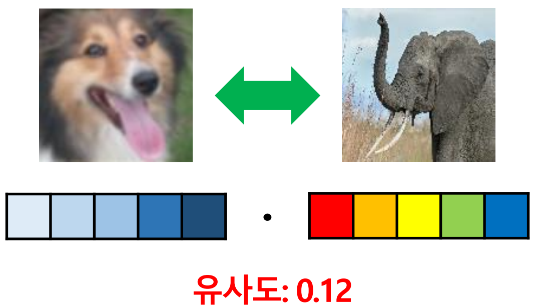
방법론

- 세부내용

❖ Details① False Negative Identification

- Step1. Anchor를 여러 번(N번) 증강한 이미지와 하나의 Negative Sample간 유사도 연산
  - ✓ Cosine 유사도 (내적) 활용





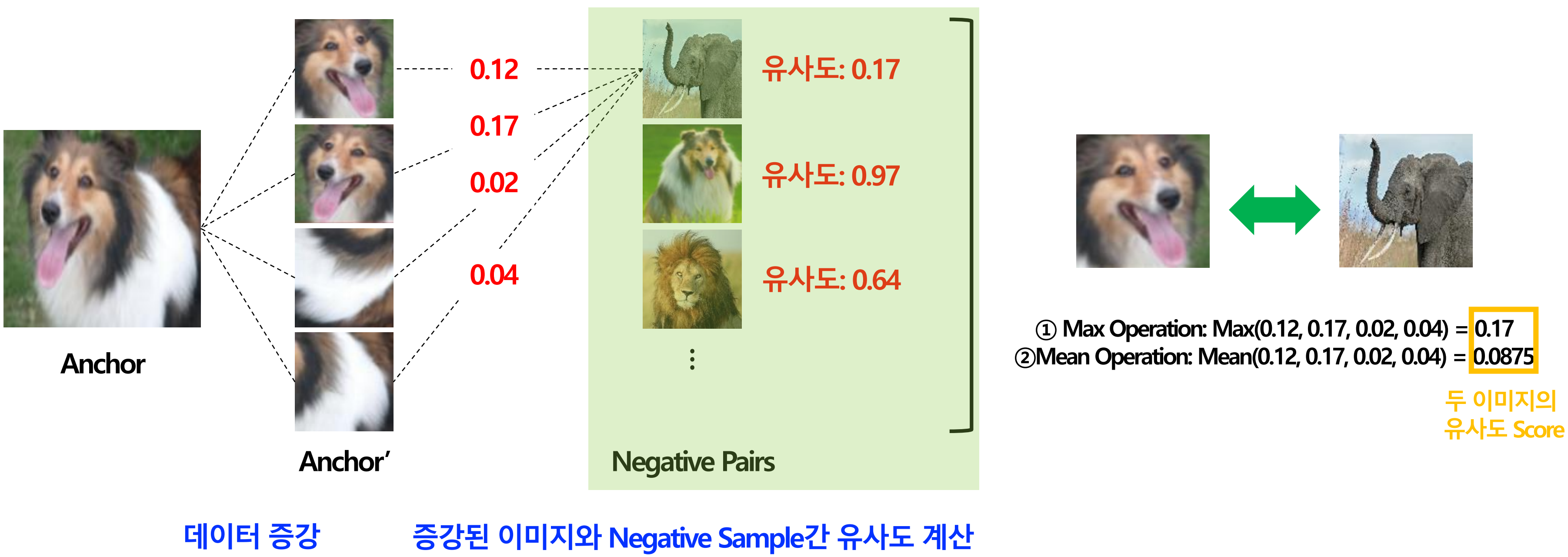
데이터 증강 증강된 이미지와 Negative Sample간 유사도 계산

방법론

- 세부내용

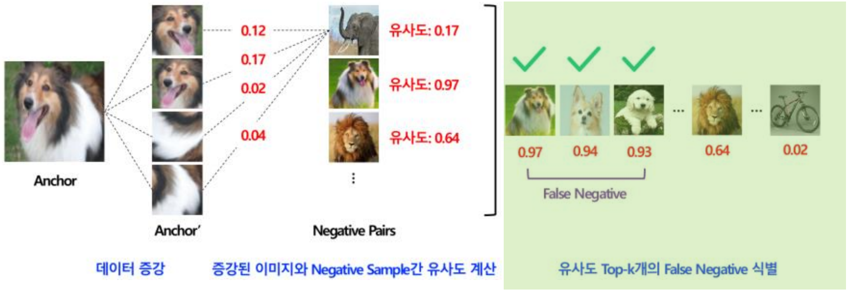
❖ Details① False Negative Identification

- Step2. N개 유사도 중 가장 높은 유사도(또는 N개 평균)를 해당 Negative Sample의 Score로 산정
  - ✓ Max Operation | Mean Operation



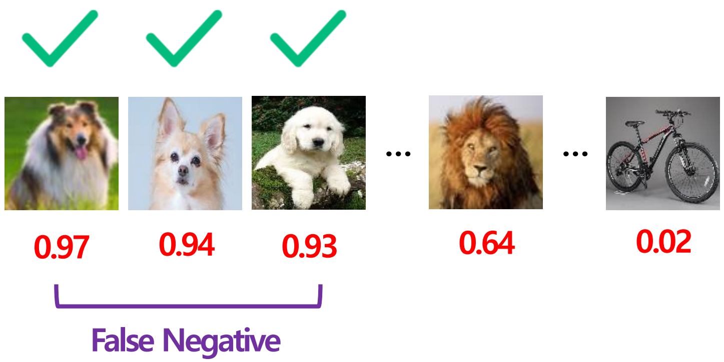
방법론

- 세부내용

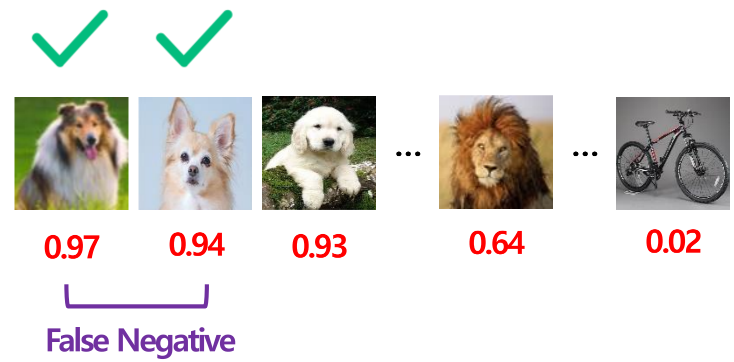


❖ Details① False Negative Identification

- Step3. Top-k개(또는 특정 Threshold 초과)의 유사도를 갖는 Sample들을 False Negative로 판단
  - ✓ Tok-k | Thresholding | Both



유사도 Top-3개의 False Negative 식별



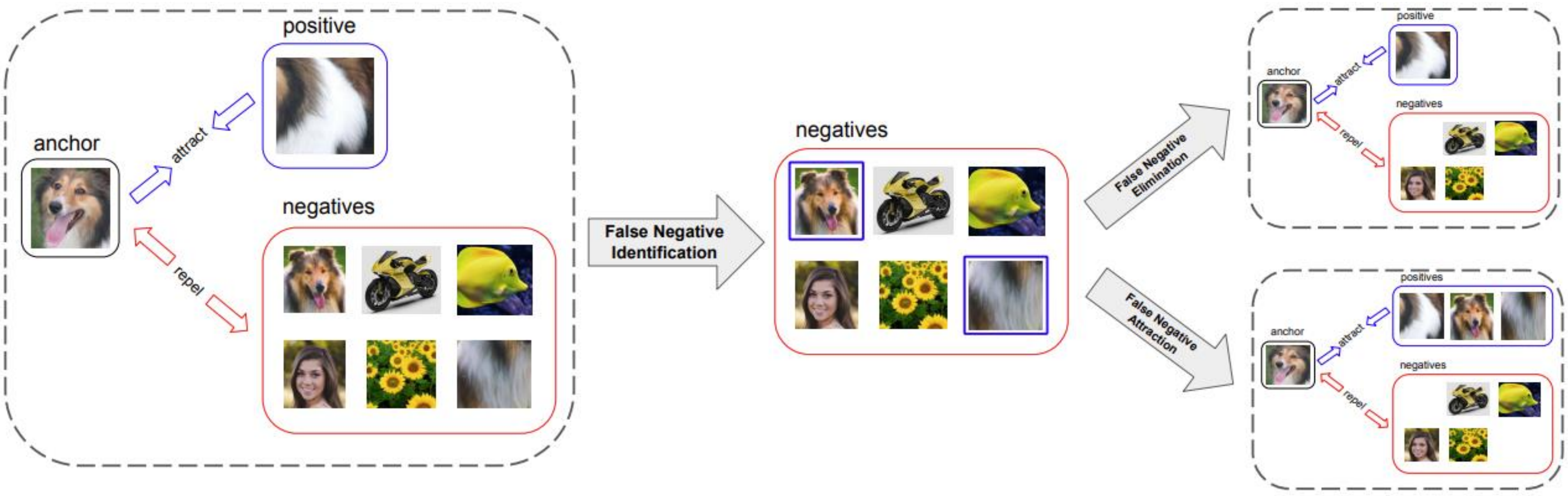
Threshold >= 0.94의 False Negative 식별

방법론

- 세부내용

❖ Details② False Negative & Contrastive Learning

- False Negative Elimination: False Negative들을 Negative로 취급하지 않음
- False Negative Attraction: False Negative들을 Positive로 취급하여 가까워지도록 학습
  - ✓ 특이점: 기존 Contrastive Loss 식에 더해주는 방식으로 연산 (단독 Loss가 아님)



방법론

- 세부내용

❖ Details② False Negative & Contrastive Learning

- False Negative Elimination: False Negative들을 Negative로 취급하지 않음
- False Negative Attraction: False Negative들을 Positive로 취급하여 가까워지도록 학습
  - ✓ 특이점: 기존 Contrastive Loss 식에 더해주는 방식으로 연산 (단독 Loss가 아님)

False Negative Elimination

$$l_i^{\text{elim}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i, k \notin \underline{\mathbb{F}_i}]} \exp(\text{sim}(z_i, z_k)/\tau)},$$

False Negative

False Negative Attraction

$$l_i^{\text{att}} = -\frac{1}{1+|\mathbb{F}_i|} \left(\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} + \sum_{f \in \mathbb{F}_i} \log \frac{\exp(\text{sim}(z_i, z_f)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \right)$$

실험결과

- Result

❖ 실험결과

- SimCLRv2에 False Negative Attraction을 적용했을 때, SOTA급 성능을 보임
  - ✓ Top-10 / Threshold: 0.7 / Augmentation: 8회 / Batch Size: 4,096 / Max Operation

Method	top-1	top-5
Supervised	76.5	
Representation Learning		
Contrastive learning		
MoCo v1 [24]	60.6	—
PIRL [35]	63.6	—
PCL [33]	65.9	—
SimCLR v1 [9]	69.3	89.0
MoCo v2 [11]	71.1	—
SimCLR v2 [10]	71.7	90.4
InfoMin [44]	73.0	91.1
FNC (ours)	***74.4***	**91.8**
Others		
BYOL [23]	74.3	91.6
SwAV [7]	**75.3**	—

Table 5. ImageNet linear evaluation.

Method	1% top-1	1% top-5	10% top-1	10% top-5
Supervised	25.4	56.4	48.4	80.4
Semi-supervised				
UDA [51]	—	68.8	—	88.5
FixMatch [42]	—	71.5	—	89.1
Representation Learning				
Contrastive learning				
PIRL [35]	30.7	60.4	57.2	83.8
PCL [33]	—	—	75.6	86.2
SimCLR v1 [9]	48.3	75.5	65.6	87.8
SimCLR v2 [10]	57.9	82.5	68.4	89.2
FNC (ours)	**63.7**	**85.3**	**71.1**	**90.2**
Others				
BYOL [23]	53.2	78.4	68.8	89.0
SwAV [7]	53.9	78.5	70.2	89.9

Table 6. ImageNet semi-supervised evaluation.

❖ 실험결과

- 사전학습에 활용하지 않은 데이터로 미세조정 시에도 준수한 성능을 보임
- Object Detection에서도 좋은 성능을 보임

	Food	CIFAR10	CIFAR100	Birdsnap	SUN397	Cars	Aircraft	VOC2007	DTD	Pets	Caltech-101	Flowers	Avg
Linear eval													
SimCLR v1 [9]	68.4	90.6	71.6	37.4	58.8	50.3	50.3	80.5	74.5	83.6	90.3	91.2	70.6
SimCLR v2 [10]	73.9	92.4	76.0	44.7	61.0	54.9	51.1	81.2	**76.5**	85.0	91.2	93.5	73.4
BYOL [23]	**75.3**	91.3	**78.4**	**57.2**	62.2	67.8	60.6	82.5	75.5	**90.4**	**94.2**	**96.1**	**77.6**
FNC (ours)	74.4	**93.0**	76.8	54.0	**63.2**	**68.8**	**61.3**	**83.0**	76.3	89.0	93.5	94.9	77.3
Finetuned													
SimCLR v1 [9]	88.2	97.7	85.9	75.9	63.5	91.3	88.1	84.1	73.2	89.2	92.1	97.0	85.5
SimCLR v2 [10]	88.2	97.5	86.0	74.9	**64.6**	91.8	87.6	84.1	74.7	89.9	92.3	97.2	85.7
BYOL [23]	**88.5**	**97.8**	86.1	**76.3**	63.7	91.6	88.1	**85.4**	**76.2**	**91.7**	**93.8**	97.0	86.3
FNC (ours)	88.3	97.7	**86.8**	**76.3**	64.2	**92.0**	**88.5**	84.7	76.0	90.9	93.6	**97.5**	**86.4**

Table 7. Transfer learning on classification task using ImageNet-pretrained ResNet models across 12 data sets.

Method	AP50
Supervised	81.3
MoCo v2 [11]	82.5
SwAV [7]	82.6
FNC (ours)	**82.8**

Table 8. Transfer learning on Pascal VOC object detection.

실험결과

- Result

❖ 실험결과

- FNC에서 제안하는 False Negative Identification이 Elimination과 Attraction에서 모두 효과적
  - ✓ Attraction에서 효과가 더 두드러짐
  - ✓ Attraction은 False Negative를 잘 선정할 수만 있다면, 성능 증가 폭이 큼
  - ✓ 현재 모델은 약 40%정도의 False Negative를 Detection 성공

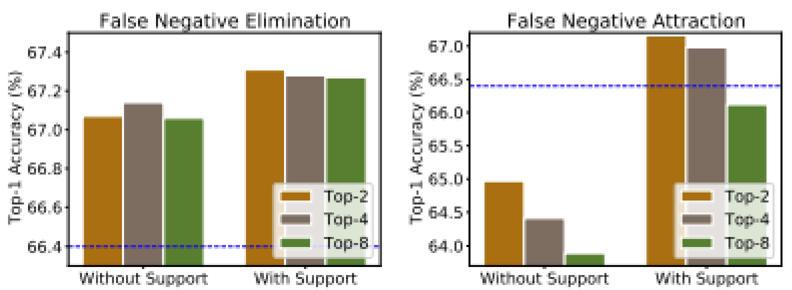


Figure 5. False negative cancellation with and without support set across top-k choices for different mitigation strategies. The dashed line denotes the performance of the SimCLR baseline. The results use mean aggregation in scoring potential false negatives.

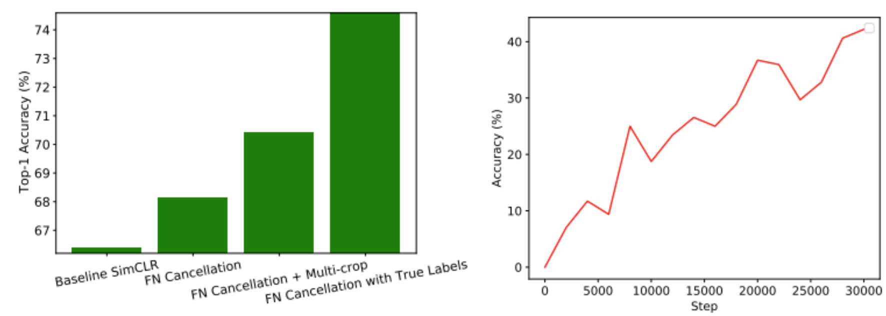


Figure 10. A visualization of (left) top-1 accuracy with false negative cancellation using detected vs. ground-truth labels and (right) the accuracy of false negative detection.

Data Mining Quality Analytics
고려대학교 KOREA UNIVERSITY

실험결과

- Result

❖ 실험결과

- FNC에서 제안하는 False Negative Identification이 Elimination과 Attraction에서 모두 효과적
  - ✓ Attraction에서 효과가 더 두드러짐
  - ✓ Attraction은 False Negative를 잘 선정할 수만 있다면, 성능 증가 폭이 큼
  - ✓ 현재 모델은 약 40%정도의 False Negative를 Detection 성공

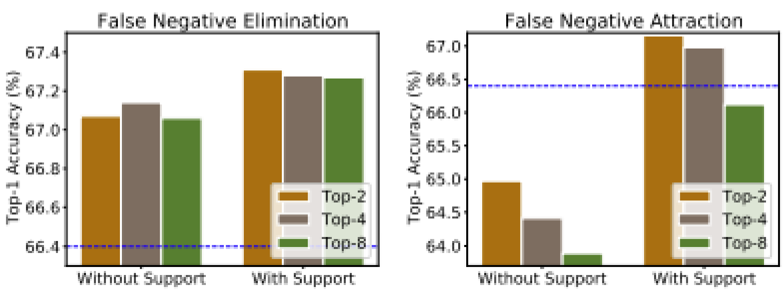


Figure 5. False negative cancellation with and without support set across top-k choices for different mitigation strategies. The dashed line denotes the performance of the SimCLR baseline. The results use mean aggregation in scoring potential false negatives.

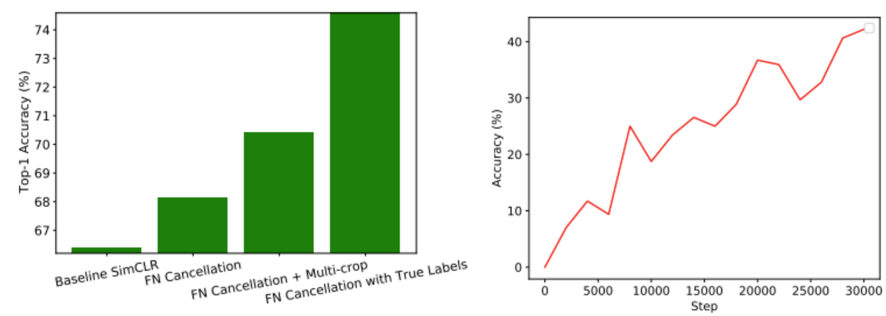
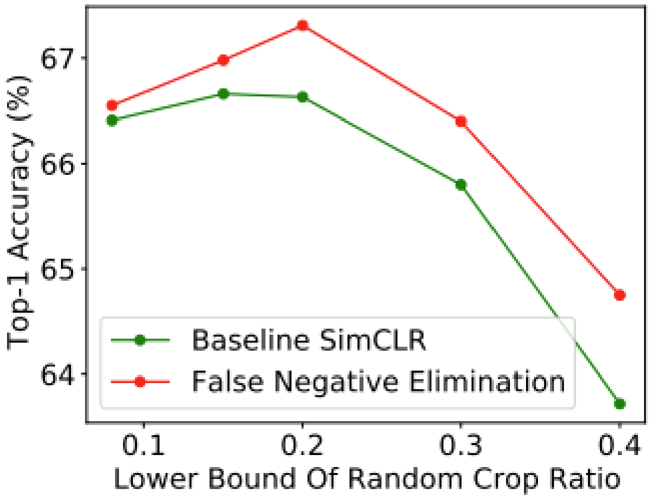


Figure 10. A visualization of (left) top-1 accuracy with false negative cancellation using detected vs. ground-truth labels and (right) the accuracy of false negative detection.

❖ 실험결과

- 데이터 증강 시 Random Crop에서 크게 Crop할수록 좋은 성능을 보임
  - ✓ 활용하는 Augmentation: Random Crop, Color Distortion, Gaussian Blur



실험결과

- Result

❖ 실험결과

- Attraction에서 Max Score가 Mean Score보다 좋은 성능을 보임
  - ✓ Elimination에서는 미미함
- Both > Top-k > Thresholding

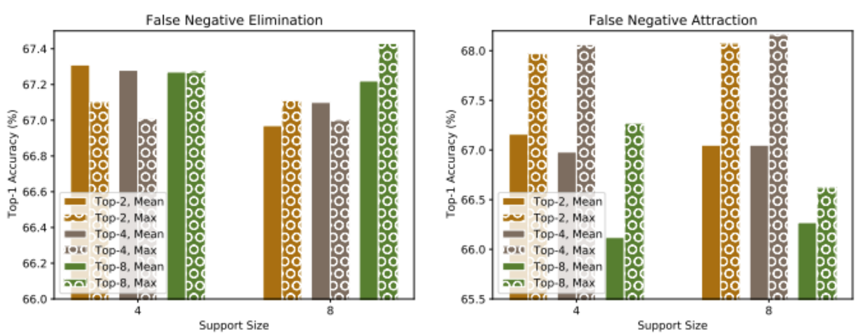


Figure 6. False negative cancellation with mean and max aggregation across support sizes and top-k for the false negative (left) elimination and (right) attraction strategies.

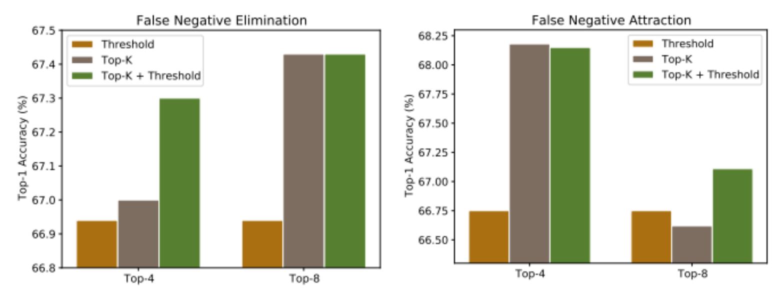


Figure 7. A comparison of top-k and threshold-based filtering for false negative (left) elimination and (right) attraction strategies.

실험결과

- Result

❖ 실험결과

- 성능도 증가하지만, 학습시간도 3배정도 증가

Model	Epochs	Time (h)	Acc. (%)
SimCLR	100	2.63	66.41
SimCLR	1000	26.22	70.34
Improved Model	100	7.50	70.42

Table 4. Computational efficiency and accuracy.