Supervised Contrastive Learning

Yoon Sang Cho

Korea University

DMQA Open Seminar

2022.1.7

발표자 소개

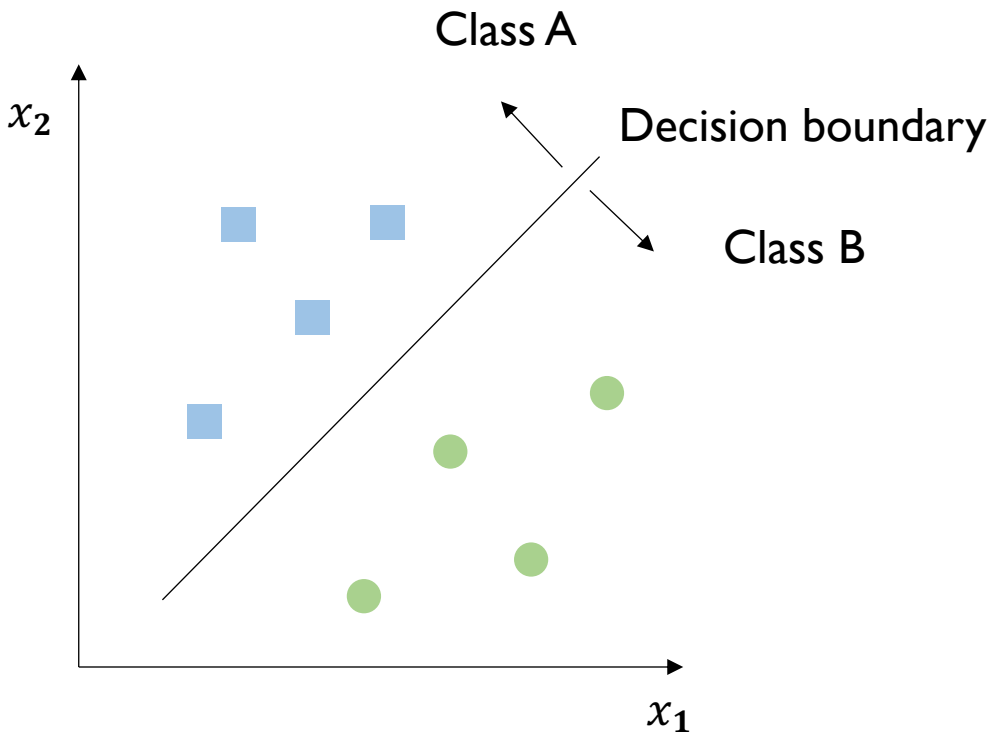
- **조윤상 (Yoon Sang Cho)**
  - 고려대학교 산업경영공학과 석·박통합과정 (2017 - 현재)
  - 한국외국어대학교 산업경영공학과 학사 (2017)
- **연구 분야**
  - Representation Learning for Multivariate Time-Series Data
  - Predictive & Explainable Modeling for Manufacturing Systems
  - Adversarial Attack and Defense Methods in AI modeling
- **이메일**
  - yscho187@korea.ac.kr

Contents

- Representation Learning
- Self-Supervised Contrastive Learning
- Supervised Contrastive Learning
- Summary

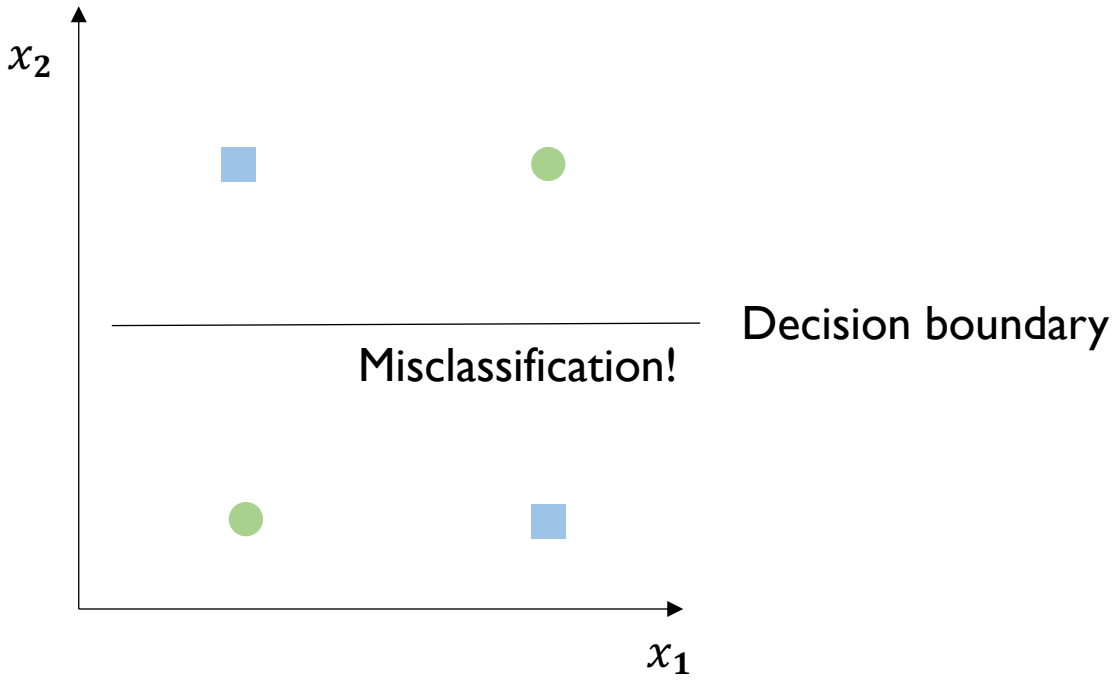
Representation Learning

- Machine learning algorithms for classification
  - linear classifiers



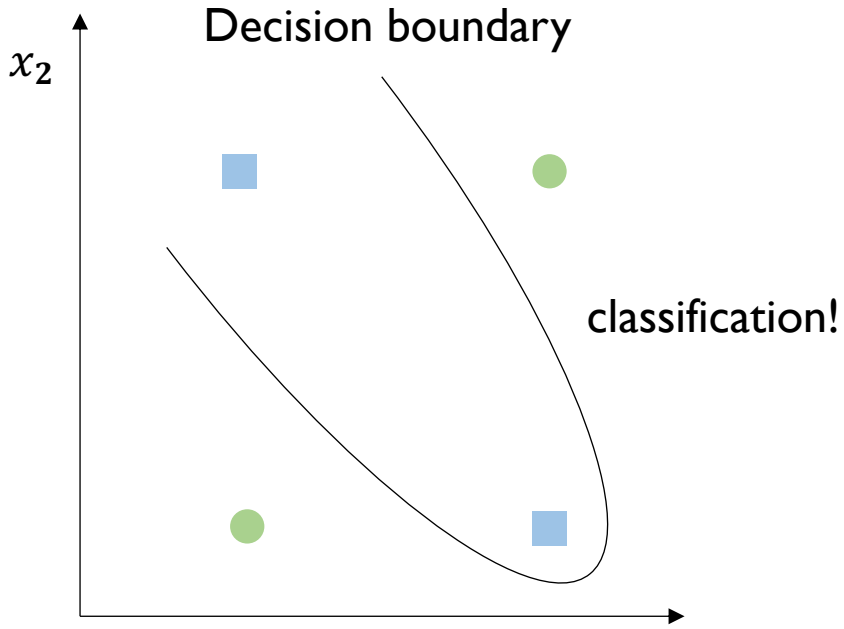
Representation Learning

- Machine learning algorithms for classification
  - linear classifiers



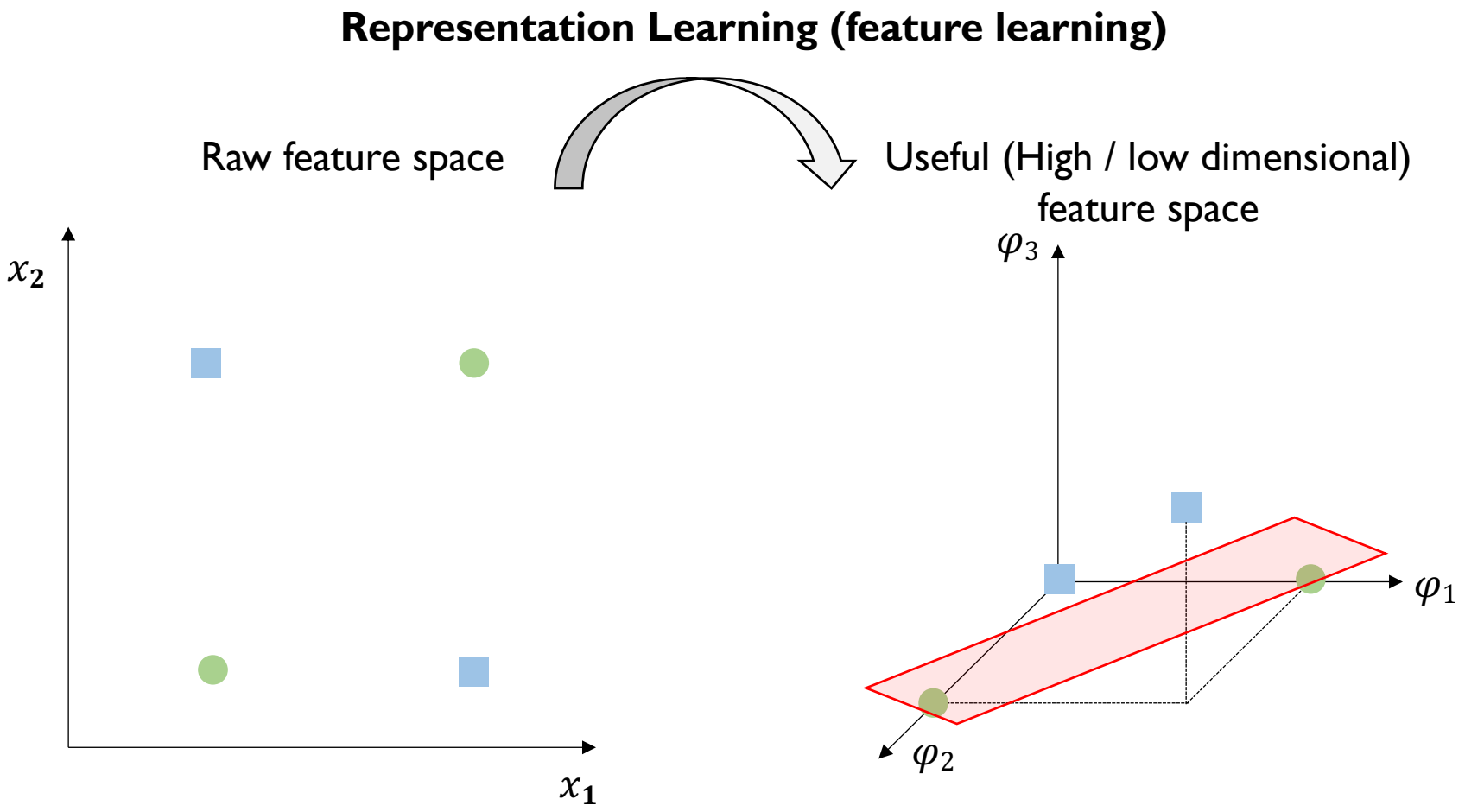
Representation Learning

- Machine learning algorithms for classification
  - linear classifiers



모든 모델은 마지막에 선형으로 분류돼야 함→ 데이터 차원(특징)을 변형하는 Representation Learning

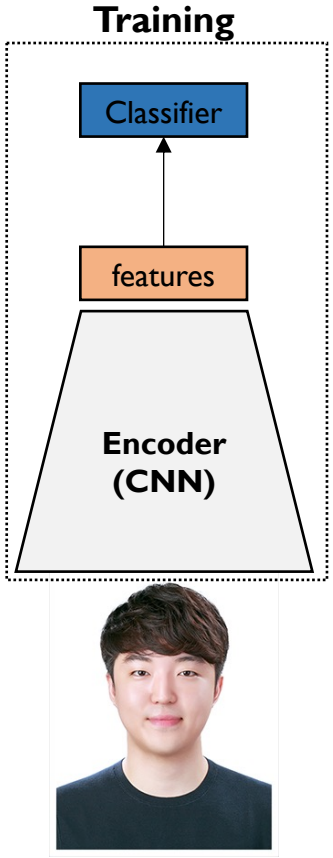
Representation Learning



Deep Learning 은 대표적인 representation learning 기법

Representation Learning

- **Representation Learning (= Feature Learning)**



- 예측 과업 (Classification / Regression)을 수행할 때 데이터를 효과적으로 구분 짓는 특징(feature)을 자동으로 추출할 수 있도록 학습하는 과정
- 추출된 Features & 정답을 softmax & cross-entropy loss에 입력

  → supervised learning
- 높은 예측성능을 기대하기 위해선 정답이 부여된 충분한 양의 데이터 필요 → 데이터 수집 비용
- 정답이 부여된 데이터가 부족한 상황에 예측성능 보장을 표현학습

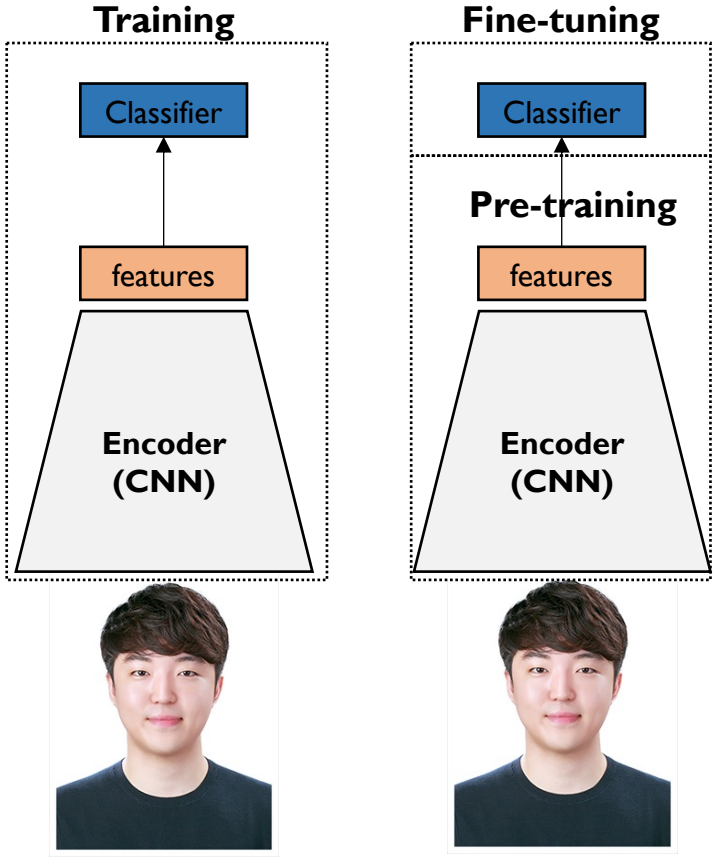
Self-Supervised Learning

Self-Supervised Contrastive Learning

(자가지도학습 기반 대조학습)

Self-Supervised Contrastive Learning

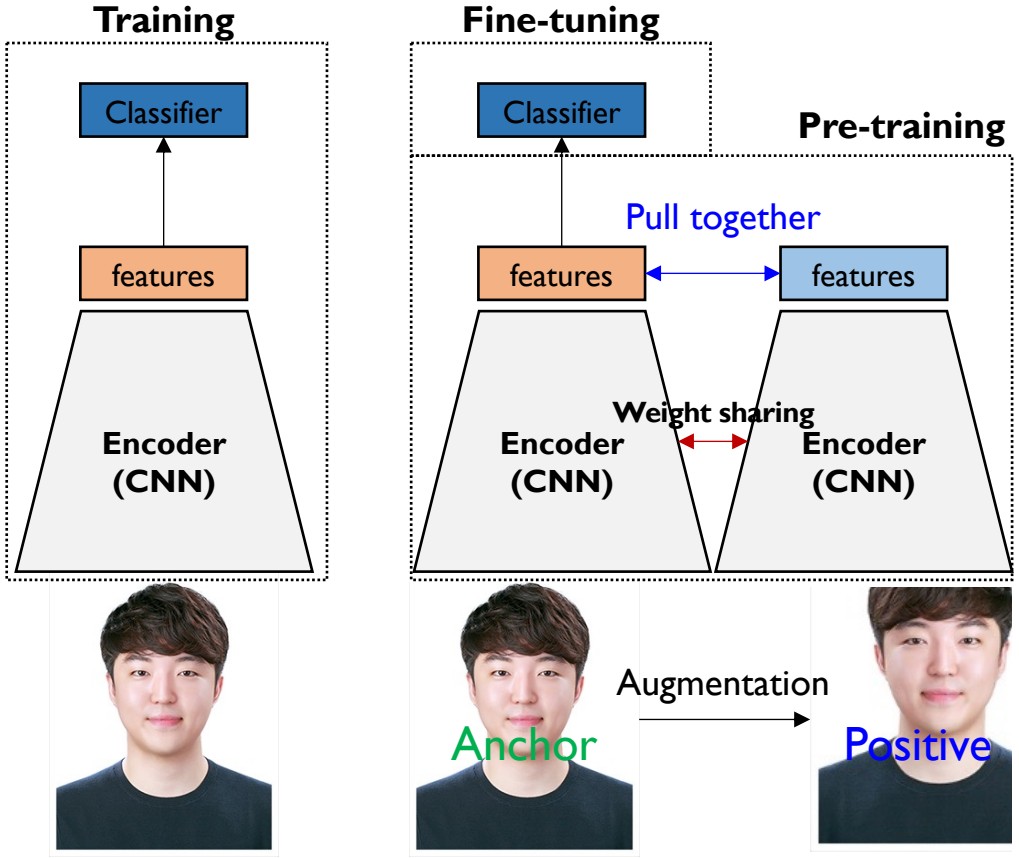
- **Supervised Learning vs. Self-Supervised Contrastive Learning**



- Pre-training (사전학습): 본 과업 목표(분류 등) 이전에 정답 없이 표현 학습
- Fine-tuning (조정학습): Pre-training 된 모델 Encoder를 사용하여 본 과업 목표(분류 등) 학습

Self-Supervised Contrastive Learning

- **Supervised Learning vs. Self-Supervised Contrastive Learning**



- Anchor: 학습 대상 데이터
- Positive: 증강 데이터
- Negative: 나머지 데이터

- Anchor를 Positive 데이터와 유사하도록 표현 학습 Pull together (나머지 Negative는 Push apart)
- 데이터를 증강하여 다양한 형태 데이터를 학습한다는 장점이 있음

Self-Supervised Contrastive Learning

- **Self-Supervised Contrastive Learning**

① Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

② Henaff, O. (2020, November). Data-efficient image recognition with contrastive predictive coding. In International Conference on Machine Learning (pp. 4182-4192). PMLR.

③ Tian, Y., Krishnan, D., & Isola, P. (2020). Contrastive multiview coding. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16 (pp. 776-794). Springer International Publishing.

④ Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., & Bengio, Y. (2018). Learning deep representations by mutual information estimation and maximization. arXiv preprint arXiv:1808.06670.

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)},$$

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \ldots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \ldots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $j(i)$: positive (anchor i 로 증강한 데이터 1개→ 유사하도록 학습)

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \dots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $j(i)$: positive (anchor i 로 증강한 데이터 1개→ 유사하도록 학습)
- $k \in A(i) \setminus \{j(i)\}$: negatives (anchor, positive 외 모든 데이터 2N-2 개 → 유사하지 않도록 학습)

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \dots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $j(i)$: positive (anchor i 로 증강한 데이터 1개→ 유사하도록 학습)
- $k \in A(i) \setminus \{j(i)\}$: negatives (anchor, positive 외 모든 데이터 2N-2 개 → 유사하지 않도록 학습)
- $a \in A(i) \equiv I \setminus \{i\}$ (anchor 외 모든 데이터 = 1개 positive + 2N-2개 negatives)

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \ldots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $j(i)$: positive (anchor i 로 증강한 데이터 1개→ 유사하도록 학습)
- $k \in A(i) \setminus \{j(i)\}$: negatives (anchor, positive 외 모든 데이터 2N-2 개 → 유사하지 않도록 학습)
- $a \in A(i) \equiv I \setminus \{i\}$ (anchor 외 모든 데이터 = 1개 positive + 2N-2개 negatives)
- $(z_i \cdot z_{j(i)})$ $(z_i \cdot z_a)$: 각각 anchor 와 positive, anchor 와 positive, negatives 간 유사도

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- Numerator $exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)$ must be maximized
- Denominator $\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)$ must be minimized

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- Numerator $exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)$ must be maximized
- Denominator $\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)$ must be minimized
  - $\tau \in \mathcal{R}^+$: temperature (hyperparameter)
    - Smaller τ benefits training more than higher ones, but extremely low temperatures are difficult to train because of numerical instability (Khosla et al., 2020).

Self-Supervised Learning

- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(z_i \cdot z_{j(i)} / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)},$$

- Numerator $exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)$ must be maximized
- Denominator $\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)$ must be minimized
  - $\tau \in \mathcal{R}^+$: temperature (hyperparameter)
    - Smaller τ benefits training more than higher ones, but extremely low temperatures are difficult to train because of numerical instability (Khosla et al., 2020).

- $log \frac{exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)} \uparrow$

Self-Supervised Learning

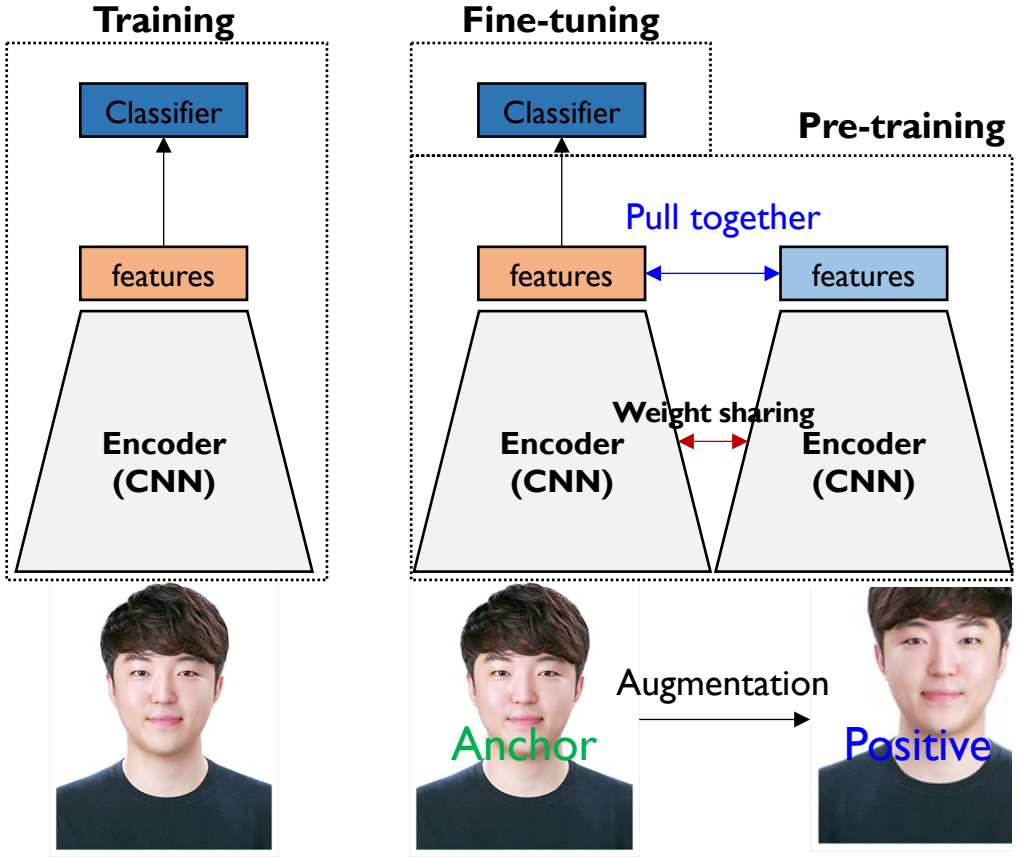
- **Self-Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)},$$

- Numerator $exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)$ must be maximized
- Denominator $\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)$ must be minimized
  - $\tau \in \mathcal{R}^+$: temperature (hyperparameter)
    - Smaller τ benefits training more than higher ones, but extremely low temperatures are difficult to train because of numerical instability (Khosla et al., 2020).
- $log \frac{exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)} \uparrow \rightarrow -\sum_{i \in I} log \frac{exp(\boldsymbol{z_i} \cdot \boldsymbol{z_{j(i)}} / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z_i} \cdot \boldsymbol{z_a} / \tau)} \downarrow$ (minimizing loss)

Self-Supervised Contrastive Learning

- **Supervised vs. Self-Supervised Contrastive Learning**



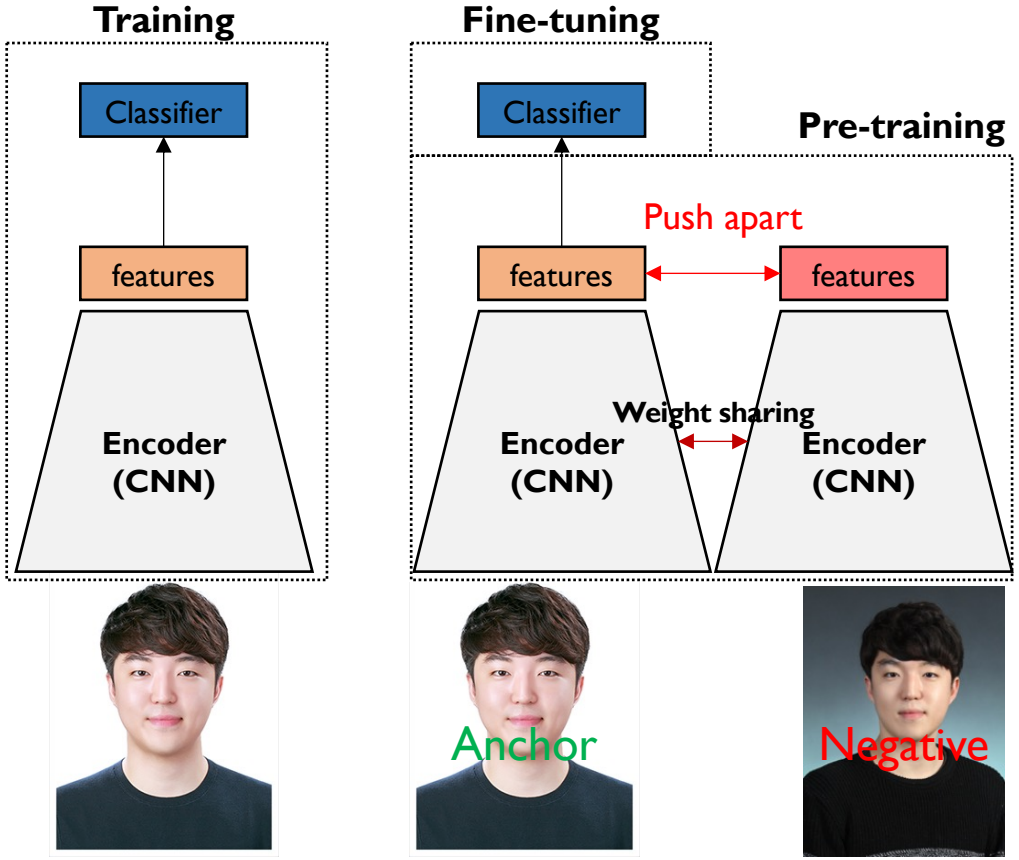
- Anchor: 학습 대상 데이터
- Positive: 증강 데이터
- Negative: 나머지 데이터

- **증강 데이터와 유사하도록 표현 학습 Pull together (나머지는 다르도록 Push apart)**

  → 데이터를 증강하여 다양한 형태 데이터를 학습한다는 장점이 있음

Self-Supervised Contrastive Learning

- **Supervised vs. Self-Supervised Contrastive Learning**



- Anchor: 학습 대상 데이터
- Positive: 증강 데이터
- Negative: 나머지 데이터

- **증강 데이터와 유사하도록 표현 학습 Pull together (나머지는 다르도록 Push apart)**

  → 하지만 같은 정답이 부여된 데이터도 Negative 로 부여한다는 단점이 존재

Supervised Contrastive Learning

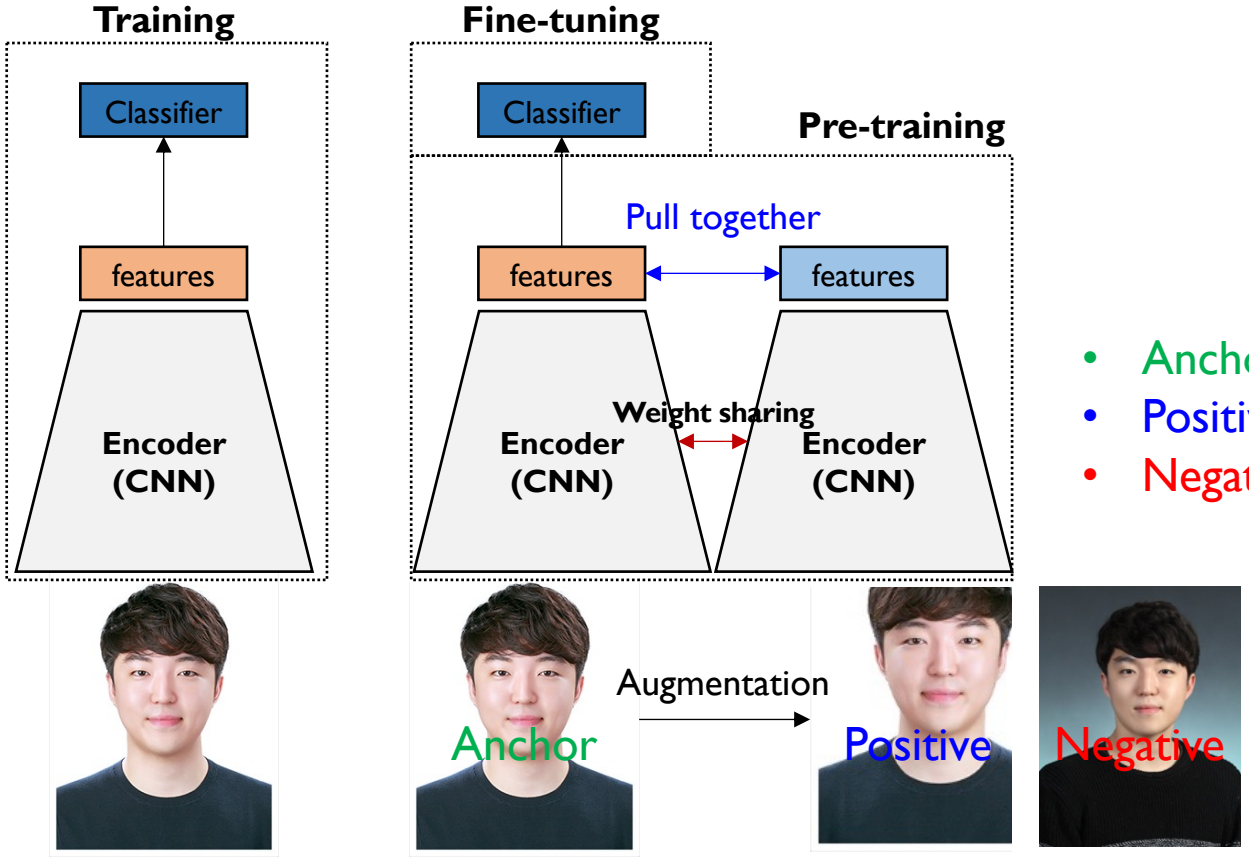
(지도학습 기반 대조학습)

Supervised Contrastive Learning

- **Supervised Learning (지도학습)**
- **Self-supervised Contrastive Learning (자가지도 대조학습)**
- **각각의 장점을 통합 활용한 기법**

Supervised Contrastive Learning

- **Supervised vs. Self-Supervised Contrastive Learning**

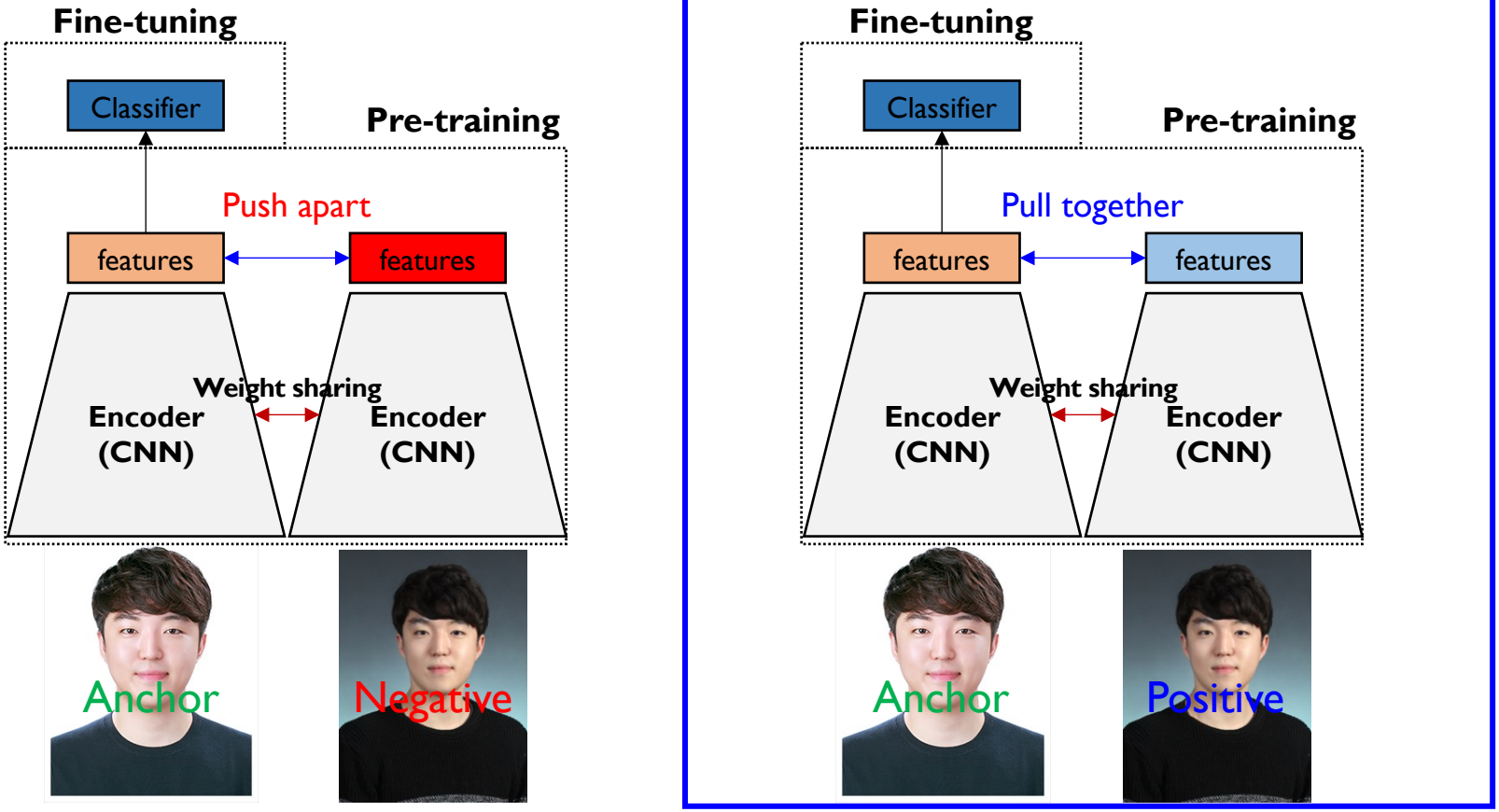


- Anchor: 학습 대상 데이터
- Positive: 증강 데이터
- Negative: 나머지 데이터

- **Supervised learning 장점: 레이블 정보를 전부 활용한다는 점**
- **Self-Supervised Contrastive Learning 장점: 데이터 증강기법 & 2단계 풍부한 표현학습**

Supervised Contrastive Learning

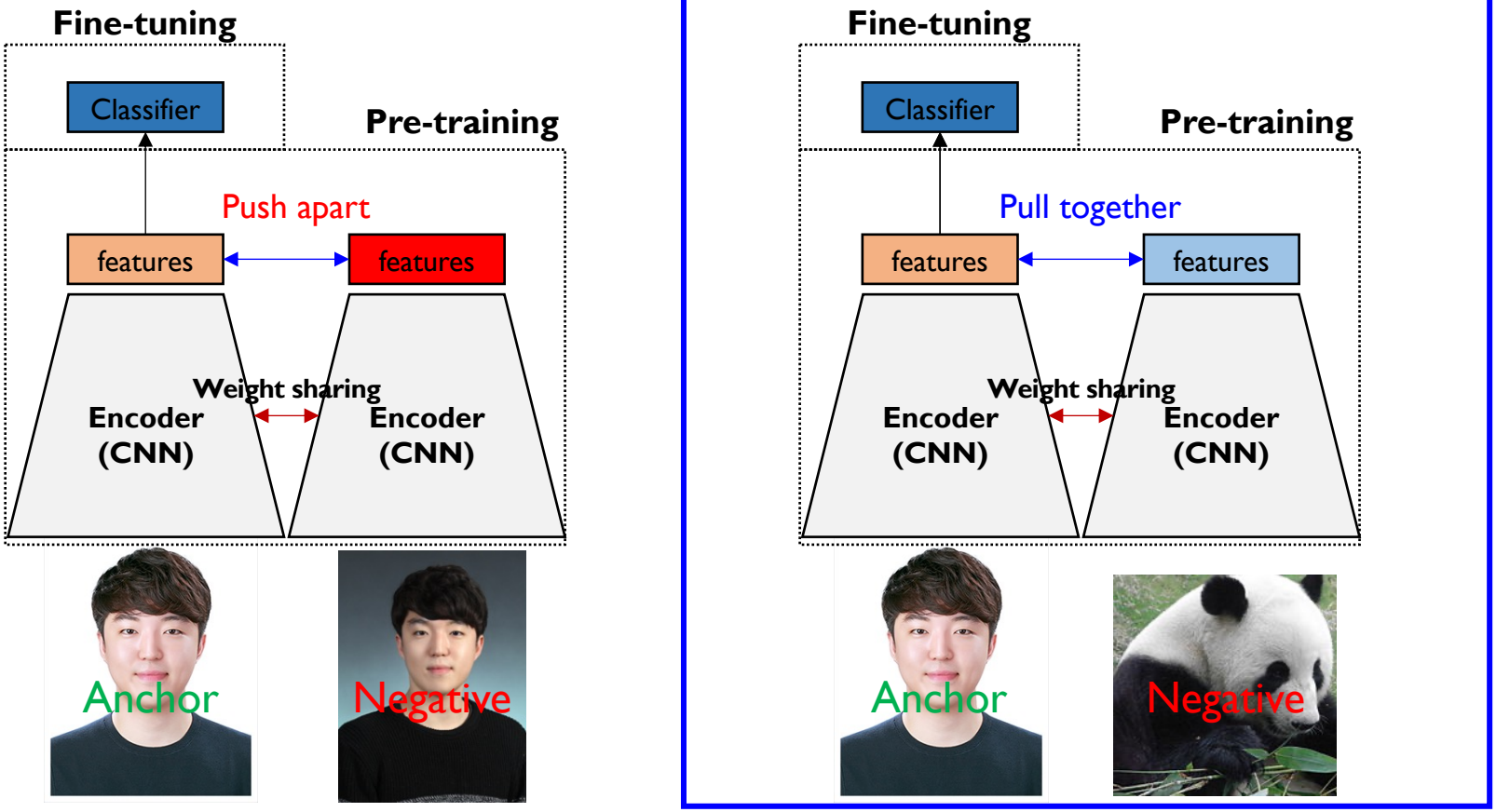
- **Self-Supervised Contrastive Learning** **vs.** **Supervised Contrastive learning**



- 같은 레이블 데이터와 유사하도록 표현 학습 → **Pull together**

Supervised Contrastive Learning

- **Self-Supervised Contrastive Learning** **vs.** **Supervised Contrastive learning**

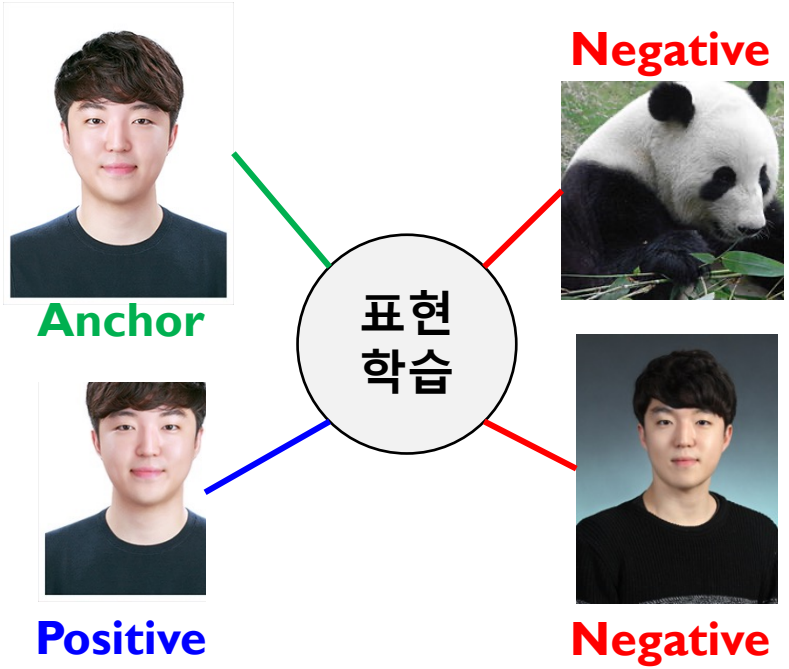


- 같은 레이블 데이터와 유사하도록 표현 학습 → **Pull together**
- 다른 레이블 데이터와 유사하지 않도록 표현 학습 → **Push apart**

Supervised Contrastive Learning

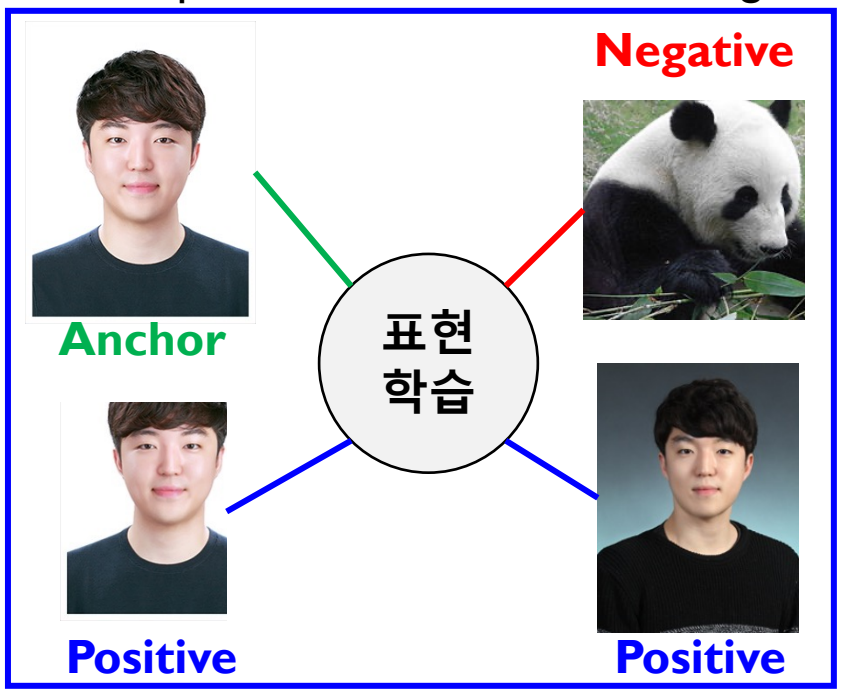
- **Supervised Contrastive Learning**
  - **Pull together (positive), Push apart (negative)**

Self-Supervised Contrastive Learning



정답이 부족한 상황에서의 표현 학습

Supervised Contrastive Learning



정답이 충분한 상황에서의 표현 학습

Supervised Contrastive Learning

- **Supervised Contrastive Learning (Loss)**

추가 Self-supervised contrastive loss

$$L = \sum_{i \in I} L_i^{sup} = \sum_{i \in I} -log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \right\},$$

Supervised Contrastive Learning

- **Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{sup} = \sum_{i \in I} -log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \right\},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \dots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)

Supervised Contrastive Learning

- **Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{sup} = \sum_{i \in I} -log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \right\},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \ldots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $P(i) \equiv \{p \in A(i) : \tilde{y}_p = \tilde{y}_i\}$ = set of indices of all positive samples
  - Positive sample 이 여러 개 → self-supervised contrastive learning은 positive 가 1개
  - Positive samples $P(i)$:Anchor의 증강 데이터,Anchor와 같은 레이블 데이터,
  - $|P(i)|$ = cardinality (the number of positive samples)

Supervised Contrastive Learning

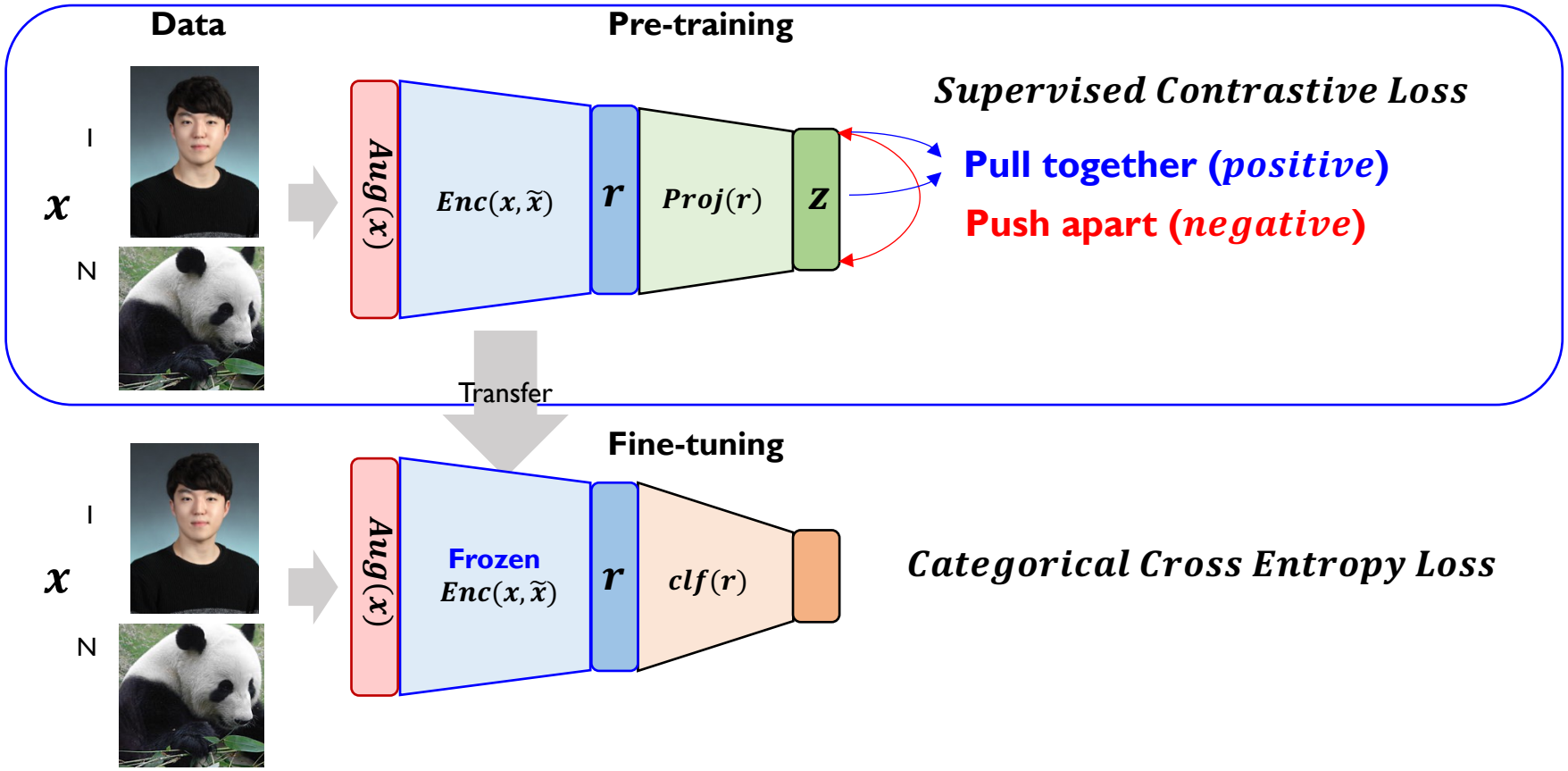
- **Supervised Contrastive Learning (Loss)**

$$L = \sum_{i \in I} L_i^{sup} = \sum_{i \in I} -log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \right\},$$

- i: Anchor (학습 대상 데이터)
- $i \in I \equiv \{1, \dots, 2N\}$ → 학습 데이터 N개 + 증강 데이터 N개 (multi-viewed data)
- $P(i) \equiv \{p \in A(i) : \tilde{y}_p = \tilde{y}_i\}$ = set of indices of all positive samples
  - Positive sample 이 여러 개 → self-supervised contrastive learning은 positive 가 1개
  - Positive samples $P(i)$:Anchor의 증강 데이터,Anchor와 같은 레이블 데이터,
  - $|P(i)|$ = cardinality (the number of positive samples)
- $a \in A(i) \equiv I \setminus \{i\}$ (anchor 외 모든 데이터 = $|P(i)|$ 개 positives + 2N − $|P(i)|$개 negatives)

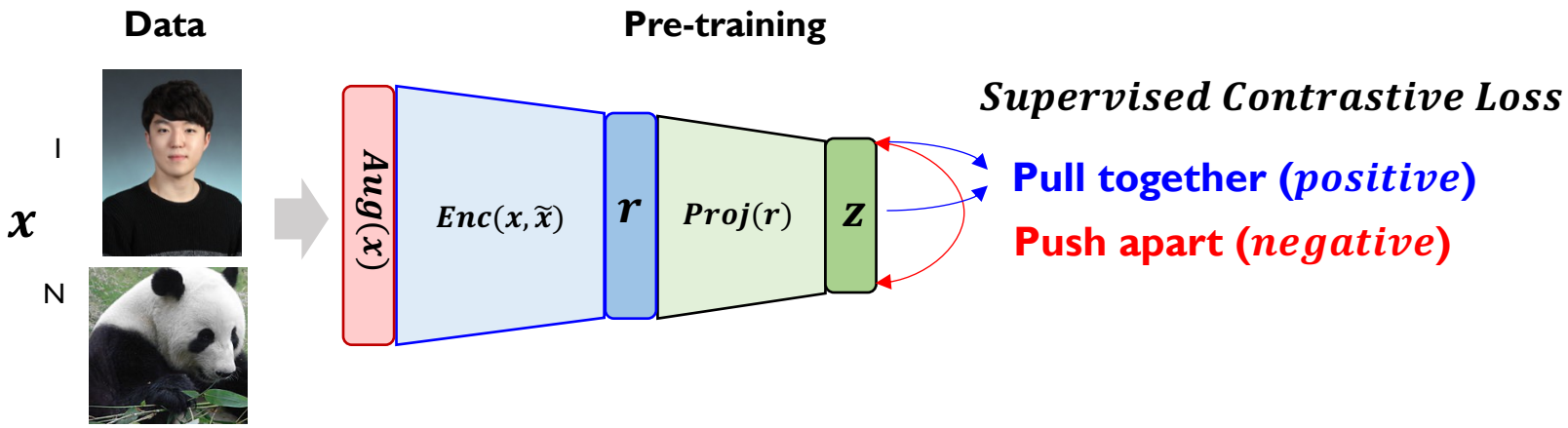
Supervised Contrastive Learning

- **Supervised Contrastive Learning**
  - 적용 과정



Supervised Contrastive Learning

- **Supervised Contrastive Learning**
  - Stage 1: Pre-training



- **Pre-training 구성 요소 3가지**

① Augmentation module $\boldsymbol{Aug}(x) = \tilde{x}$ → 데이터 증강 기법 적용하여 다양한 패턴 반영

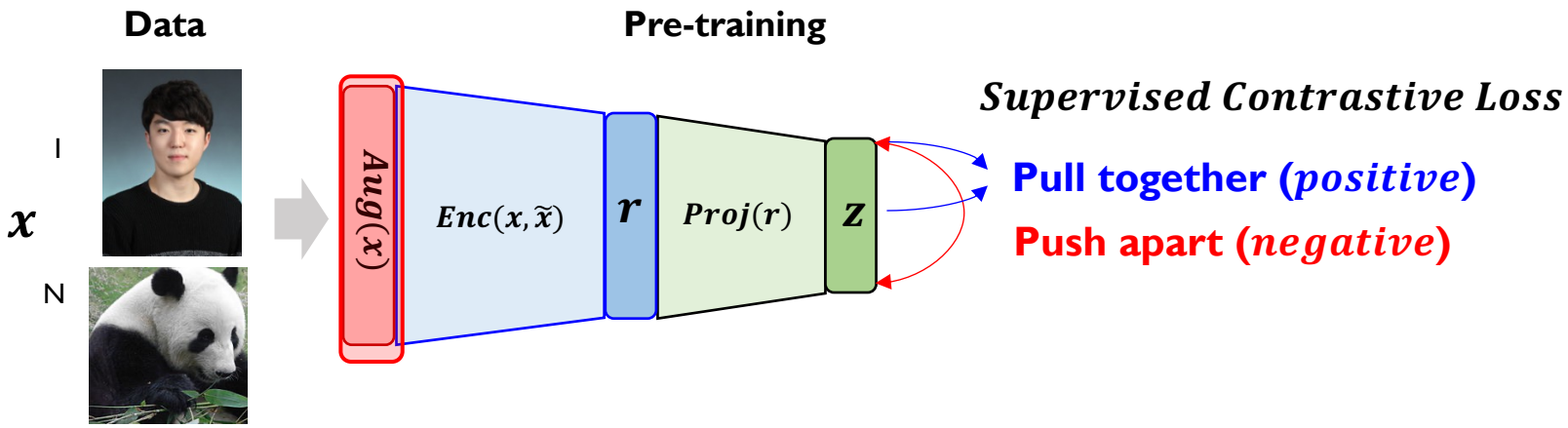
② Encoder $\boldsymbol{Enc}(\tilde{x}) = \boldsymbol{r}$ → CNN 모델 구조 적용하여 특징 추출

③ Projection head $\boldsymbol{Proj}(\boldsymbol{r}) = \boldsymbol{z}$ → DNN 모델 및 L2정규화 한 특징 추출

  - Supervised contrastive loss 적용

Supervised Contrastive Learning

- **Supervised Contrastive Learning**
  - Stage 1: Pre-training



- **Pre-training 구성 요소 3가지**

① Augmentation module $\boldsymbol{Aug}(x) = \tilde{x}$ → 데이터 증강 기법 적용하여 다양한 패턴 반영

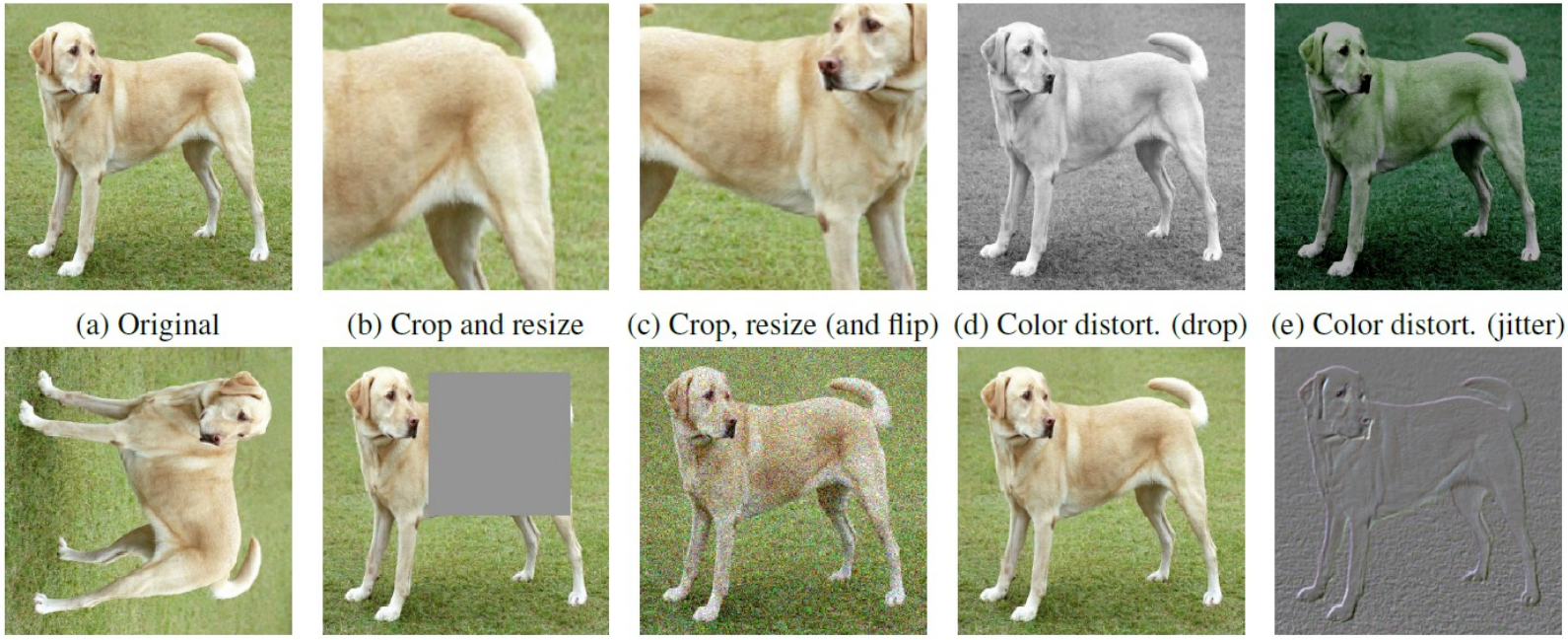
② Encoder $\boldsymbol{Enc}(\tilde{x}) = \boldsymbol{r}$ → CNN 모델 구조 적용하여 특징 추출

③ Projection head $\boldsymbol{Proj}(\boldsymbol{r}) = \boldsymbol{z}$ → DNN 모델 및 L2정규화 한 특징 추출

  - Supervised contrastive loss 적용

Supervised Contrastive Learning

- Augmentation module $\boldsymbol{Aug}(x) = \tilde{x}$ → 데이터 증강 기법 적용하여 다양한 패턴 반영

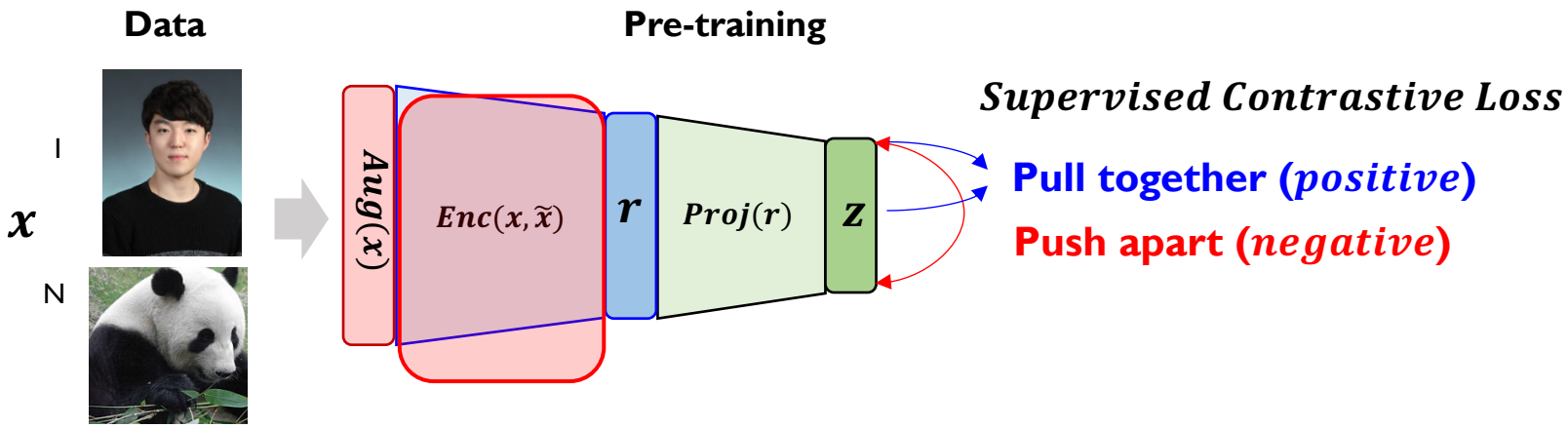


- 매 학습(epoch)마다 증강기법 종류나 변형 크기를 Random하게 적용
- 여러 종류 Data Augmentation 기법을 동시에 적용할 수 있음

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

Supervised Contrastive Learning

- **Supervised Contrastive Learning**
  - Stage 1: Pre-training



- **Pre-training 구성 요소 3가지**

① Augmentation module $\boldsymbol{Aug}(x) = \tilde{x}$ → 데이터 증강 기법 적용하여 다양한 패턴 반영

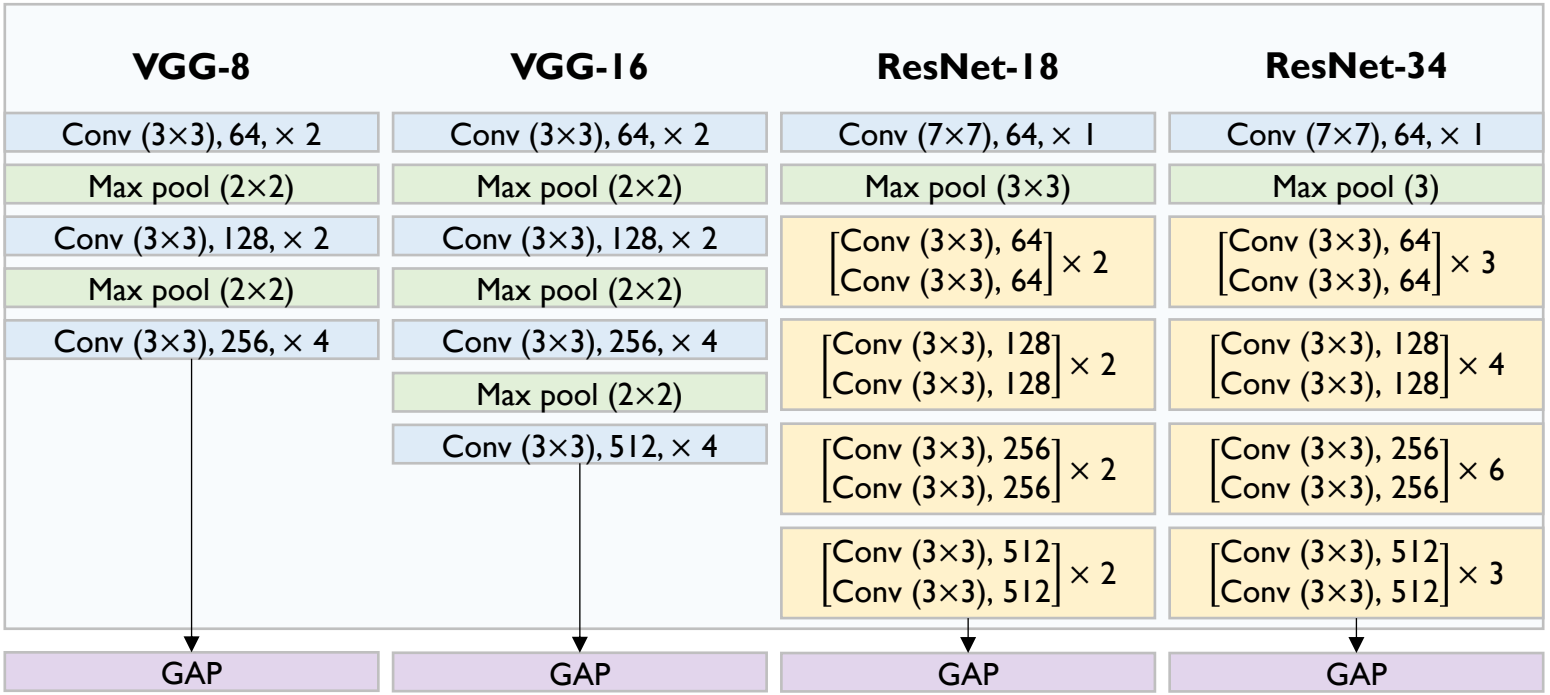
② Encoder $\boldsymbol{Enc}(\tilde{x}) = \boldsymbol{r}$ → CNN 모델 구조 적용하여 특징 추출

③ Projection head $\boldsymbol{Proj}(\boldsymbol{r}) = \boldsymbol{z}$ → DNN 모델 및 L2정규화 한 특징 추출

  - Supervised contrastive loss 적용

Supervised Contrastive Learning

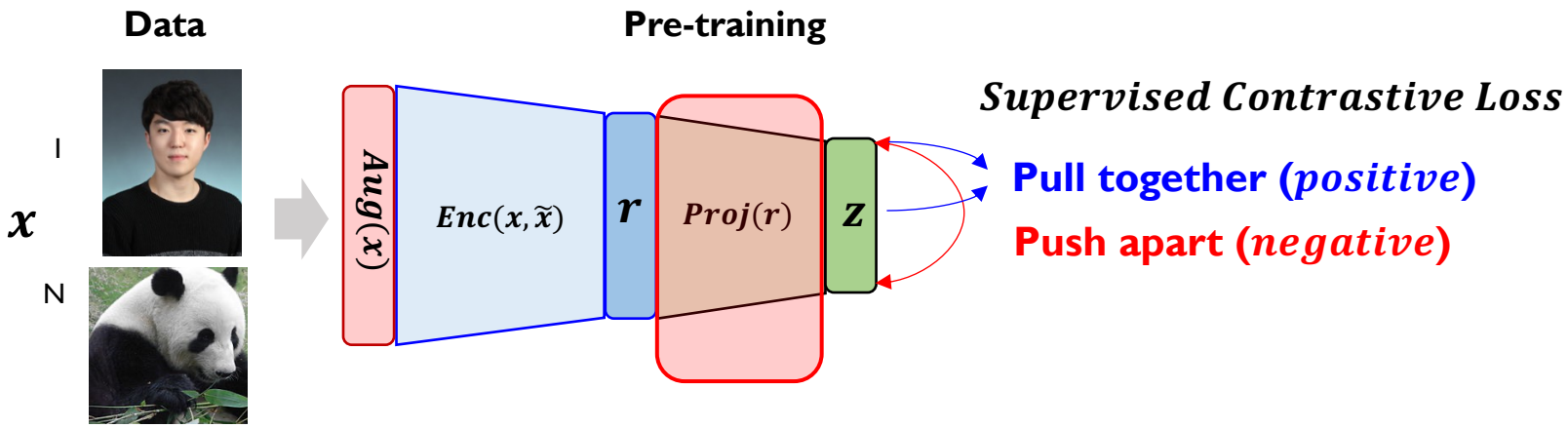
- **Supervised Contrastive Learning:** Stage 1: Pre-training
  - Encoder $\boldsymbol{Enc}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \boldsymbol{r}$ → CNN 모델 구조 적용하여 특징 추출



- CNN 모델 구조(feature extraction, classification network)에서 feature extraction만 사용
- Feature extraction 이후 Global average pooling을 추가하여 특징백터를 추출하도록 $\boldsymbol{Enc}(\cdot)$ 구성

Supervised Contrastive Learning

- **Supervised Contrastive Learning**
  - Stage 1: Pre-training



- **Pre-training 구성 요소 3가지**

① Augmentation module $Aug(x) = \tilde{x}$ → 데이터 증강 기법 적용하여 다양한 패턴 반영

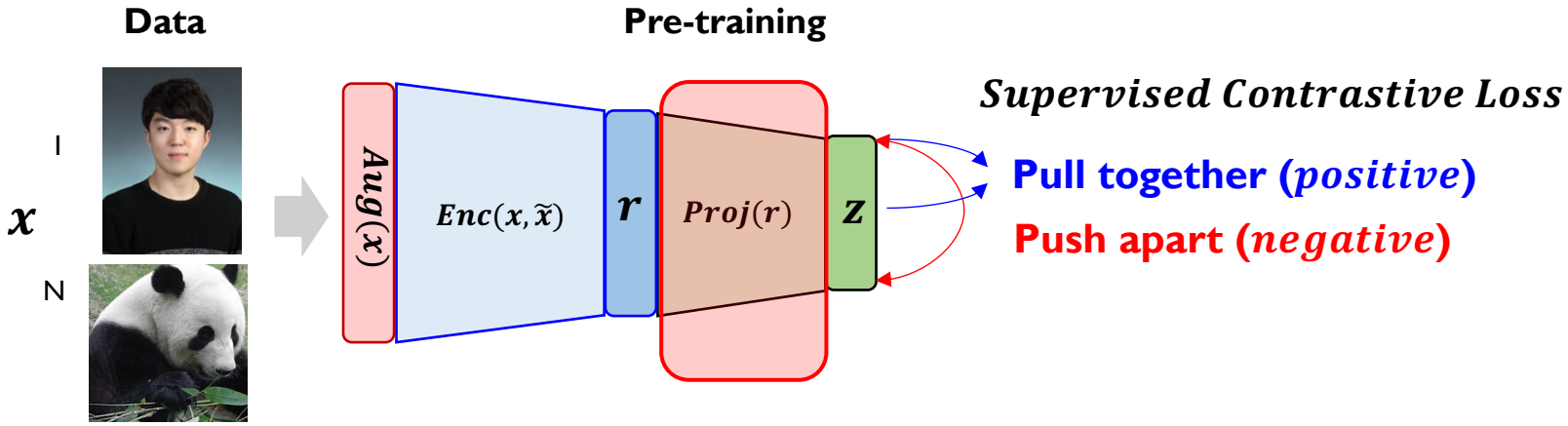
② Encoder $Enc(\tilde{x}) = r$ → CNN 모델 구조 적용하여 특징 추출

③ Projection head $Proj(r) = z$ → DNN 모델 및 L2정규화 한 특징 추출

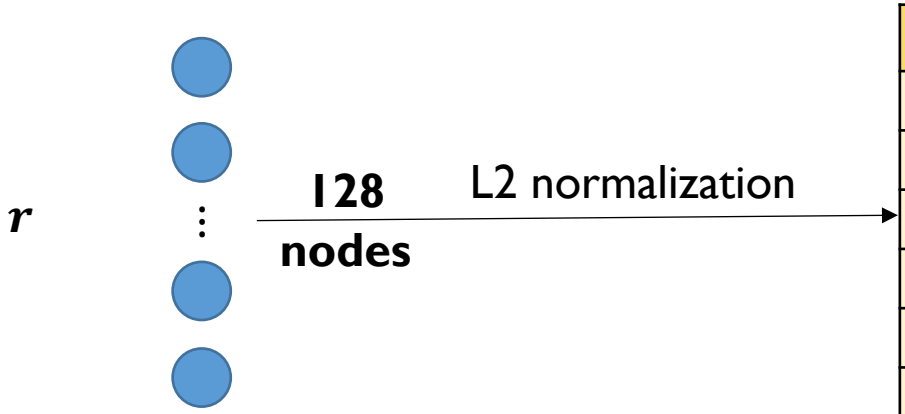
  - Supervised contrastive loss 적용

Supervised Contrastive Learning

- **Supervised Contrastive Learning:** Stage 1: Pre-training



- Projection Head (Dense layer 1개, node 128개, L2 normalization layer)



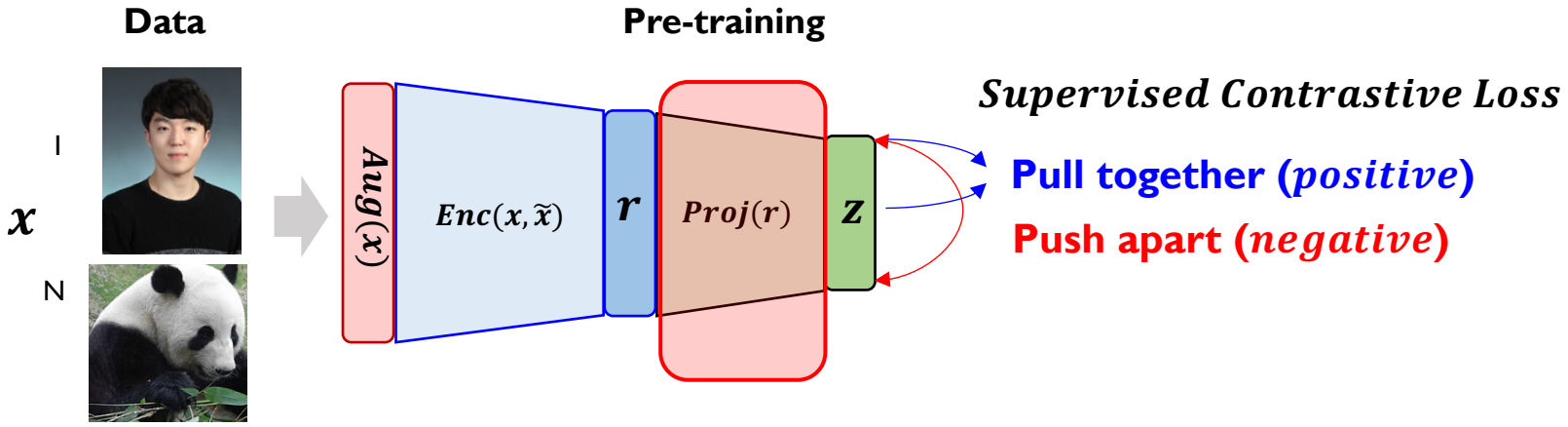
Batch	Feature 1	Feature 2	…	Feature 127	Feature 128
1	…	…	…	…	…
…	…	…	…	…	…
N	…	…	…	…	…
N+1	…	…	…	…	…
…	…	…	…	…	…
2N	…	…	…	…	…

Normalized Features (원본 데이터)

Normalized Features (증강 데이터)

Supervised Contrastive Learning

- **Supervised Contrastive Learning:** Stage 1: Pre-training



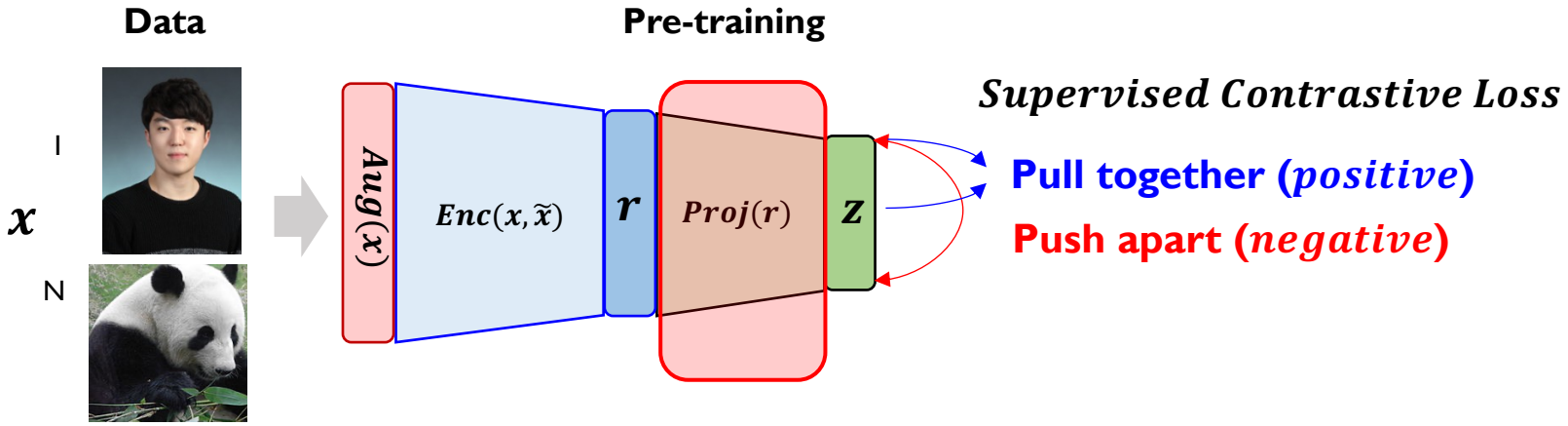
Features Matrix Multiplication **Features^T**

Batch	Feature 1	Feature 2	…	Feature 127	Feature 128
1	…	…	…	…	…
…	…	…	…	…	…
N	…	…	…	…	…
N+1	…	…	…	…	…
…	…	…	…	…	…
2N	…	…	…	…	…

Batch	1	…	N	N+1	…	2N
Feature 1	…	…	…	…	…	…
Feature 2	…	…	…	…	…	…
…	…	…	…	…	…	…
Feature 127	…	…	…	…	…	…
Feature 128	…	…	…	…	…	…

Supervised Contrastive Learning

- **Supervised Contrastive Learning:** Stage 1: Pre-training



$$\mathbf{z} = (features \cdot features^{T})$$

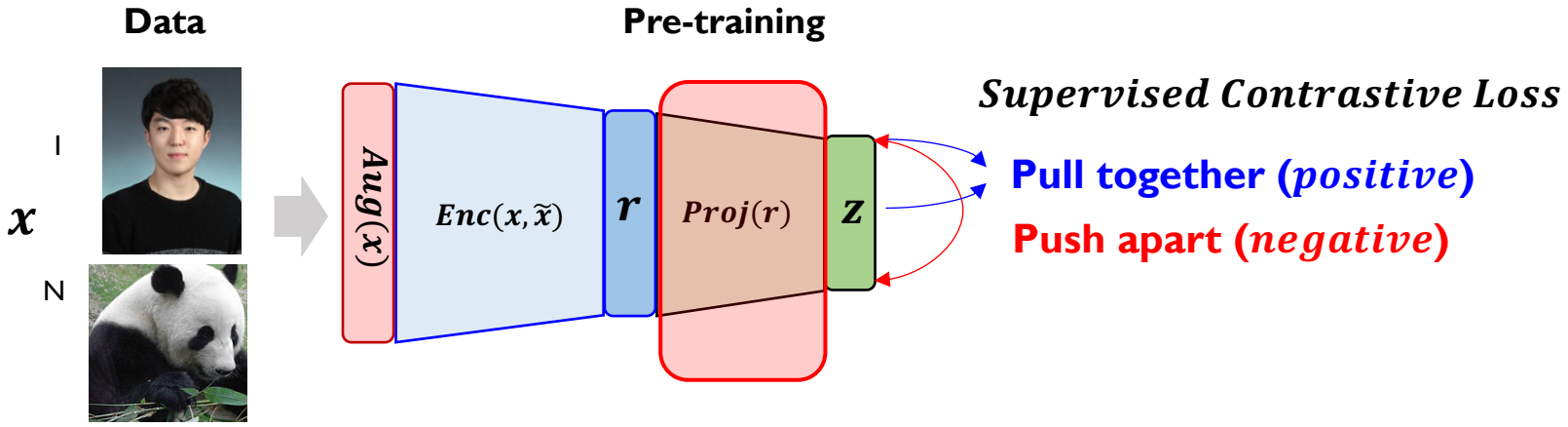
$\mathbf{z}$	z_1	$\cdots$	z_N	z_{N+1}	$\cdots$	z_{2N}
z_1	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
z_N	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
z_{N+1}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
z_{2N}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Supervised Contrastive Loss

$$\sum_{i \in I} -log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \right\}$$

Supervised Contrastive Learning

- **Supervised Contrastive Learning:** Stage 1: Pre-training



$$\mathbf{z} = (features \cdot features^T)$$

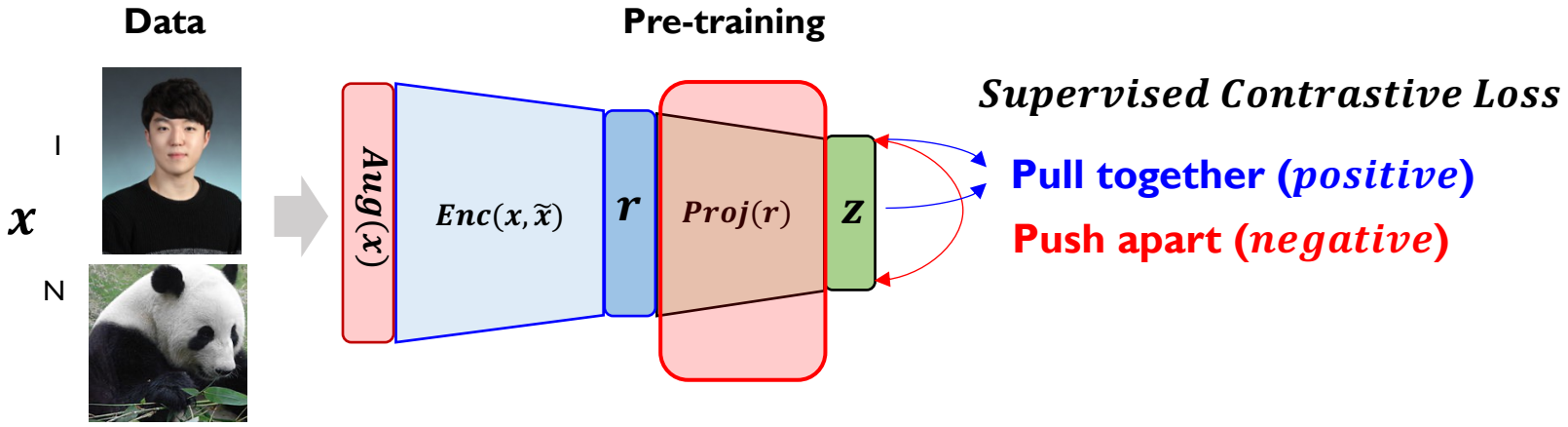
$\mathbf{z}$	z_1	⋯	z_N	z_{N+1}	⋯	z_{2N}
z_1	⋯	⋯	⋯	⋯	⋯	⋯
⋯	⋯	⋯	⋯	⋯	⋯	⋯
z_N	⋯	⋯	⋯	⋯	⋯	⋯
z_{N+1}	⋯	⋯	⋯	⋯	⋯	⋯
⋯	⋯	⋯	⋯	⋯	⋯	⋯
z_{2N}	⋯	⋯	⋯	⋯	⋯	⋯

Class
A
⋯
A
A
⋯
B

- Batch 내 1번 관측치 (anchor) 가 다른 관측치와의 유사도

Supervised Contrastive Learning

- **Supervised Contrastive Learning:** Stage 1: Pre-training



$$\mathbf{z} = (features \cdot features^T)$$

$\mathbf{z}$	z_1	⋯	z_N	z_{N+1}	⋯	z_{2N}
z_1	⋯	⋯	⋯	⋯	⋯	⋯
⋯	⋯	⋯	⋯	⋯	⋯	⋯
z_N	⋯	⋯	⋯	⋯	⋯	⋯
z_{N+1}	⋯	⋯	⋯	⋯	⋯	⋯
⋯	⋯	⋯	⋯	⋯	⋯	⋯
z_{2N}	⋯	⋯	⋯	⋯	⋯	⋯

Class
A
⋯
A
A
⋯
B

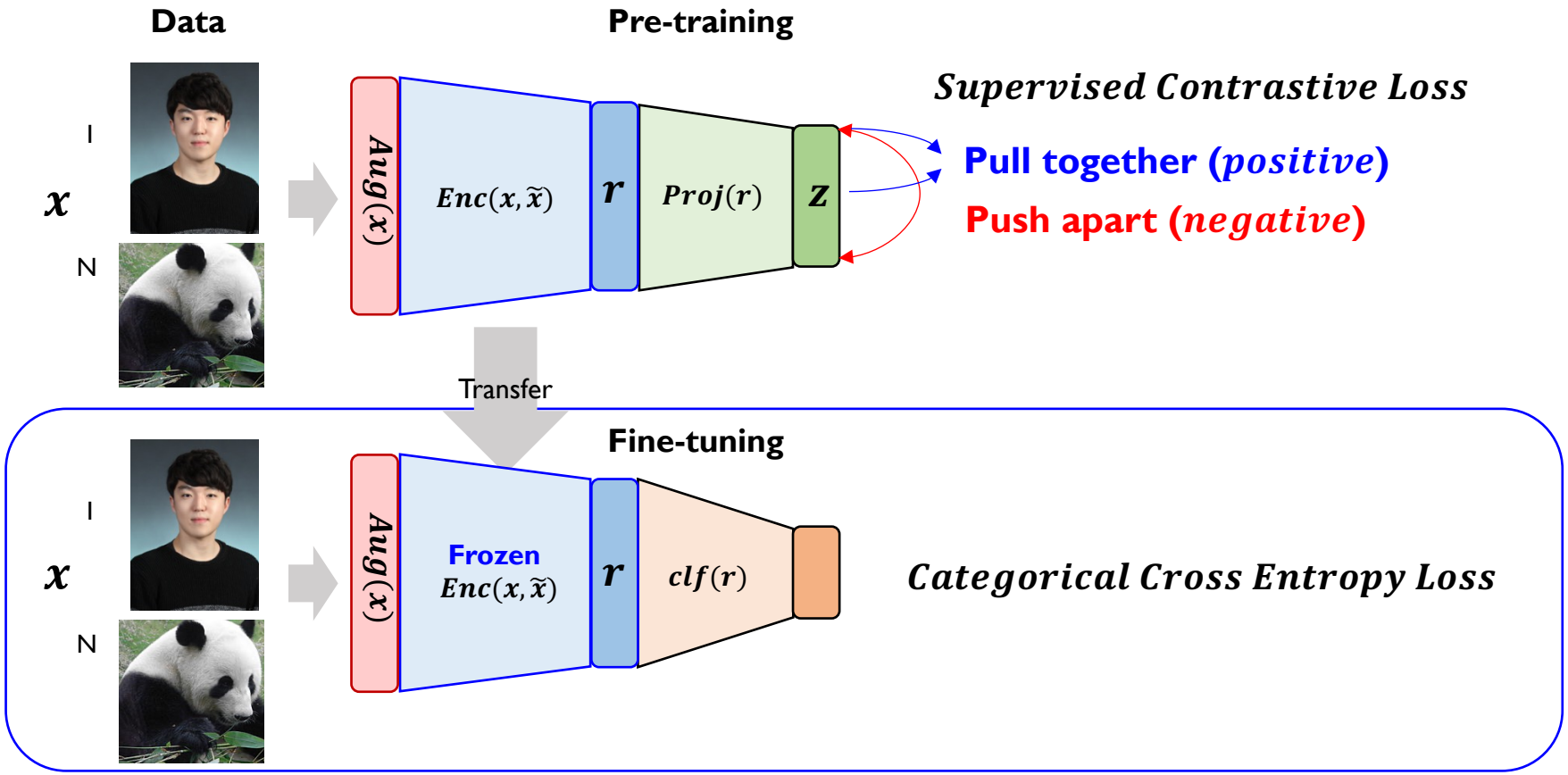
- Batch 내 1번 관측치 (anchor) 가 다른 관측치와의 유사도

→ **Pull together (*positive*)**

→ **Push apart (*negative*)**

Supervised Contrastive Learning

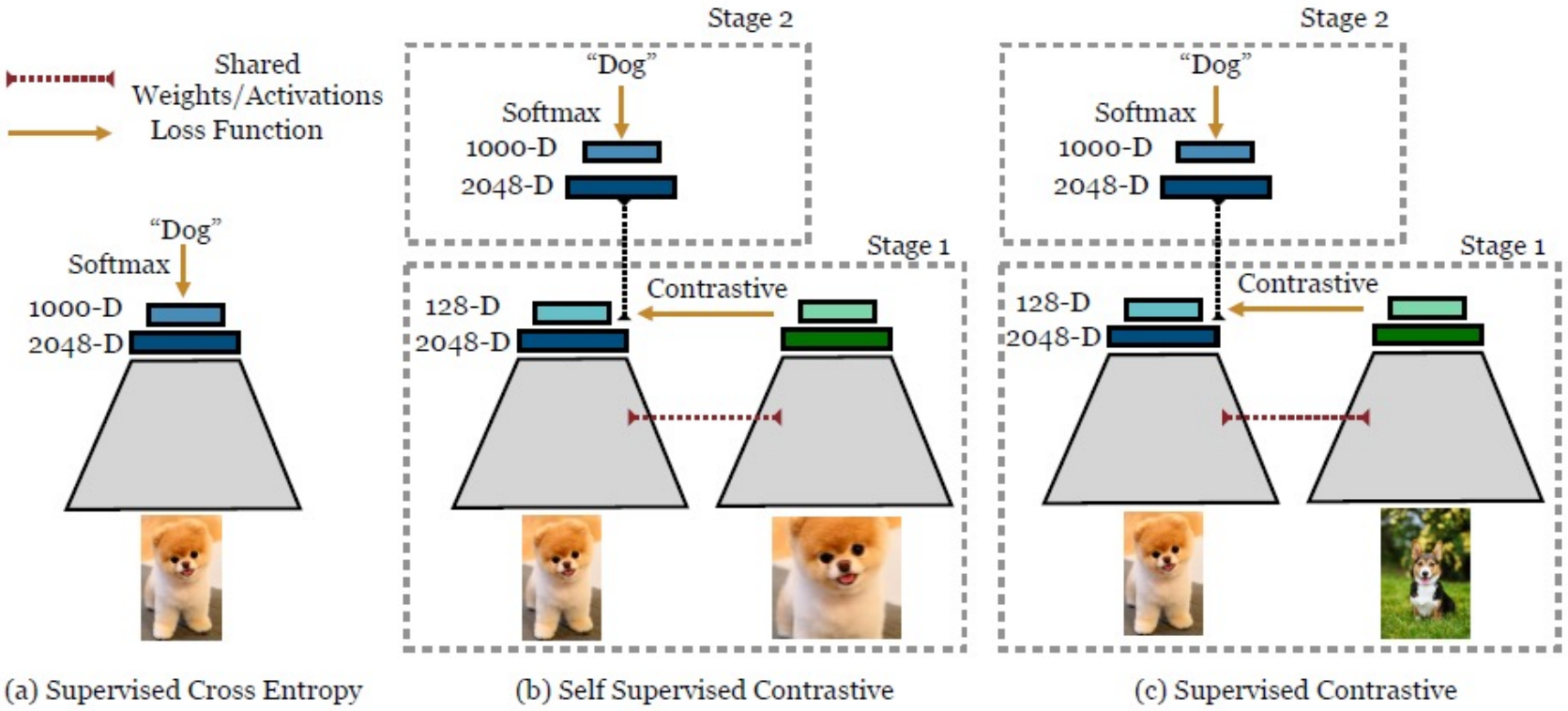
- **Supervised Contrastive Learning: stage 2 (fine-tuning)**



- 학습된 Encoder를 고정하고 Supervised Learning으로 Fine-tuning 진행
- 더 Deep한 $clf(r)$는 정확도 향상을 기대할 수 있지만 $Supervised\ Contrastive\ Loss$ 만의 성능을 살펴보기 위해 실험적으로는 1층 신경망(혹은 선형모델)으로 구성

Supervised Contrastive Learning

- **Supervised Contrastive Learning 을 활용한 연구 (1) → Image Data**
  - 사전학습(pre-training) 단계에서 Supervised Contrastive Learning 수행
  - 조정학습(fine-tuning) 단계에서 이미지 데이터 분류



Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*.

Supervised Contrastive Learning

- **Supervised Contrastive Learning 을 활용한 연구 (1) → Image Data**
  - 사전학습(pre-training) 단계에서 Supervised Contrastive Learning 수행
  - 조정학습(fine-tuning) 단계에서 이미지 데이터 분류
    - **전통적 지도학습, 최신 자가지도학습 대비 예측성능 향상**

Dataset	SimCLR[3]	Cross-Entropy	Max-Margin [32]	SupCon
CIFAR10	93.6	95.0	92.4	**96.0**
CIFAR100	70.7	75.3	70.5	**76.5**
ImageNet	70.2	78.2	78.0	**78.7**

Table 2: Top-1 classification accuracy on ResNet-50 [17] for various datasets. We compare cross-entropy training, unsupervised representation learning (SimCLR [3]), max-margin classifiers [32] and SupCon (ours). We re-implemented and tuned hyperparameters for all baseline numbers except margin classifiers where we report published results. Note that the CIFAR-10 and CIFAR-100 results are from our PyTorch implementation and ImageNet from our TensorFlow implementation.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*.

Supervised Contrastive Learning

- **Supervised Contrastive Learning 을 활용한 연구 (1) → Image Data**
  - 사전학습(pre-training) 단계에서 Supervised Contrastive Learning 수행
  - 조정학습(fine-tuning) 단계에서 이미지 데이터 분류
    - **ImageNet 데이터 기반 여러 학습/증강기법별 비교에서 역시 최고 성능**

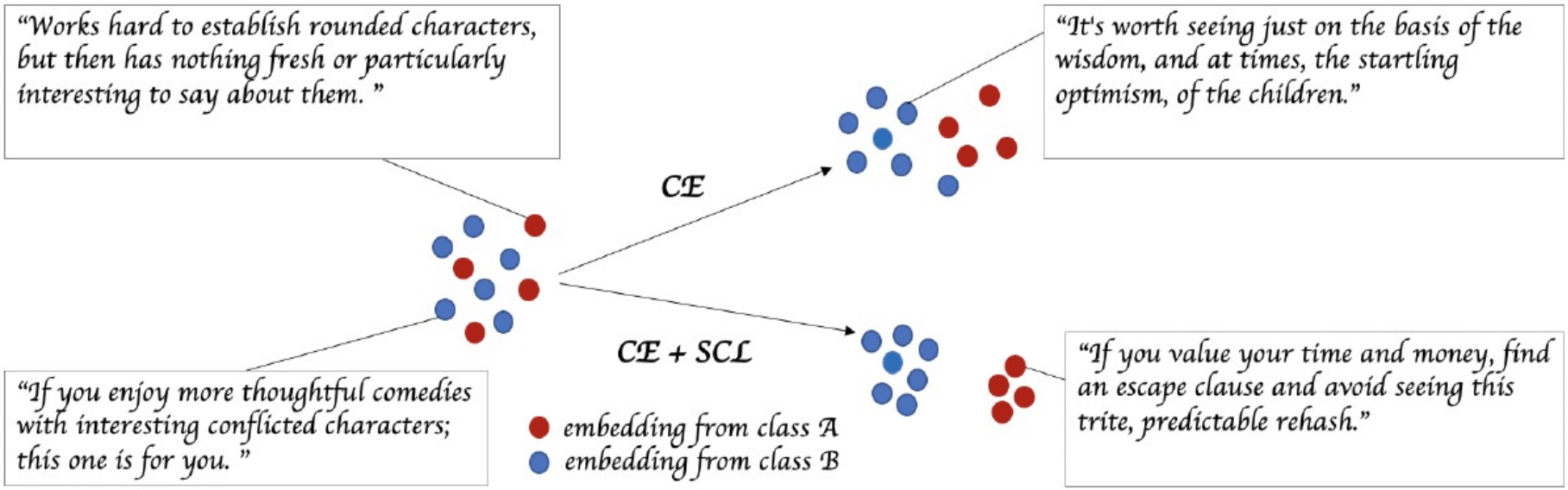
Loss	Architecture	Augmentation	Top-1	Top-5
Cross-Entropy (baseline)	ResNet-50	MixUp [61]	77.4	93.6
Cross-Entropy (baseline)	ResNet-50	CutMix [60]	78.6	94.1
Cross-Entropy (baseline)	ResNet-50	AutoAugment [5]	78.2	92.9
Cross-Entropy (our impl.)	ResNet-50	AutoAugment [30]	77.6	95.3
SupCon	ResNet-50	AutoAugment [5]	**78.7**	**94.3**
Cross-Entropy (baseline)	ResNet-200	AutoAugment [5]	80.6	95.3
Cross-Entropy (our impl.)	ResNet-200	Stacked RandAugment [49]	80.9	95.2
SupCon	ResNet-200	Stacked RandAugment [49]	**81.4**	**95.9**
SupCon	ResNet-101	Stacked RandAugment [49]	80.2	94.7

Table 3: Top-1/Top-5 accuracy results on ImageNet for AutoAugment [5] with ResNet-50 and for Stacked RandAugment [49] with ResNet-101 and ResNet-200. The baseline numbers are taken from the referenced papers, and we also re-implement cross-entropy.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*.

Supervised Contrastive Learning

- **Supervised Contrastive Learning 을 활용한 연구 (2) → Text Data**
  - 사전학습(pre-training) 된 Language 모델을 활용
  - 조정학습(fine-tuning) 단계에서 Cross entropy (CE) + SCL 활용하여 성능향상



Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2020). Supervised contrastive learning for pre-trained language model fine-tuning. arXiv preprint arXiv:2011.01403.

Supervised Contrastive Learning

- **Supervised Contrastive Learning 을 활용한 연구 (2) → Text Data**
  - 사전학습(pre-training) 된 Language 모델을 활용
  - 조정학습(fine-tuning) 단계에서 Cross entropy (CE) + SCL 활용하여 성능향상

Model	Loss	N	SST-2	QNLI	MNLI
$\text{RoBERTa}_{\text{Large}}$	CE	20	85.9±2.1	65.0±2.0	39.3±2.5
$\text{RoBERTa}_{\text{Large}}$	CE + SCL	20	**88.1±3.3**	**75.7±4.8**	**42.7±4.6**
	p-value		5e-10	1e-46	1e-8
$\text{RoBERTa}_{\text{Large}}$	CE	100	91.1±1.3	81.9±0.4	59.2±2.1
$\text{RoBERTa}_{\text{Large}}$	CE + SCL	100	**92.8±1.3**	**82.5±0.4**	**61.1±3.0**
	p-value		3e-17	1e-20	2e-4
$\text{RoBERTa}_{\text{Large}}$	CE	1000	94.0±0.6	89.2±0.6	81.4±0.2
$\text{RoBERTa}_{\text{Large}}$	CE + SCL	1000	**94.1±0.5**	**89.8±0.4**	**81.5±0.2**
	p-value		0.6	1e-12	0.5

Table 2: Few-shot learning test results on the GLUE benchmark where we have N=20,100,1000 labeled examples for training. Reported results are the mean and the standard deviation of the test accuracies of the top 3 models based on validation accuracy out of 10 random training set samples, along with p-values for each experiment.

Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2020). Supervised contrastive learning for pre-trained language model fine-tuning. arXiv preprint arXiv:2011.01403.

Summary

- **Supervised Learning**
  - Representation learning 대표 기법 Deep Learning
  - 정답이 부여된 충분한 데이터는 높은 예측성능을 기대할 수 있음

- **Self-Supervised Contrastive Learning**
  - 데이터가 부족한 상황을 극복하기 위한 Representation learning
  - 다양한 데이터 패턴을 고려할 수 있는 증강기법을 통한 사전학습 및 조정학습

- **Supervised Contrastive Learning (SCL)**
  - 정답 활용 Supervised Learning & 증강기법 포함 사전학습 및 조정학습 Contrastive Learning
  - **SCL ➔ A fully-supervised version of contrastive learning**
  - 이미지, 텍스트 데이터 예측 성능 향상을 보임

감사합니다.

EOD