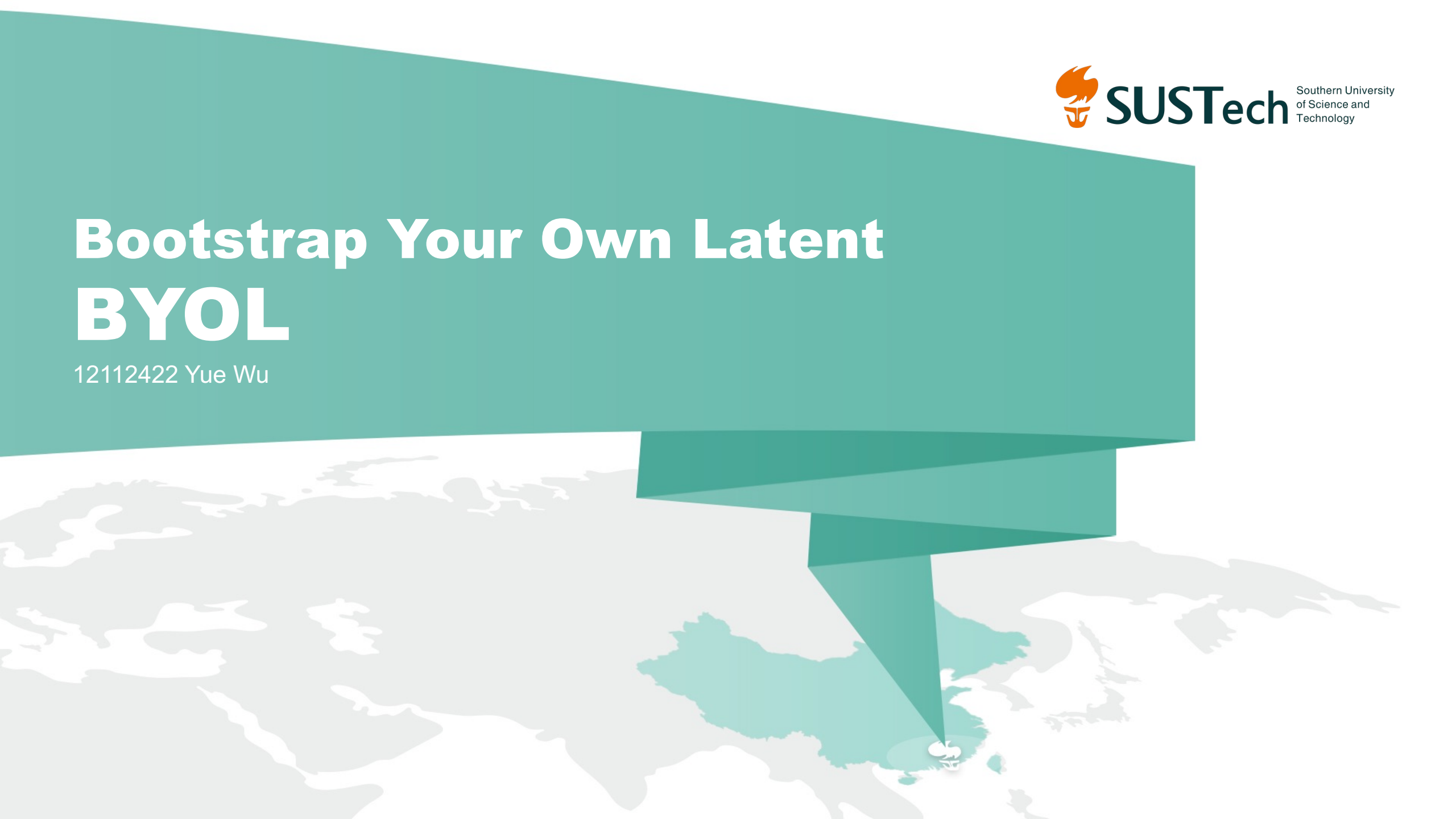
SUSTech Southern University of Science and Technology

Bootstrap Your Own Latent
BYOL

12112422 Yue Wu



SUSTech Southern University of Science and Technology

Abstract

A new approach to self-supervised image representation learning

Two neural network: online & target: online predict target

No negative part

Good perform

SUSTech Southern University of Science and Technology

CATALOGUE

SUSTech Southern University of Science and Technology

01

Background

Background

SUSTech Southern University of Science and Technology

- Learning good image representation is a key challenge in computer vision
- Wide application scenarios
- Can solve few-shot and zero-shot learning problems
- ……

Background - Generative

SUSTech Southern University of Science and Technology

- Generative
  - representation learning build a distribution over data and latent embedding and use the learned embeddings as image representations
  - Auto-encoder of images
  - Adversial learning

Background - Generative

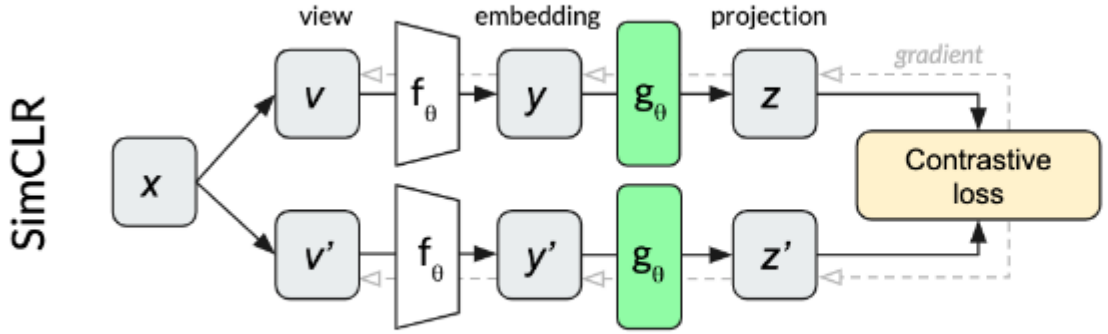
SUSTech Southern University of Science and Technology

- **Generative - Cons**
  - **Operate in pixel space**
  - **Computationally expensive**
  - **high level of detail required for image generation may not be necessary for representation learning**

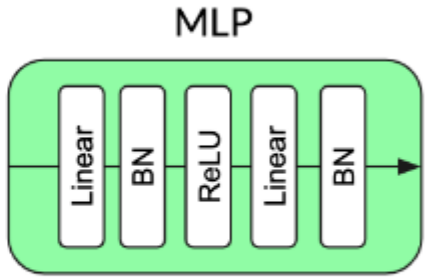
Background - Contrastive

SUSTech Southern University of Science and Technology

- **Contrasive methods : Same closer!**
  - **Positive pairs & Negative pairs & Collapse - trival solution**



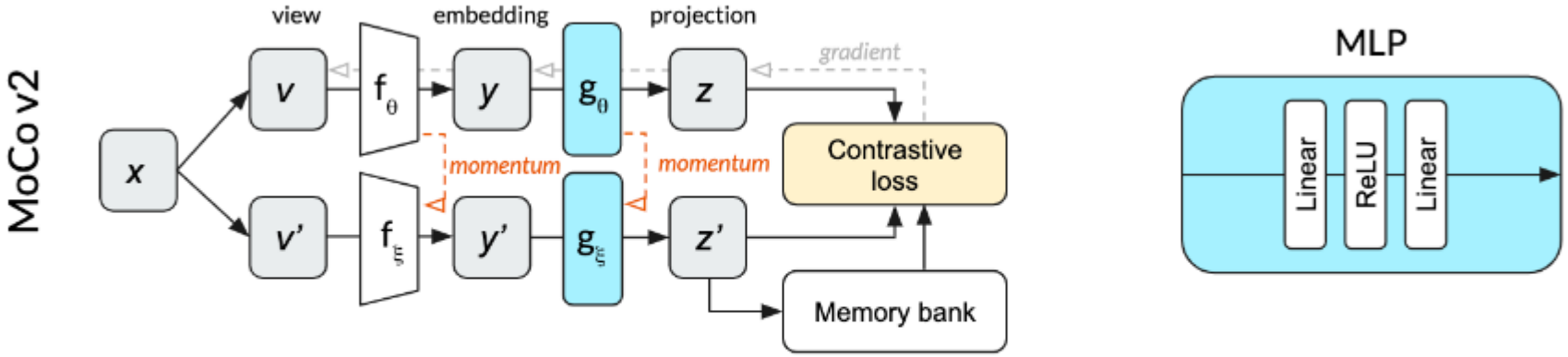
SimCLR architecture.



Background - Contrastive

SUSTech Southern University of Science and Technology

- **Many ways to avoid Collapse**



MoCo v2 architecture. Top row is online encoder, bottom row is momentum encoder.

Background - Contrastive

SUSTech Southern University of Science and Technology

- **Contrastive – Cons**
  - **Contrastive methods often require comparing each example with many other examples to work well --- computation cost**
  - **Whether using negative pairs is necessary ?**

SUSTech Southern University of Science and Technology

02

Method

Method

SUSTech Southern University of Science and Technology

$$\mathcal{L}_{\theta,\xi} \triangleq \left\| \overline{q_\theta}(z_\theta) - \overline{z}'_\xi \right\|_2^2 = 2 - 2 \cdot \frac{\langle q_\theta(z_\theta), z'_\xi \rangle}{\left\| q_\theta(z_\theta) \right\|_2 \cdot \left\| z'_\xi \right\|_2}.$$

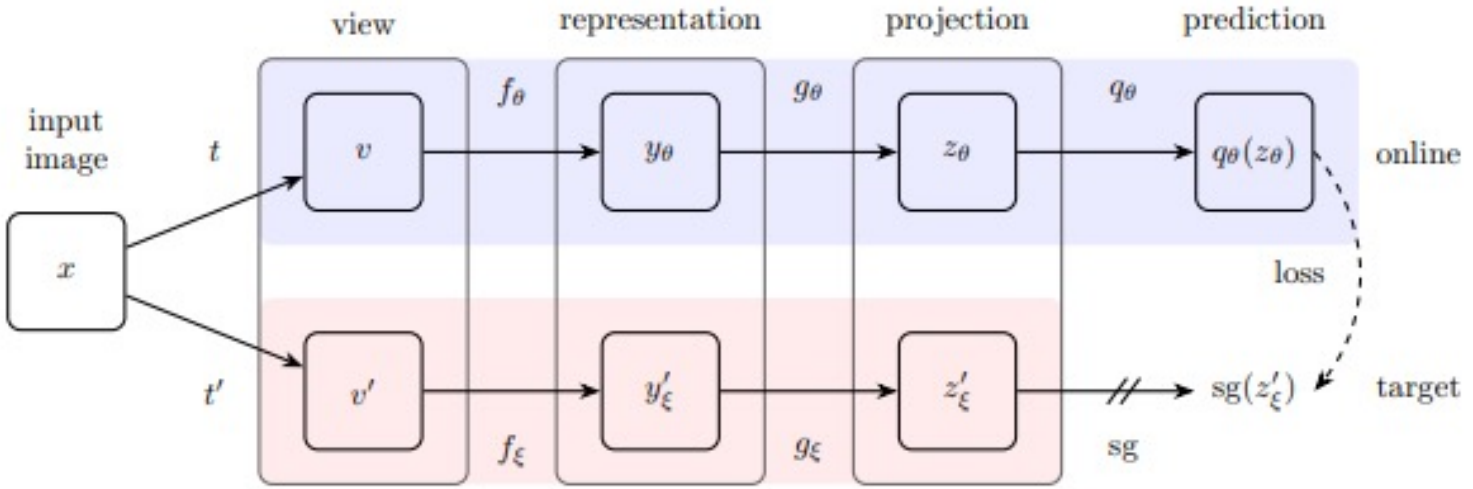
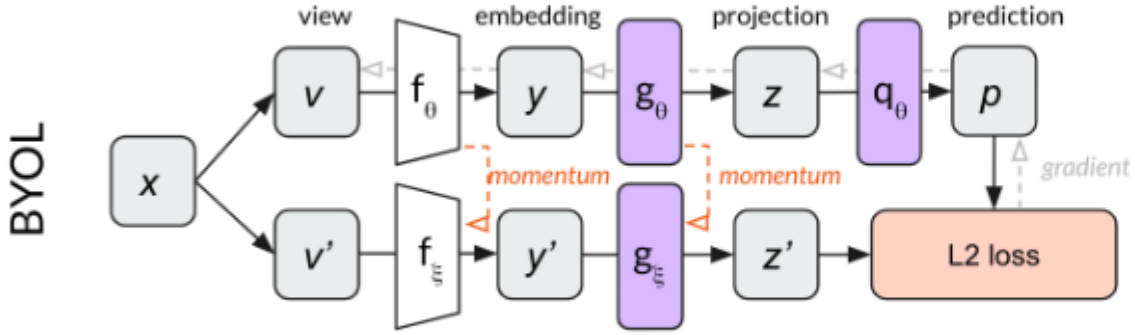


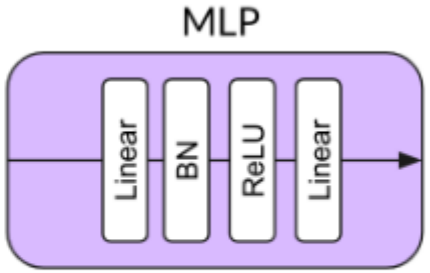
Figure 2: BYOL's architecture. BYOL minimizes a similarity loss between $q_\theta(z_\theta)$ and $\mathrm{sg}(z'_\xi)$, where θ are the trained weights, ξ are an exponential moving average of θ and sg means stop-gradient. At the end of training, everything but f_θ is discarded, and y_θ is used as the image representation.

$$\theta \leftarrow \mathrm{optimizer}\left(\theta, \nabla_\theta \mathcal{L}^{\mathtt{BYOL}}_{\theta,\xi}, \eta\right) \quad \text{and} \quad \xi \leftarrow \tau\xi + (1-\tau)\theta,$$

Method

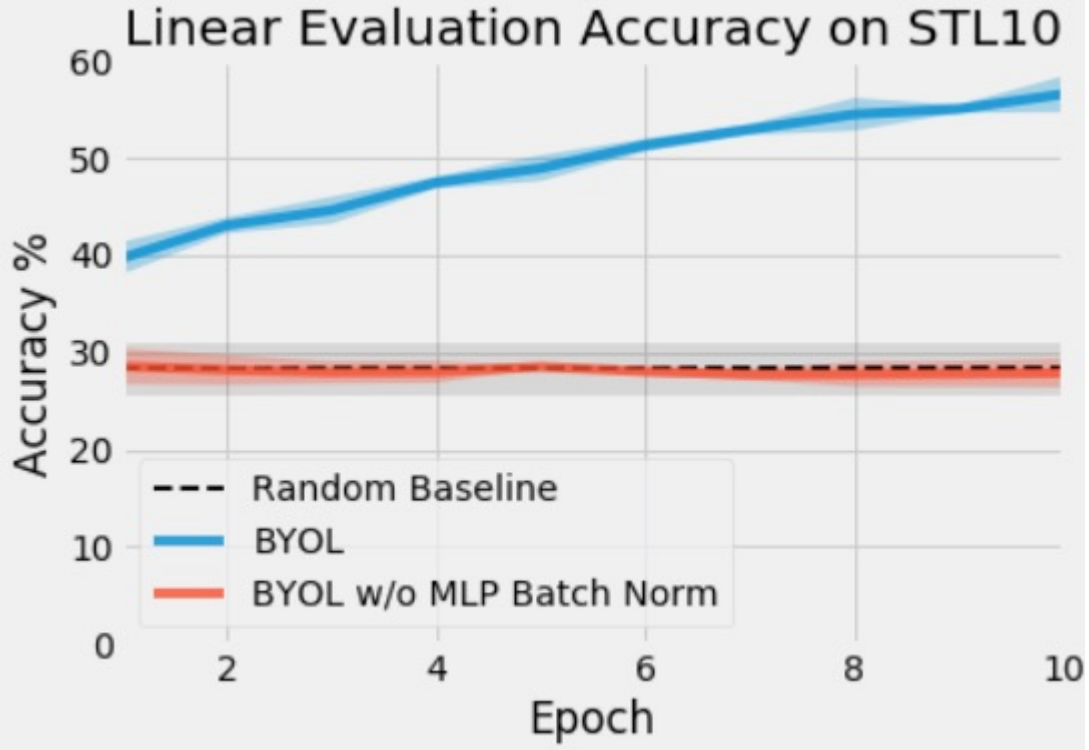


BYOL architecture.



- BN important
- BYOL can learn without explicitly contrasting between multiple different images.
- we only keep the encoder fθ

Method



- **BN important**
- **BN is another Contrastive**
- **A Constraint**
- **Avoid Collapse**

Method

SUSTech Southern University of Science and Technology

Name	Projection MLP Norm	Prediction MLP Norm	Loss Function	Contrastive	Performance [5]
Contrastive Loss	None	None	Cross Entropy	Explicit	44.1
BYOL	Batch Norm	Batch Norm	L2	Implicit	57.7
Projection BN Only	Batch Norm	None	L2	Implicit	55.3
Prediction BN Only	None	Batch Norm	L2	Implicit	48
No Normalization	None	None	L2	None	28.3
Layer Norm	Layer Norm	Layer Norm	L2	None	29.4
Random	—	—	—	None	28.8

- BN important
- Avoid Collapse

- https://generallyintelligent.com/research/2020-08-24-understanding-self-supervised-contrastive-learning/

Method

SUSTech Southern University of Science and Technology

- Why projection：
  - The back layer of the neural network - Get more generalized features
  - High-dimensional mapping to low-latitude
  - Cosine similarity

Method

SUSTech Southern University of Science and Technology

- Distillation？
  - Self-distillation：Self-distillation is a model compression technique that uses a larger teacher model to guide the training of a smaller student model.
  - two views of one image via augmentation

SUSTech Southern University of Science and Technology

03

Experiment

Experiment

SUSTech Southern University of Science and Technology

- **Linear evaluation**
  - **Training a linear classifier on top of the frozen representation**

Method	Top-1	Top-5
Local Agg.	60.2	-
PIRL [35]	63.6	-
CPC v2 [32]	63.8	85.3
CMC [11]	66.2	87.0
SimCLR [8]	69.3	89.0
MoCo v2 [37]	71.1	-
InfoMin Aug. [12]	73.0	91.1
BYOL (ours)	**74.3**	**91.6**

(a) ResNet-50 encoder.

Method	Architecture	Param.	Top-1	Top-5
SimCLR [8]	ResNet-50 (2×)	94M	74.2	92.0
CMC [11]	ResNet-50 (2×)	94M	70.6	89.7
BYOL (ours)	ResNet-50 (2×)	94M	77.4	93.6
CPC v2 [32]	ResNet-161	305M	71.5	90.1
MoCo [9]	ResNet-50 (4×)	375M	68.6	-
SimCLR [8]	ResNet-50 (4×)	375M	76.5	93.2
BYOL (ours)	ResNet-50 (4×)	375M	78.6	94.2
BYOL (ours)	ResNet-200 (2×)	250M	**79.6**	**94.8**

(b) Other ResNet encoder architectures.

Experiment

SUSTech Southern University of Science and Technology

- **Semi-supervised training**
  - **finetuning BYOL's representation on a classification task with a small subset of train set**

Method	Top-1		Top-5	
	1%	10%	1%	10%
Supervised [77]	25.4	56.4	48.4	80.4
InstDisc	-	-	39.2	77.4
PIRL [35]	-	-	57.2	83.8
SimCLR [8]	48.3	65.6	75.5	87.8
BYOL (ours)	**53.2**	**68.8**	**78.4**	**89.0**

(a) ResNet-50 encoder.

Method	Architecture	Param.	Top-1		Top-5	
			1%	10%	1%	10%
CPC v2 [32]	ResNet-161	305M	-	-	77.9	91.2
SimCLR [8]	ResNet-50 (2×)	94M	58.5	71.7	83.0	91.2
BYOL (ours)	ResNet-50 (2×)	94M	62.2	73.5	84.1	91.7
SimCLR [8]	ResNet-50 (4×)	375M	63.0	74.4	85.8	92.6
BYOL (ours)	ResNet-50 (4×)	375M	69.1	75.7	87.9	92.5
BYOL (ours)	ResNet-200 (2×)	250M	**71.2**	**77.7**	**89.5**	**93.7**

(b) Other ResNet encoder architectures.

Experiment

SUSTech Southern University of Science and Technology

- **Transfer to other classification tasks**

Method	Food101	CIFAR10	CIFAR100	Birdsnap	SUN397	Cars	Aircraft	VOC2007	DTD	Pets	Caltech-101	Flowers
Linear evaluation:												
BYOL (ours)	**75.3**	91.3	**78.4**	**57.2**	**62.2**	**67.8**	60.6	82.5	75.5	90.4	94.2	**96.1**
SimCLR (repro)	72.8	90.5	74.4	42.4	60.6	49.3	49.8	81.4	**75.7**	84.6	89.3	92.6
SimCLR [8]	68.4	90.6	71.6	37.4	58.8	50.3	50.3	80.5	74.5	83.6	90.3	91.2
Supervised-IN [8]	72.3	**93.6**	78.3	53.7	61.9	66.7	**61.0**	**82.8**	74.9	**91.5**	**94.5**	94.7
Fine-tuned:												
BYOL (ours)	**88.5**	**97.8**	86.1	**76.3**	63.7	91.6	**88.1**	**85.4**	**76.2**	91.7	**93.8**	97.0
SimCLR (repro)	87.5	97.4	85.3	75.0	63.9	91.4	87.6	84.5	75.4	89.4	91.7	96.6
SimCLR [8]	88.2	97.7	85.9	75.9	63.5	91.3	88.1	84.1	73.2	89.2	92.1	97.0
Supervised-IN [8]	88.3	97.5	**86.4**	75.8	**64.3**	**92.1**	86.0	85.0	74.6	**92.1**	93.3	**97.6**
Random init [8]	86.9	95.9	80.2	76.1	53.6	91.4	85.9	67.3	64.8	81.5	72.6	92.0

Table 3: Transfer learning results from ImageNet (IN) with the standard ResNet-50 architecture.

Experiment

SUSTech Southern University of Science and Technology

- **Transfer to other vision tasks**

Method	AP_{50}	mIoU
Supervised-IN [9]	74.4	74.4
MoCo [9]	74.9	72.5
SimCLR (repro)	75.2	75.2
BYOL (ours)	**77.5**	**76.3**

(a) Transfer results in semantic segmentation and object detection.

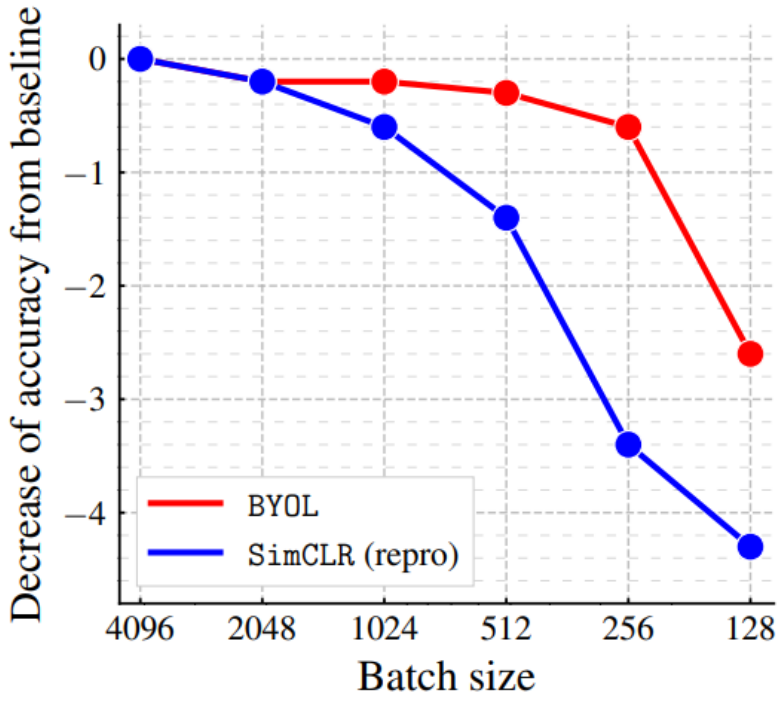
Method	Higher better			Lower better	
	pct.< 1.25	pct.$< 1.25^2$	pct.$< 1.25^3$	rms	rel
Supervised-IN [83]	81.1	95.3	98.8	0.573	**0.127**
SimCLR (repro)	83.3	96.5	99.1	0.557	0.134
BYOL (ours)	**84.6**	**96.7**	**99.1**	**0.541**	0.129

(b) Transfer results on NYU v2 depth estimation.

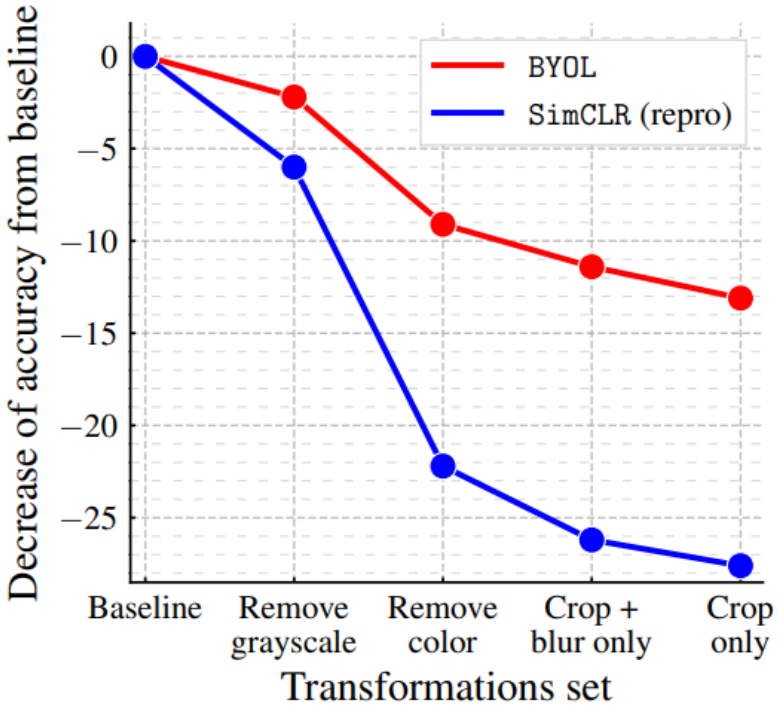
Experiment

SUSTech Southern University of Science and Technology

- **Batch size & Image augmentation**



(a) Impact of batch size



(b) Impact of progressively removing transformations

SUSTech Southern University of Science and Technology

04

Conclusion

Conclusion

SUSTech Southern University of Science and Technology

- No negative pairs
- Faster & Good perform
- Robust

- Good experiment design

SUSTech Southern University of Science and Technology

Q&A

THANK YOU