# Contrastive Learning with Synthetic Positive

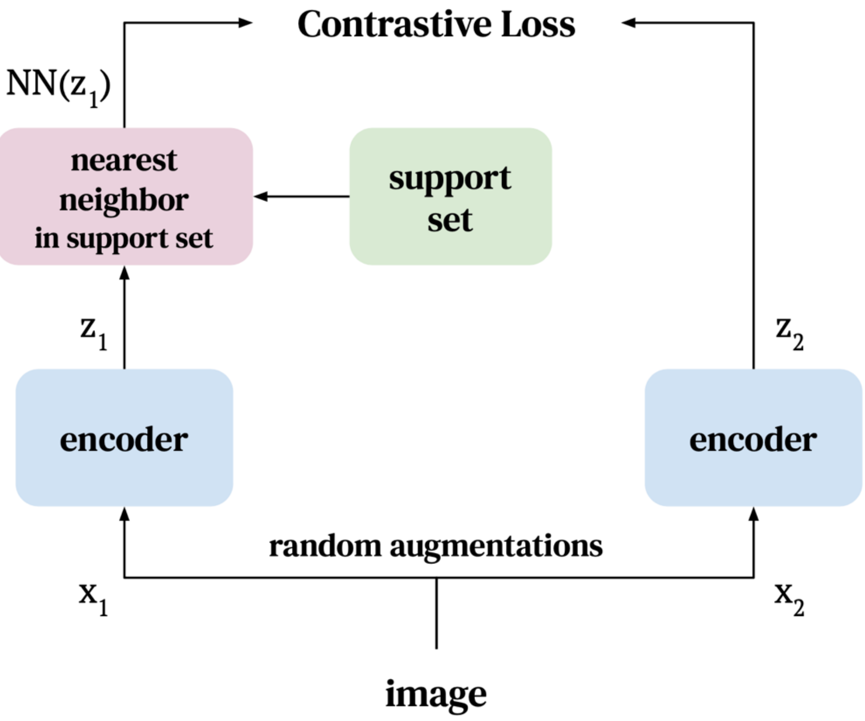
Dewen Zeng[1], Xinrong Hu[1], Yawen Wu[1], Xiaowei Xu[2] and Yiyu Shi[1]

[1] University of Notre Dame

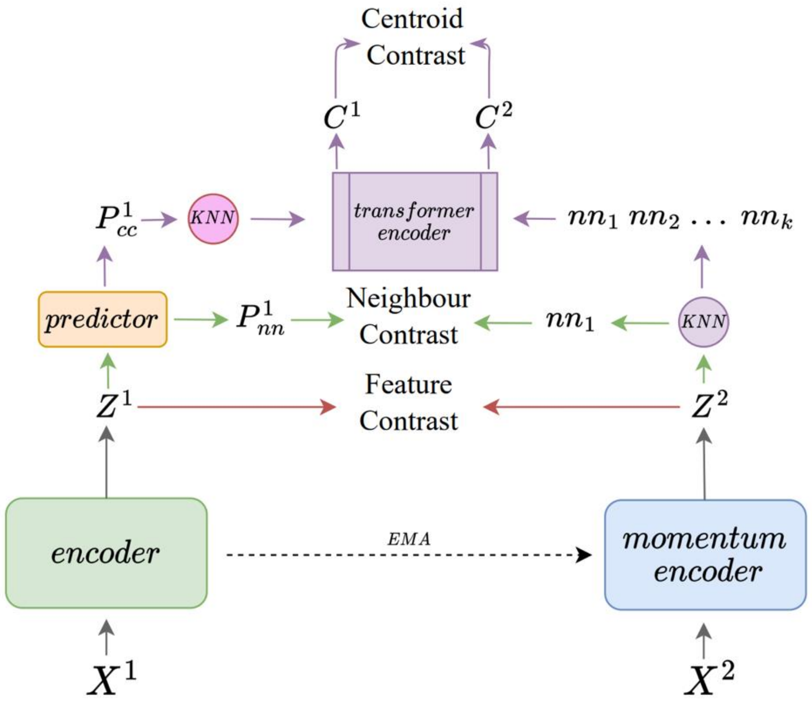
[2] Guangdong Provincial People's Hospital

# Contrastive Learning Benefits from Positives Beyond Simple Data augmentation





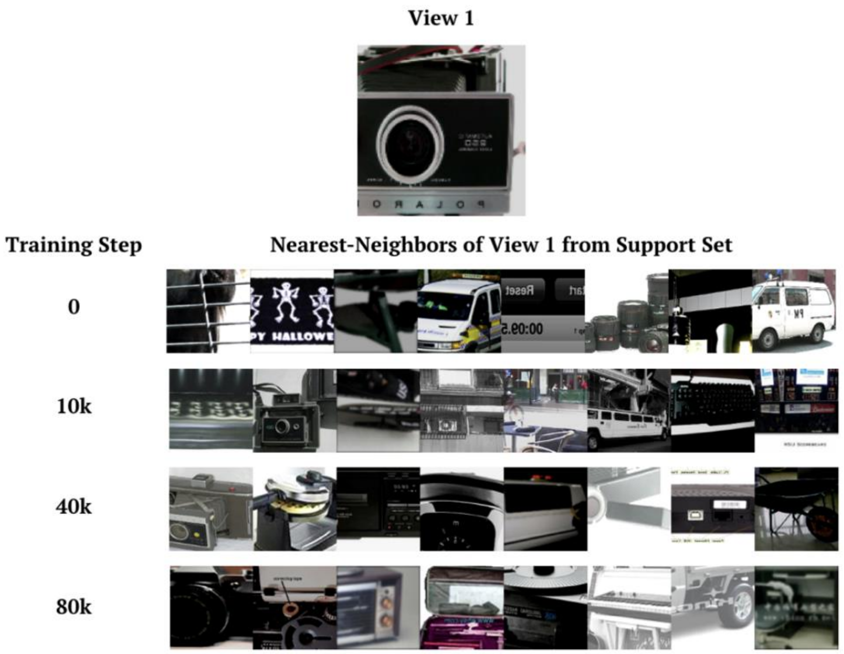

NNCLR (ICCV'21), using the nearest neighbor as positive




All4One (ICCV'23), using centroid of multiple nearest neighbors

# Limitations of Nearest Neighbors

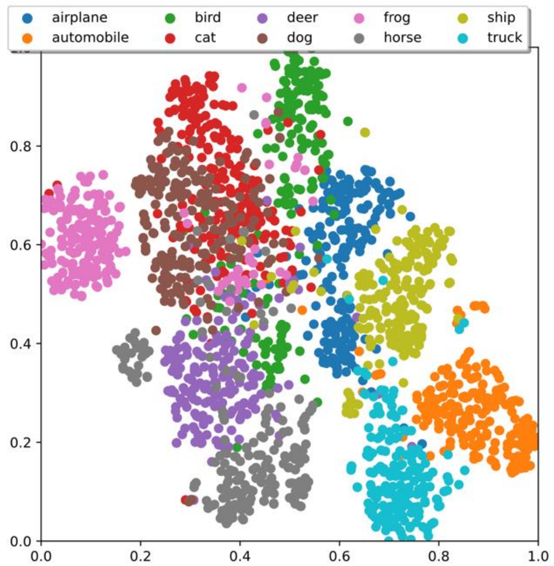
- False positives in the early training stage
- "Easy" positives because the model can already generate close embeddings for these examples



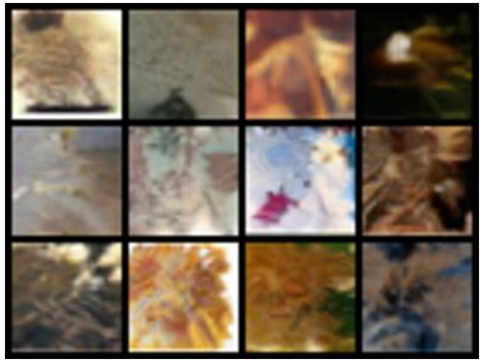

NNs at different training stages

# Motivation

- Unconditional diffusion model already learns good semantic embedding
- Semantic and background features are decoupled in different layers
- Control the semantics of the generated images by modifying the latent embedding in random sampling

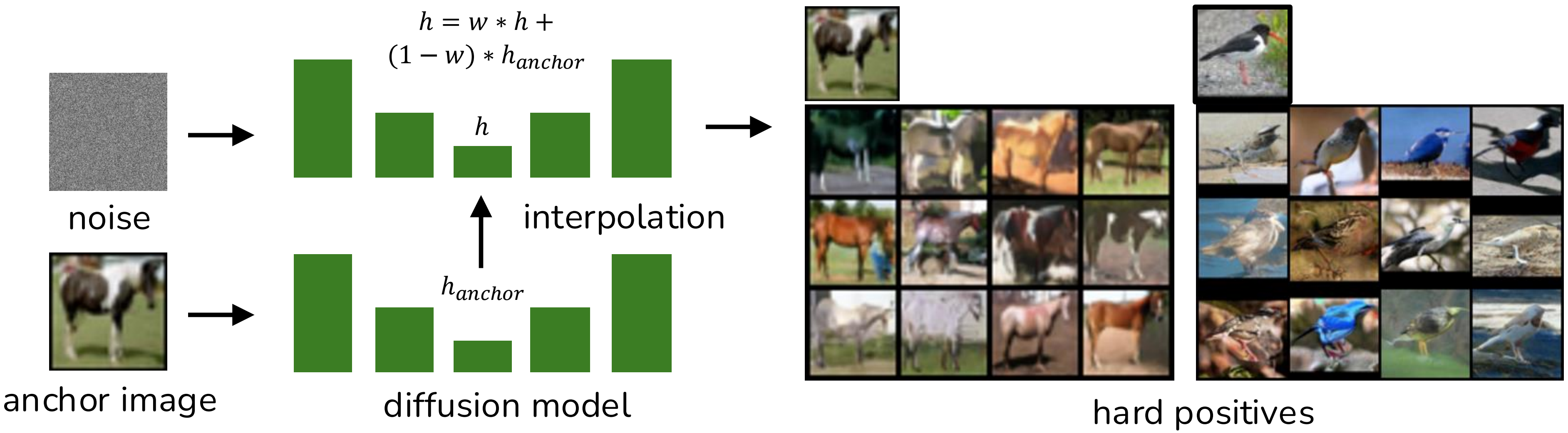



t-SNE plot from latent features of diffusion model trained on Cifar10



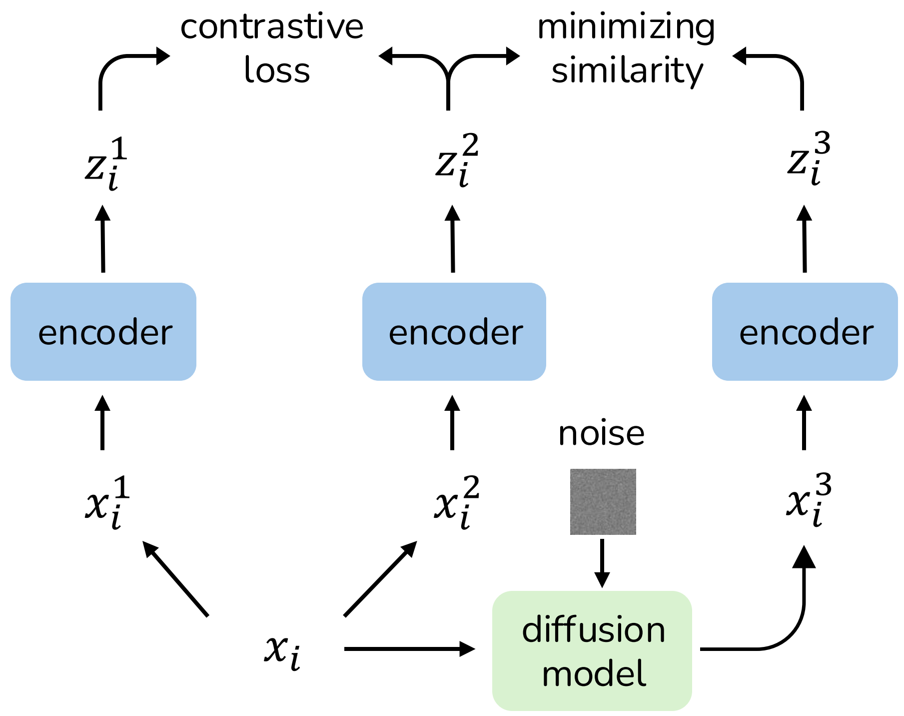
Setting latent features to 0 during random sampling (Semantic is missing)

# Hard Positive Generation by Feature Interpolation




$$h = w * h + (1 - w) * h_{anchor}$$

# Contrastive Learning with Synthetic Positives (CLSP)




$$\mathcal{L}_{clsp} = \mathcal{L}_{simclr} + \lambda \sum_{i \in [1,N]} \left\| z_i^2 - z_i^3 \right\|_2^2$$

# Linear Evaluation Results

EUROPEAN CONFERENCE ON COMPUTER VISION

| | Epoch | CIFAR10 Linear | CIFAR10 kNN | CIFAR100 Linear | CIFAR100 kNN |
|---|---|---|---|---|---|
| Barlow Twins | 1000 | 92.10 | 90.60 | 70.90 | 66.40 |
| BYOL | 1000 | 92.58 | 90.78 | 70.46 | 67.22 |
| DINO | 1000 | 89.52 | 88.40 | 66.76 | 62.07 |
| MoCoV2 | 1000 | 92.94 | 91.45 | 69.89 | 66.91 |
| MoCoV3 | 1000 | 93.10 | 91.00 | 68.83 | 65.07 |
| NNCLR | 1000 | 91.88 | 90.61 | 69.62 | 65.98 |
| SimCLR | 1000 | 91.47 | 88.57 | 65.78 | 62.69 |
| SupCon | 1000 | 93.82 | 92.60 | 70.38 | 69.28 |
| SwAV | 1000 | 89.17 | 87.42 | 64.88 | 60.65 |
| VIbCReg | 1000 | 91.18 | 89.25 | 67.37 | 63.16 |
| VICReg | 1000 | 92.07 | 89.97 | 68.54 | 65.09 |
| All4One | 1000 | 93.24 | 91.38 | **72.17** | 67.83 |
| DCL [53] | 200 | 85.90 | 84.40 | 58.90 | 54.30 |
| CL-GAN [48] | 300 | 92.94 | - | 67.41 | - |
| SimCLR-FT [58] | 1000 | 91.77 | 91.45 | 66.43 | 65.07 |
| MoCoV2-FT [58] | 1000 | 92.58 | 92.60 | 69.40 | 68.79 |
| CLSP-noaug | 1000 | 92.86 | 92.46 | 54.46 | 53.53 |
| SimCLR-aug | 1000 | 91.96 | 91.44 | 68.30 | 66.75 |
| MoCoV2-aug | 1000 | 93.04 | 92.82 | 69.13 | 68.21 |
| CLSP-SimCLR | 300 | 93.10 | 92.44 | 67.94 | 66.04 |
| CLSP-SimCLR | 1000 | 94.37 | 93.59 | 72.01 | **70.19** |
| CLSP-MoCoV2 | 1000 | **94.41** | **93.90** | 71.76 | 69.75 |

# Transfer Learning Results

Pre-trained on Cifar10, Cifar100, and STL10

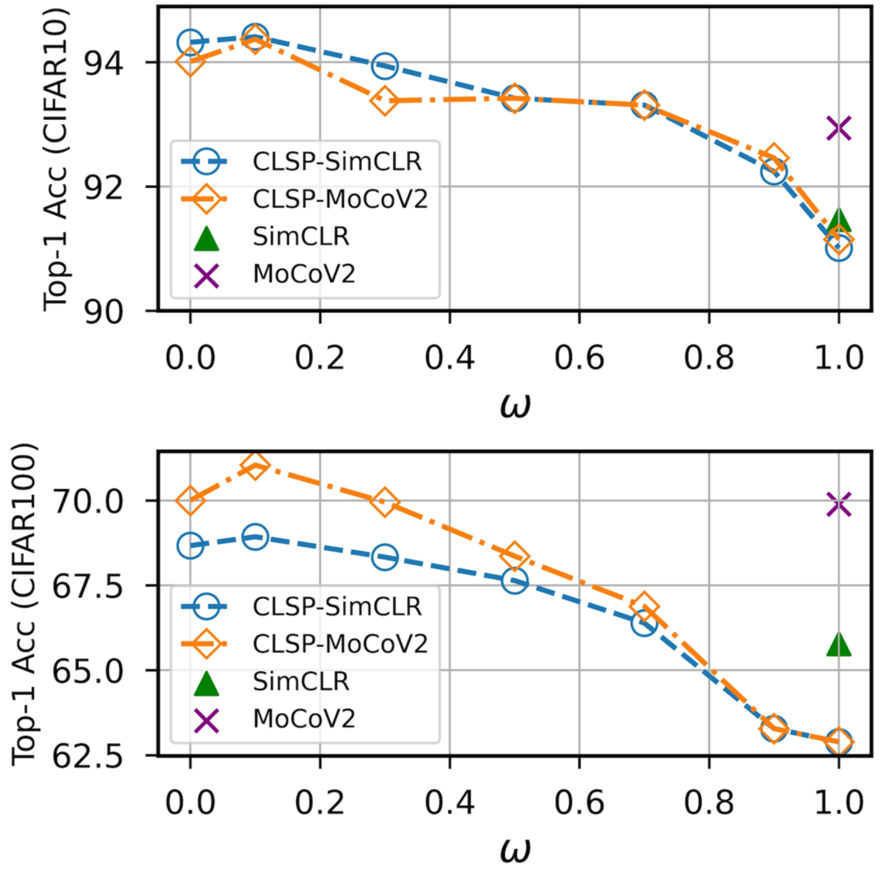
| Source | CIFAR10 | CIFAR100 | STL10 | | | | | |
|---|---|---|---|---|---|---|---|---|
| Target | CIFAR100 | CIFAR10 | CIFAR10 | CIFAR100 | Pets | Flowers | DTD | Caltech101 |
| SimCLR | 47.97 | 73.81 | 81.58 | 52.79 | 53.97 | 53.23 | 45.74 | 73.42 |
| MoCoV2 | 54.90 | 78.57 | 83.39 | 55.41 | 56.37 | 54.32 | 46.23 | 76.13 |
| BYOL | 53.62 | 78.69 | 82.88 | 55.34 | 55.93 | 57.81 | 45.59 | 78.57 |
| NNCLR | 53.27 | 78.79 | 81.60 | 53.94 | 56.31 | **61.93** | 43.51 | 78.24 |
| All4One | 56.17 | 79.38 | 83.33 | 55.49 | 56.65 | 57.52 | 46.13 | 76.62 |
| SimCLR-FT | 55.36 | 80.89 | 81.62 | 53.89 | 54.70 | 51.75 | 42.71 | 73.58 |
| SimCLR-aug | 54.42 | 79.78 | 81.83 | 54.57 | 56.66 | 56.14 | 45.37 | 75.16 |
| CLSP-SimCLR | **61.64** | **83.14** | 84.98 | 56.83 | **56.75** | 57.52 | **46.76** | **80.01** |
| CLSP-MoCoV2 | 61.19 | 83.13 | **86.58** | **57.97** | 55.82 | 50.20 | 44.68 | 79.42 |

Pre-trained on ImageNet100

| | CIFAR10 | CIFAR100 | Pets | Flowers | DTD | Caltech101 | Food101 | Cars |
|---|---|---|---|---|---|---|---|---|
| All4One | 86.58 | 65.42 | 71.25 | 72.21 | 59.04 | 84.93 | **63.30** | **33.14** |
| CLSP-SimCLR | **87.36** | **66.50** | **73.97** | **72.65** | **59.95** | **85.24** | 59.87 | 32.81 |

# Impact of Feature Interpolation Weight

- $h = w * h + (1 - w) * h_{anchor}$
- $\omega$ controls the degree of randomness in the semantic during sampling

# Positive Candidate Set Size and Loss Weight

Accuracy changes with different candidate set size

| $k$ | 1 | 2 | 4 | 8 | 16 | 32 |
|---|---|---|---|---|---|---|
| CIFAR10 | 92.24 | 93.29 | 93.95 | 94.37 | 94.38 | 94.33 |
| CIFAR100 | 65.84 | 68.33 | 70.65 | 72.01 | 71.92 | 72.04 |

$$\mathcal{L}_{clsp} = \mathcal{L}_{simclr} + \lambda \sum_{i \in [1,N]} \left\| z_i^2 - z_i^3 \right\|_2^2$$

Accuracy changes with different $\lambda$

| $\lambda$ | 0.01 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 | 1.6 |
|---|---|---|---|---|---|---|---|---|---|
| CIFAR10 | 91.80 | 91.97 | 93.16 | 93.84 | 94.23 | 94.37 | 94.27 | 94.21 | 94.59 |
| CIFAR100 | 65.19 | 68.08 | 70.68 | 71.36 | 71.81 | 72.01 | 70.44 | 70.00 | 69.07 |

# Thank you!

Code: https://github.com/dewenzeng/clsp