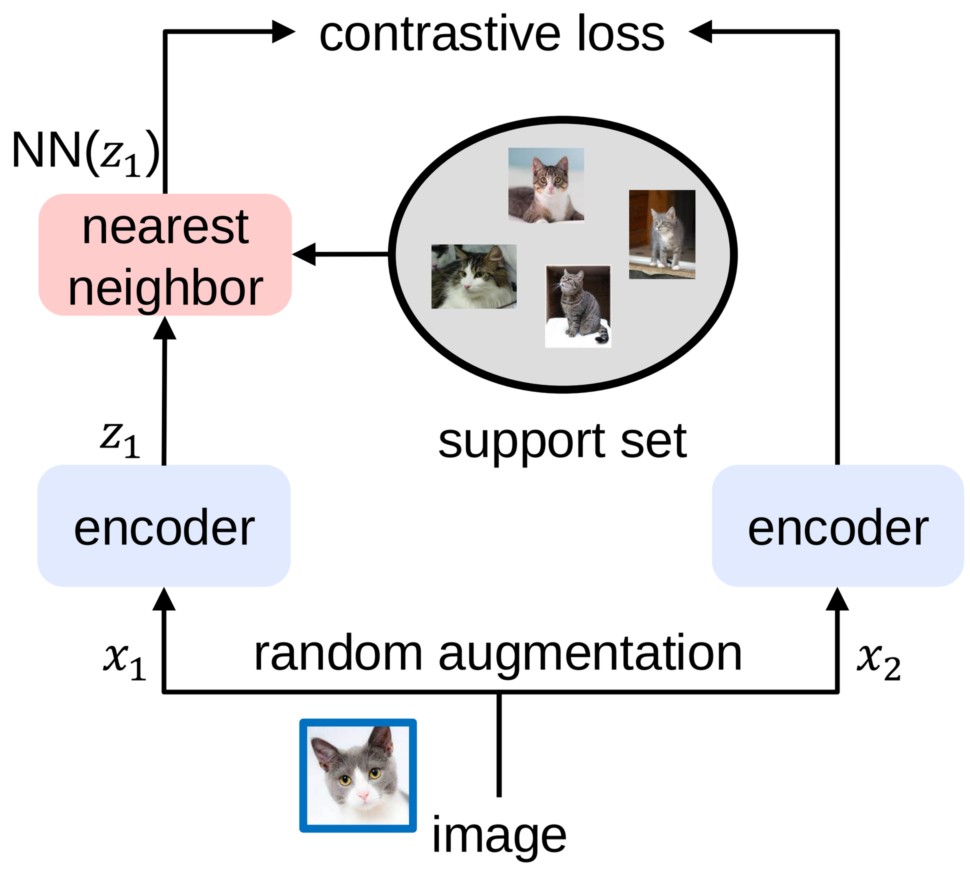
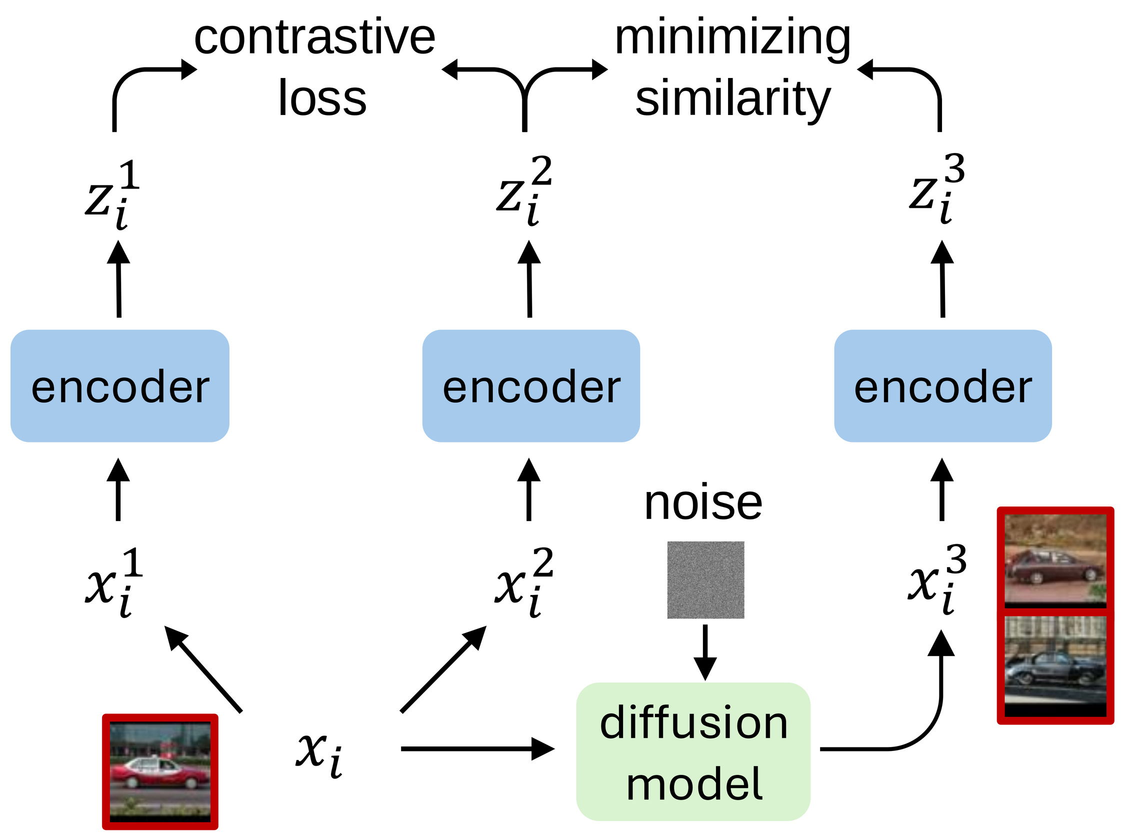
Contrastive Learning with Synthetic Positives

Dewen Zeng[1], Yawen Wu[1], Xinrong Hu[1], Xiaowei Xu[2] and Yiyu Shi[1]
[1] University of Notre Dame, [2] Guangdong Provincial People's Hospital

Contrastive learning with positives beyond simple data augmentation



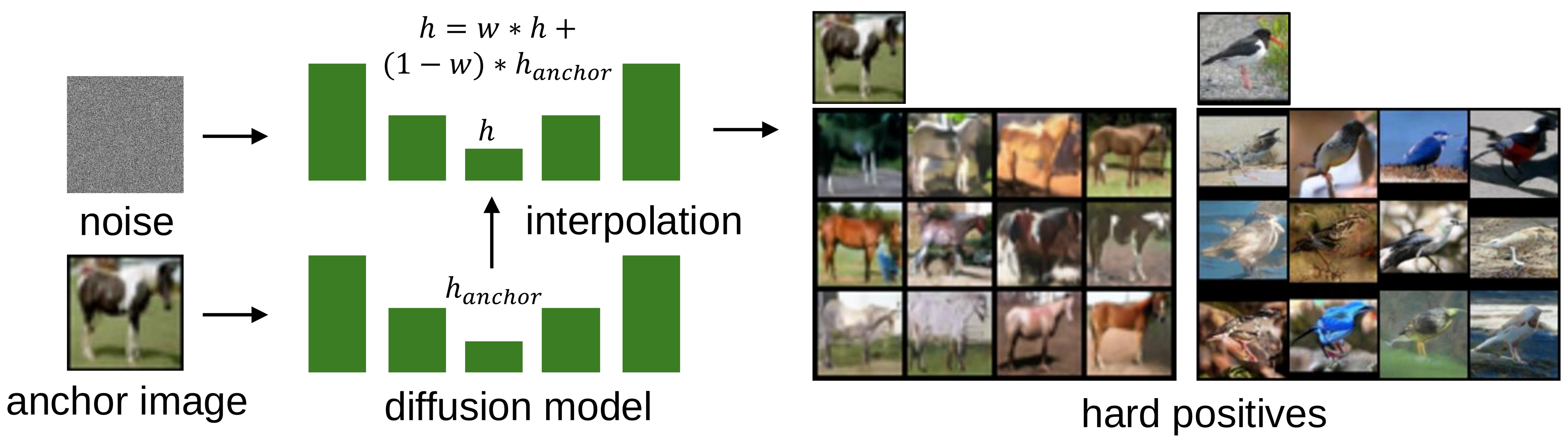
(a) Using nearest neighbor identified on-the-fly [1]



(b) Using synthetic positives generated by unconditional diffusion model (Ours)

- Contrastive learning benefits from additional positives
- Latent embedding contains semantic information in **unconditional** diffusion model, which enables embedding-guided diffusion sampling
- The synthetic positives generated by unconditional diffusion model are diverse and "**harder**"

Hard positive generation with feature interpolation



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

Experiments

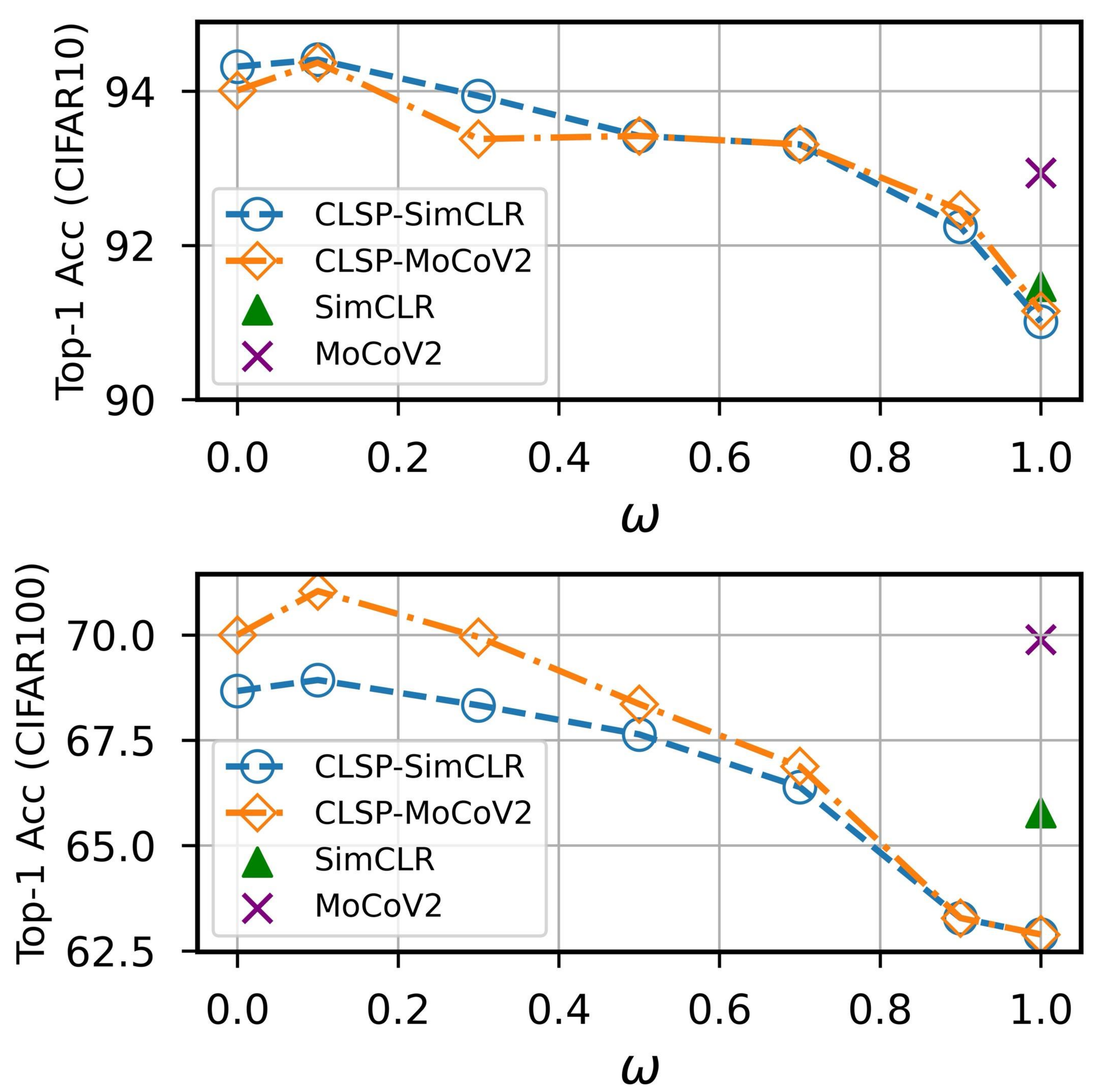
- *linear evaluation*

Method	Cifar10	Cifar100	STL10	ImageNet100
SimCLR	91.47	65.78	90.63	75.75
MoCoV2	92.94	69.89	92.20	78.00
NNCLR	91.88	69.62	90.78	77.43
All4One	93.24	**72.17**	92.21	79.09
CLSP-SimCLR	94.37	72.01	**93.74**	**79.62**
CLSP-MoCoV2	**94.41**	71.76	93.69	79.11

- *transfer to downstream (pre-trained on STL-10)*

Method	Cifar10	Cifar100	Pets	Flowers	DTD	Caltech101
NNCLR	81.60	53.94	56.31	**61.93**	43.51	78.24
All4One	83.33	55.49	56.65	57.52	46.13	76.62
CLSP-SimCLR	84.98	56.83	**56.75**	57.52	**46.76**	**80.01**
CLSP-MoCoV2	**86.58**	**57.97**	55.82	50.20	44.68	79.42

- *ablation on the feature interpolation weight ω*



[1] With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. ICCV'21