

7/7 meeting

應名宥

training cost

- asymmetric loss structure
 - ~ 7m 05s (4 kimg)
- asymmetric loss structure (preprocess ResNet features)
 - ~ 6m 30s (4 kimg)
- symmetric loss structure
 - ~ 7m 23s (4 kimg)
- symmetric loss structure (preprocess ResNet features)
 - ~ 6m 57s(4 kimg)

tuning hyper-parameters

1. structure parameter
2. image-text parameter
3. image-image parameter
4. heterologous parameter

structure parameter

- structure parameters ratio
 - Discriminator : 1
 - CLIP : 2
 - ResNet : 1

image-text parameter

- expected tuning range
 - [0.5, 1.25, 2.5, 3.75, 5]

image-image parameter

- expected tuning range
 - [0.5, 1.25, 2.5, 3.75, 5]

heterologous parameter

- expected tuning range
 - [0.1, 0.2, 0.5, 1, 2]

example (discriminator)

$$w_D^{ii} = S_D \cdot C_{ii} \cdot 1$$

$$w_D^{ii'} = S_D \cdot C_{ii} \cdot h$$

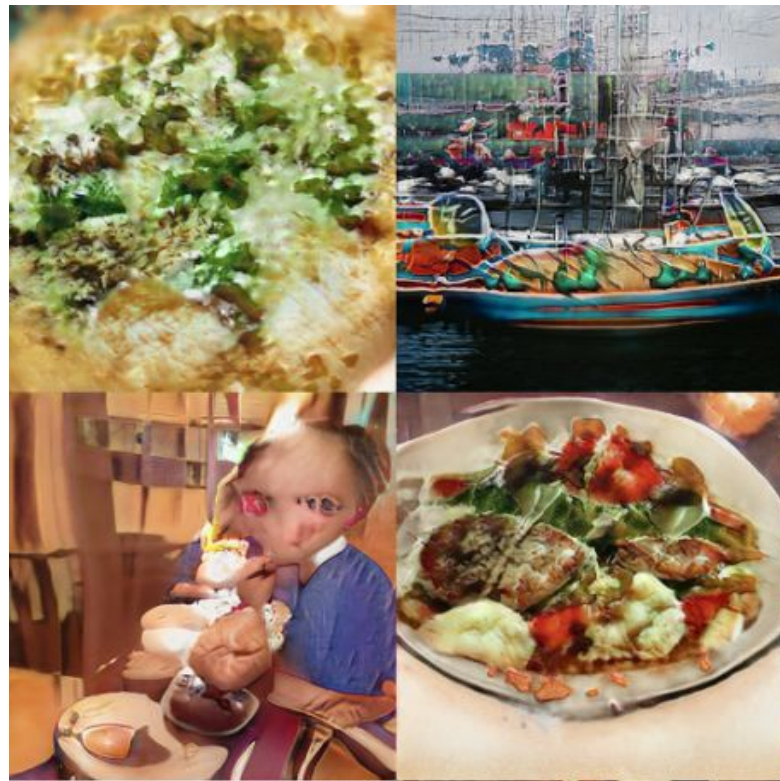
$$w_D^{it} = S_D \cdot C_{it} \cdot 1$$

$$w_D^{it'} = S_D \cdot C_{it} \cdot h$$

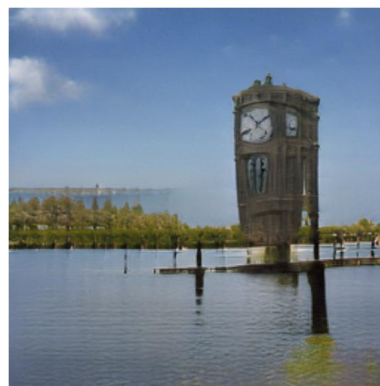
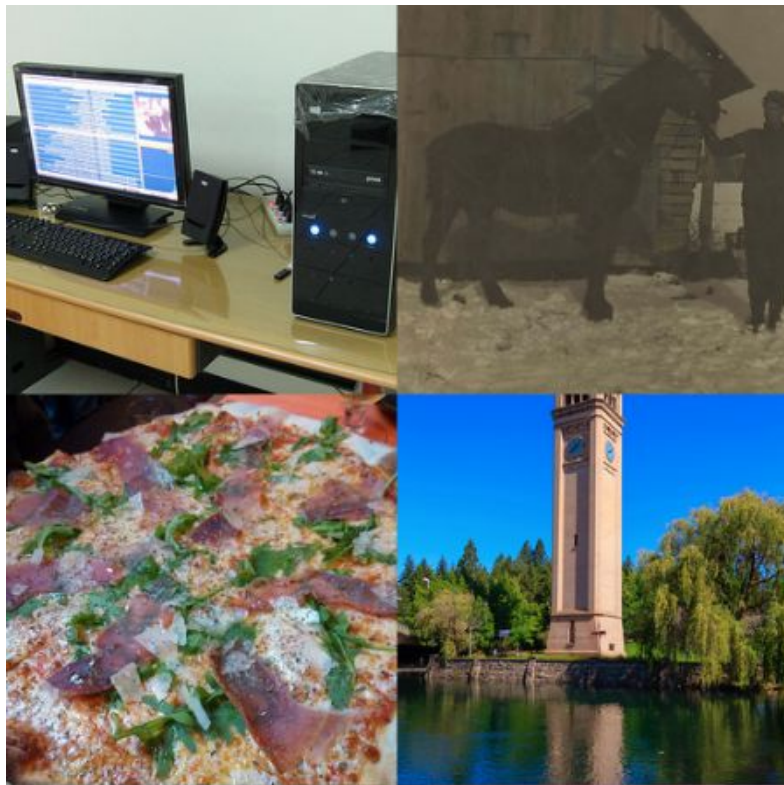
pretrain model



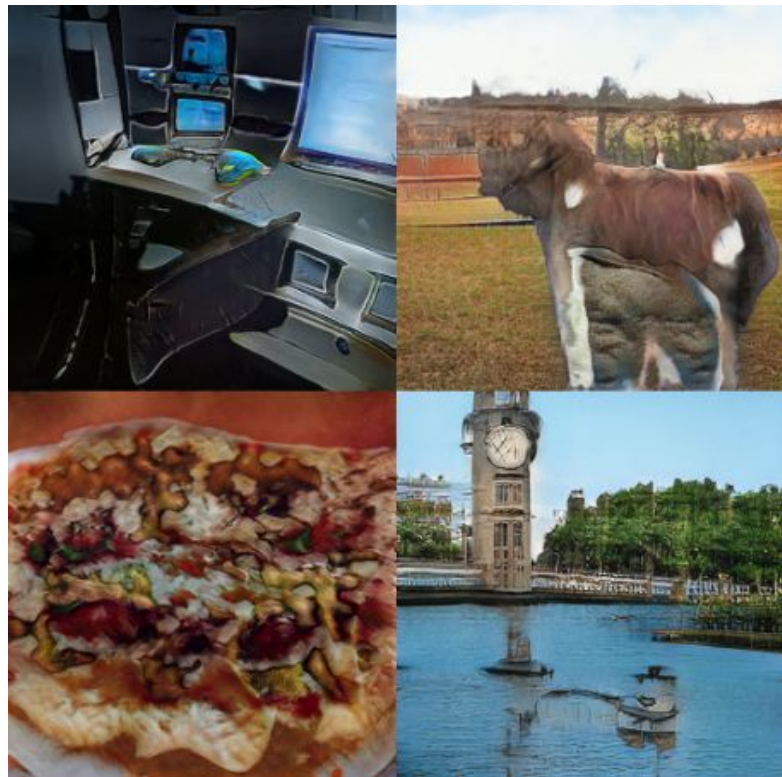
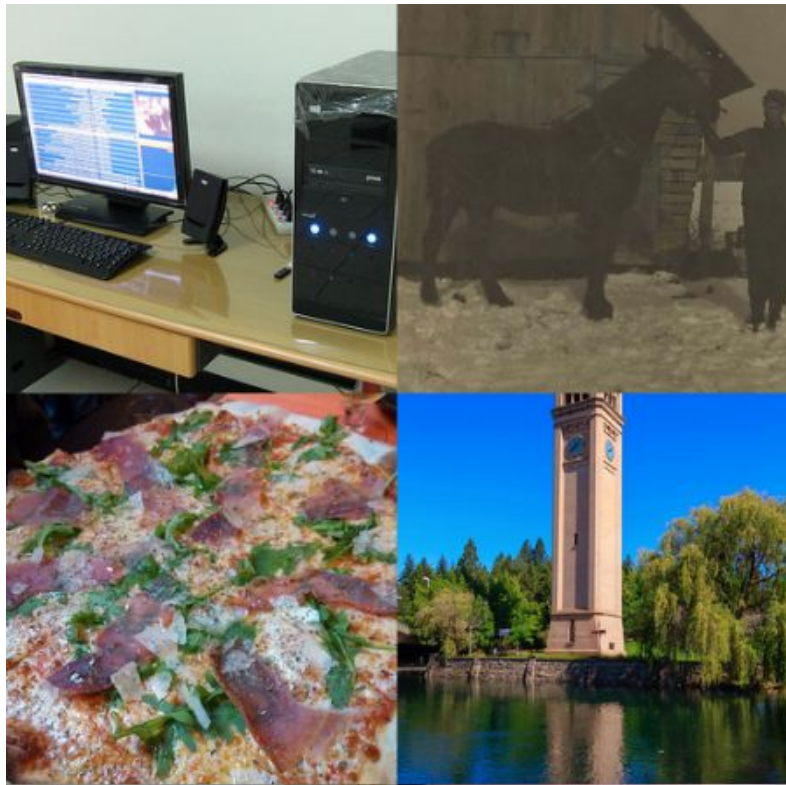
training result



pretrain model



training result



calculate FID

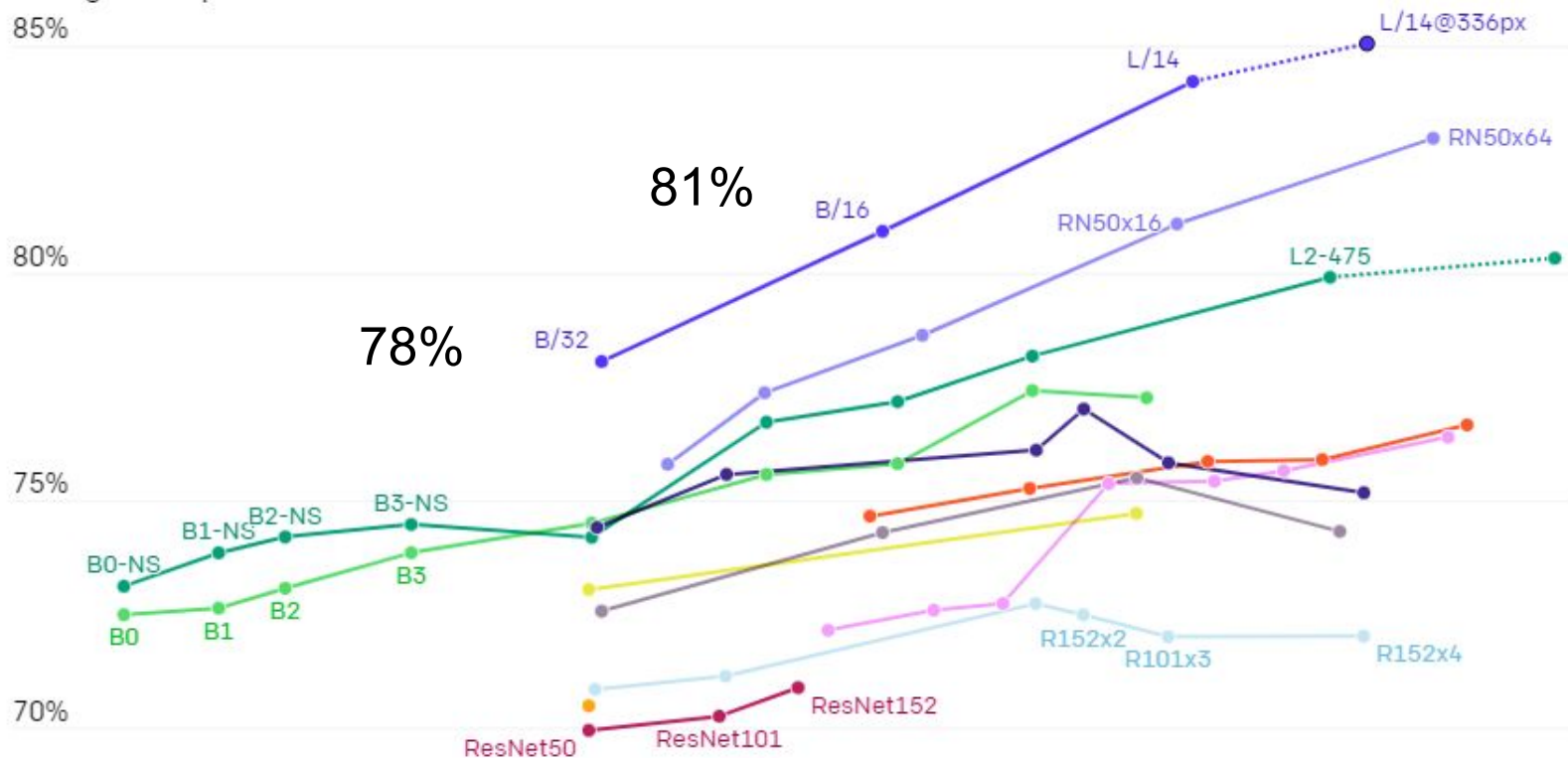
- origin (VIT-B32)
 - 20.741864805036187 (2000 kimg)
 - 16.797841481759033 (3000 kimg)
- modified (VIT-L14@336px)
 - 17.078982249532714 (2300 kimg)

clip model

Average linear probe score across 27 datasets

85%

85%



some advice

- tune hyper-parameters from 1~50 (itd, itc... etc)
- do not config ada to compare with previous methods fairly
- Increase batch size for contrastive learning may lead to performance improvement

experiments

- test old CLIP model (ViT-B16)
- test symmetric loss structure
- large model capacity

study

- SimCLR (google research)
 - [A Simple Framework for Contrastive Learning of Visual Representations \(mlr.press\)](#)
 - break through in image classification (resnet)
- MSG-GAN
 - [CVPR 2020 Open Access Repository \(thecvf.com\)](#)
 - stylegan2 generator structure (base)
- Weakly Supervised Contrastive Learning
 - [ICCV 2021 Open Access Repository \(thecvf.com\)](#)
 - supervised contrastive learning framework