5/5 進度報告

應名宥

text - image similarity

score : 35.16

score : 40.59

add image realistic score

only clip rating

clip rating and vgg19 score

NSGA2

rand 1000 sample

population : 100
#offspring : 25
generation : 50

win

a cat is lay on the table.

result of failure

151.25

a street scene with a double-decker bus on the road.

166.59

analysis

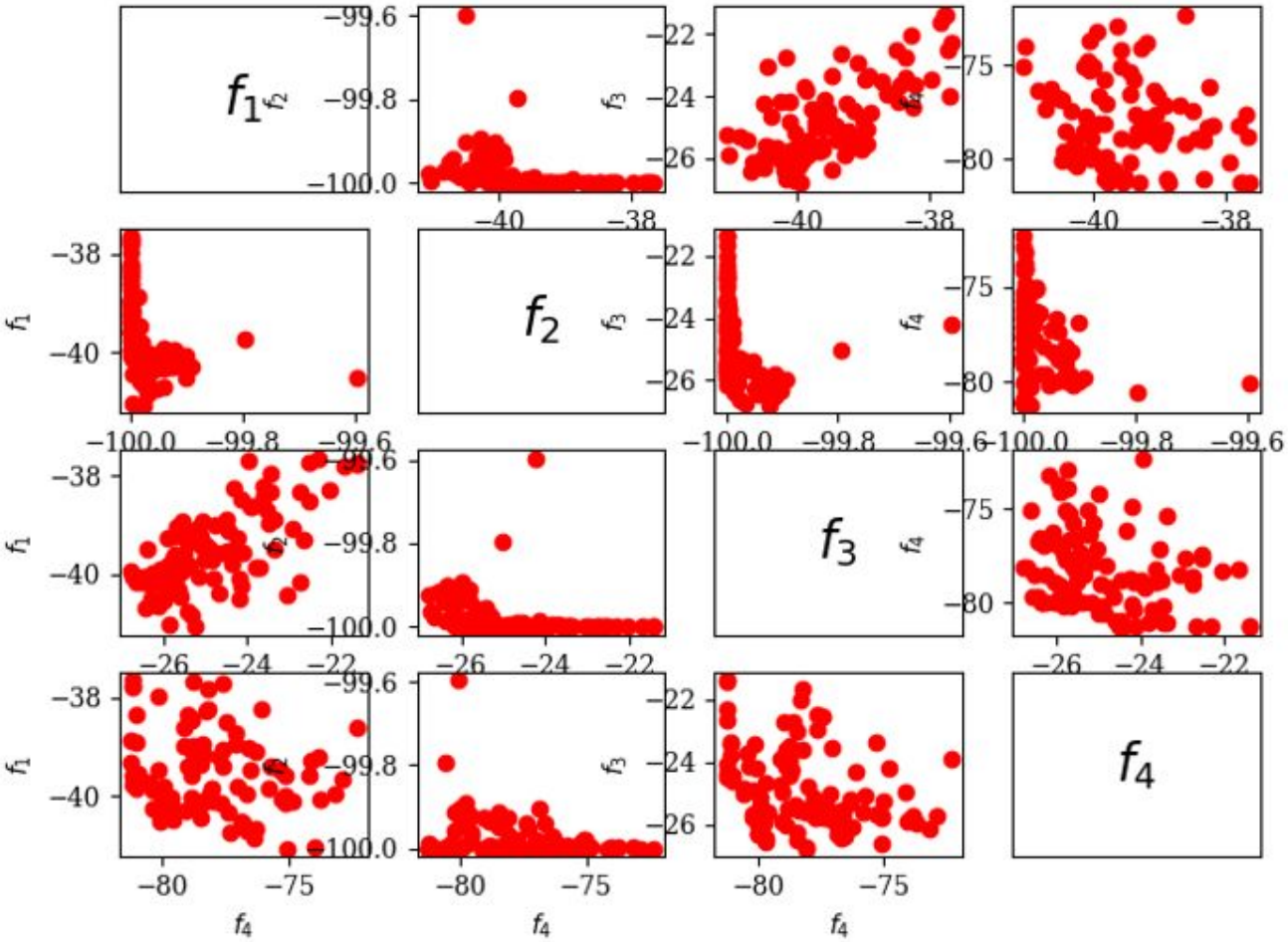
```
txt = a street scene with a double-decker bus on the road.
description - text similarity score = 35.97
image realistic score = 89.17
predict label similarity score = 26.11
label = trolleybus , probability = 0.33
label = streetcar , probability = 0.25
label = minibus , probability = 0.17
label = passenger car , probability = 0.11
label = amphibian , probability = 0.04
```

```
txt = a street scene with a double-decker bus on the road.
description - text similarity score = 38.34
image realistic score = 99.82
predict label similarity score = 28.43
label = trolleybus , probability = 0.84
label = streetcar , probability = 0.14
label = minibus , probability = 0.01
label = passenger car , probability = 0.01
label = school bus , probability = 0.00
```

a zebra is walking in the snow.

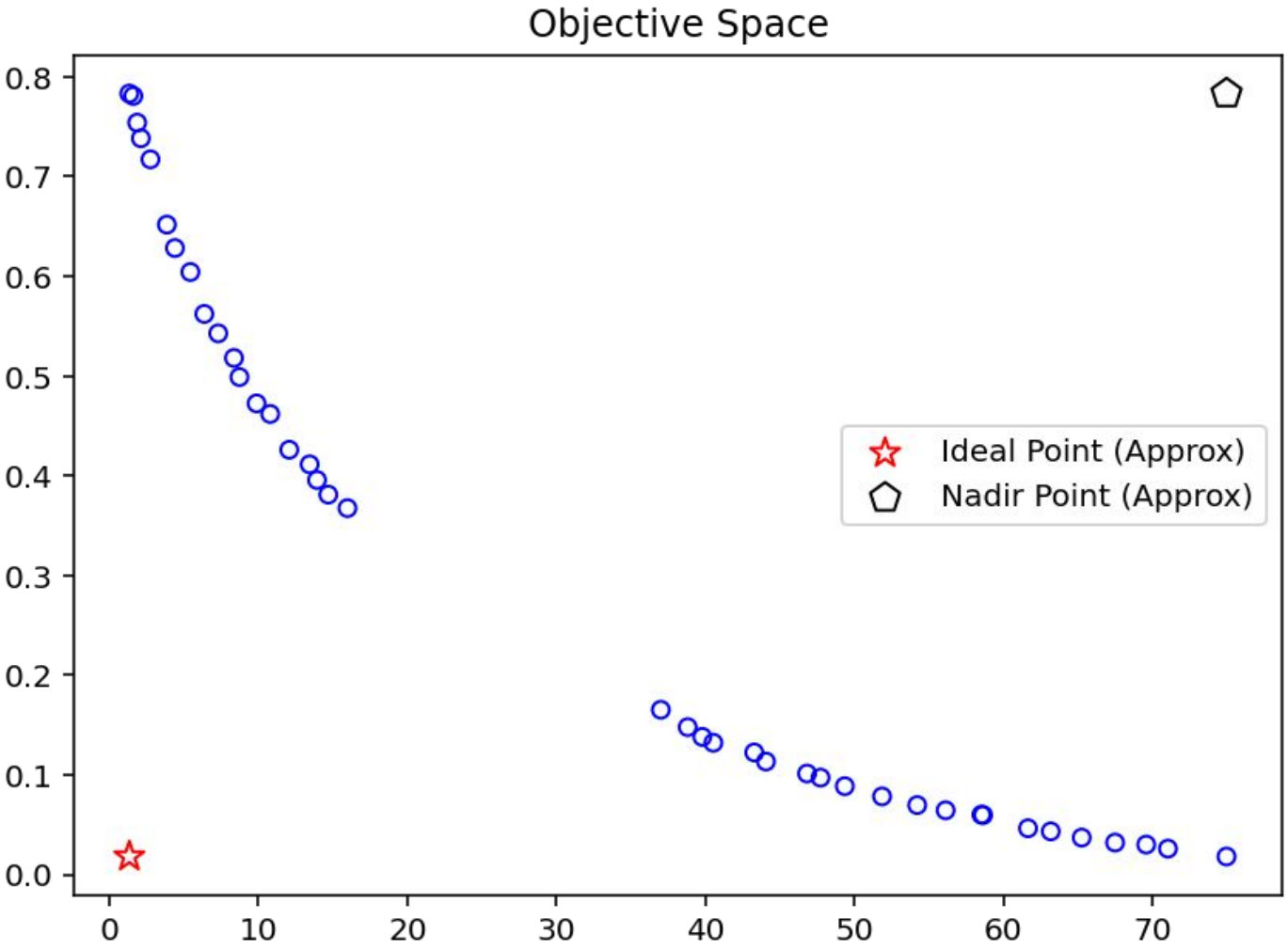
population : 100
#offspring : 100
generation : 50

1. text-image-simularity
2. image realistic score
3. image-label-simularity
4. word-label-simularity



MCDM

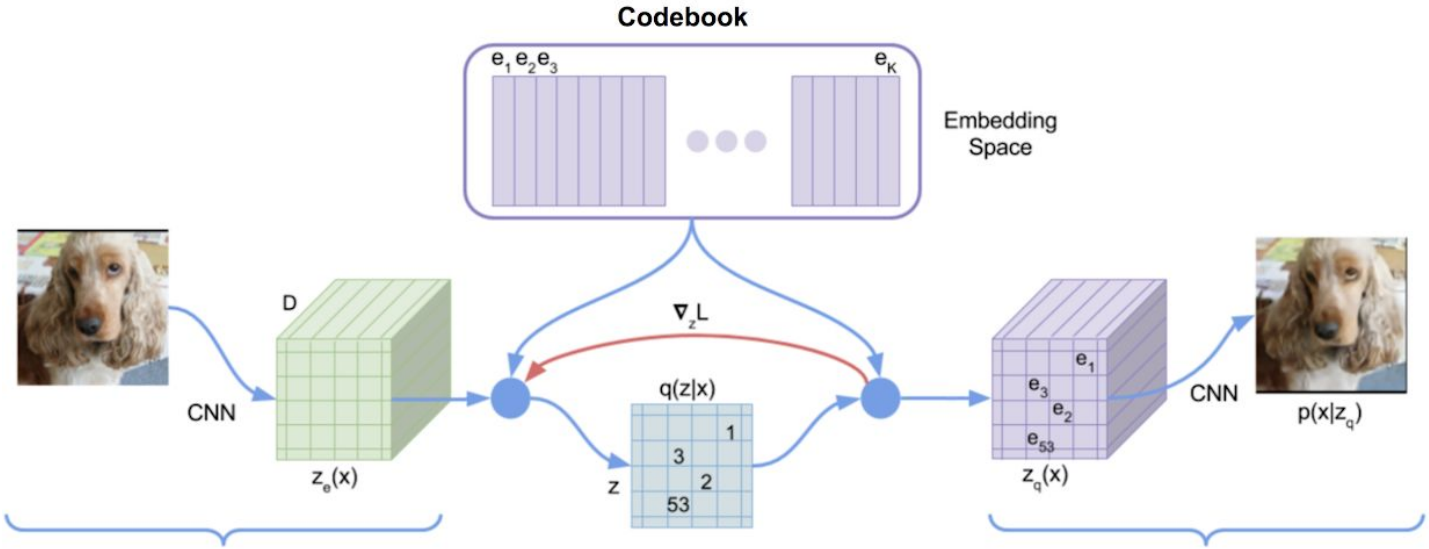
```python
def mcdm(res = None):
    F = res.F
    approx_ideal = F.min(axis=0)
    approx_nadir = F.max(axis=0)

    # normalization
    nF = (F - approx_ideal) / (approx_nadir - approx_ideal)

    # setting weigth
    weights = np.array([0.30,0.40,0.30])
    decomp = get_decomposition("asf")

    # inverse
    I = decomp.do(nF, 1/weights).argmin()

    print("Best regarding decomposition: ASF {} - {}".format(I, -F[I]))
    return I
```

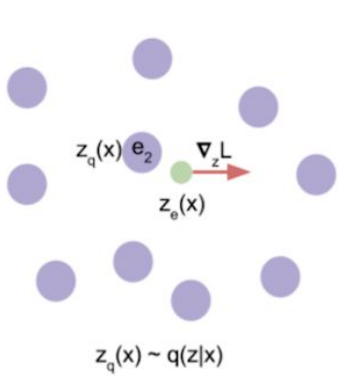

Objective Space
Ideal Point (Approx)
Nadir Point (Approx)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
10
20
30
40
50
60
70

explain

- Normalizing the obtained objective values regarding the boundary points.
- you might be wondering why the weights are not passed directly, but 1/weights. For ASF, different formulations exist, one where the values are divided and one where they are multiplied. In pymoo, we divide, which does not reflect the idea of the user’s criteria. Thus, the inverse needs to be applied.

noise codebook





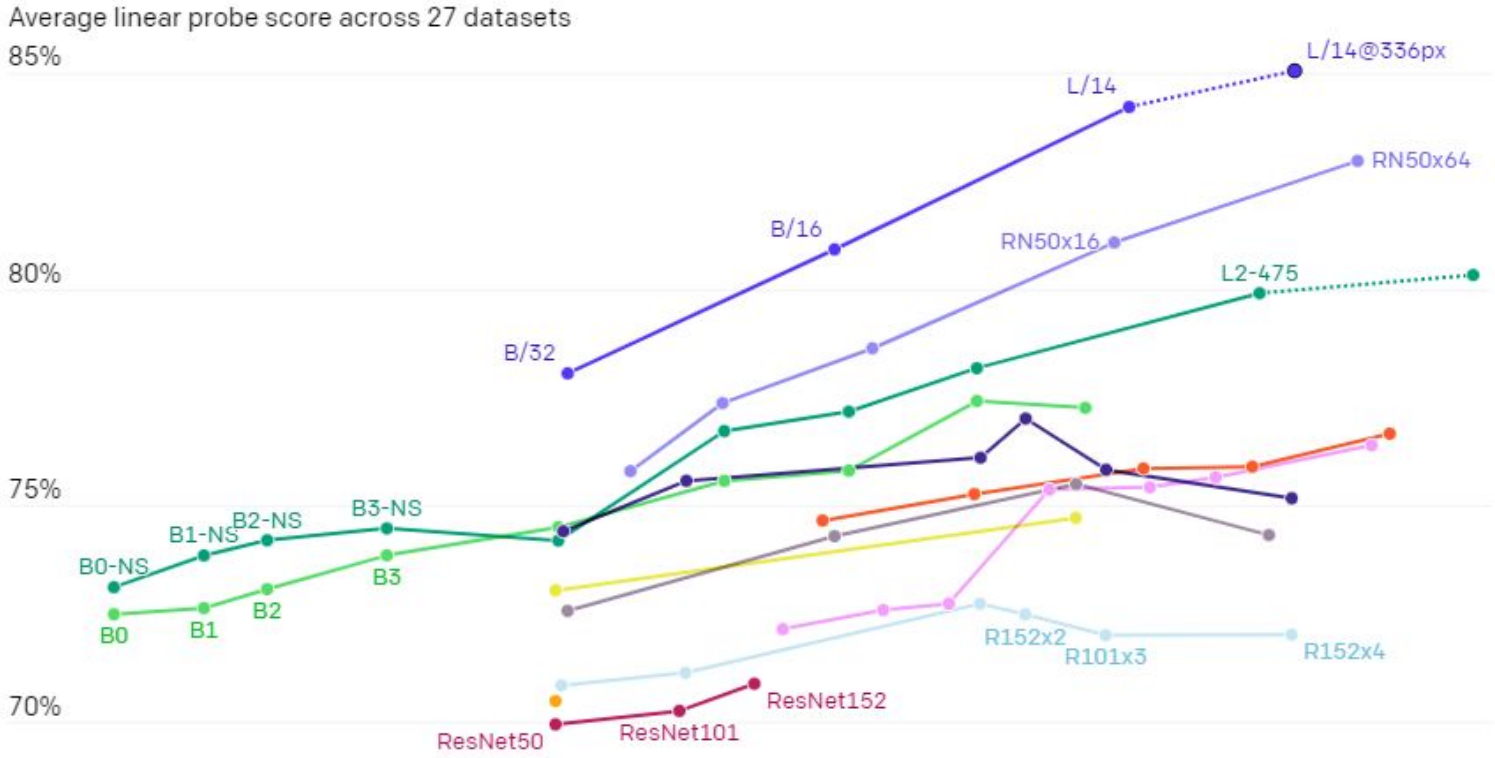
Posterior categorical distribution:

$$q(\mathbf{z}=\mathbf{e}_k|\mathbf{x}) = \begin{cases} 1 & \text{if } k = \arg\min_i \|\mathbf{z}_e(\mathbf{x}) - \mathbf{e}_i\|_2 \\ 0 & \text{otherwise.} \end{cases}$$

codebook benefit

- 範圍可控制
- 穩定
- 生成結果更好

large clip model



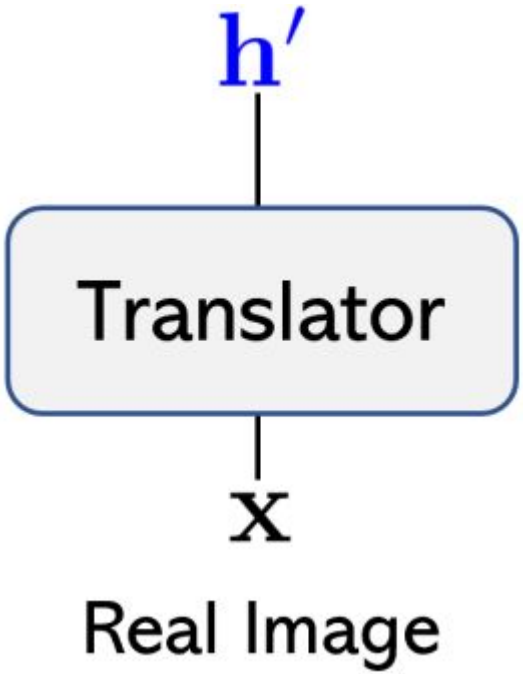
datasets

- COCO2017
- Image-net (classification)

```
templates = [
    'a photo of a {}.',
    'a photo of the {}.',
]
```

- 3. Datasets
  - Caltech-UCSD Bird (CUB)
  - Oxford-102 Flower
  - MS-COCO
  - Multi-Modal-CelebA-HQ
  - CelebA-Dialog
  - FFHQ-Text
  - CelebAText-HQ

pseudo text feature



Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

ref : https://arxiv.org/pdf/2105.05233.pdf

study

- CLIP
- VQGAN (Taming Transformers)
- Diffusion Models Beat GANs on Image Synthesis
- Guided Diffusion
- Make-A-Scene