MaskCon: Masked Contrastive Learning for Coarse-Labelled Dataset

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Feng, Chen, and Ioannis Patras. "MaskCon: Masked contrastive learning for coarse-labelled dataset." (CVPR 2023)

AI 5100 Deep Learning

Presenters:
Raj Popat (CS23MTECH14009)
Supriya Rawat (CS23MTECH11019)

Content

- Problem Statement
- Terminologies
- Drawbacks of previous approach
- Maskcon Framework
- Experiments
- Conclusion

Problem Statement

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- Annotating large-scale datasets accurately and efficiently, especially in specialized domains requires fine-grained labels, is costly and challenging.
- Obtaining coarse labels is easier but does not suffice for finer classification.
- **MaskCon**, a masked contrastive learning method, leverages coarse-labelled data for addressing **finer labelling problems.**

Terminologies

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- **Self-supervised contrastive learning:**
  Makes a model learn representations by contrasting different views of the same data, without labeled annotations,
- **Supervised contrastive learning:**
  - Model learns representations by contrasting similar and dissimilar pairs of data points, using labeled annotations to guide the learning process.
  - Improves discrimination between classes and enhances performance in classification tasks.

$$R(f,h) = \sum_{i=1}^{N} L_{con}(\boldsymbol{x}_i, \boldsymbol{z}_i; f, h),$$

where the contrastive loss L_{con} is defined as follows:

$$L_{con}(\boldsymbol{x}_i, \boldsymbol{z}_i; f, h) = -\sum_{n=1}^{N} \boldsymbol{z}_i^n \log \boldsymbol{q}_i^n.$$

- In the contrastive learning the following empirical risk will be optimized.
- where the contrastive loss is defined as follows.
- Where zi is inter sample relationship.

Drawbacks of previous approach

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- In the fine-grained labeling problem, previous approaches have utilized self-contrastive learning as an auxiliary method, along with supervised contrastive learning and cross-entropy under coarse labels. However, all of these methods have struggled to learn ideal representations for fine labels.
- Using only cross-entropy or supervised contrastive learning tends to result in under clustering issues. Therefore, researchers have combined them with self-supervised contrastive learning to improve performance. Despite these efforts, there remains a significant performance gap with these methods.
- In MaskCon, researchers address this issue by updating the inter-sample relations of contrastive learning. Specifically, they set the probability of similarity to one for samples with the same coarse label, zero for samples with different coarse labels, and maintain it as one for the sample itself.

Maskcon Framework

- Prioritizes learning inter-sample relations through cosine similarity.
- Establish confidently positive relations with oneself and **estimate relations with other samples through soft labels derived from contrasting augmented views**.

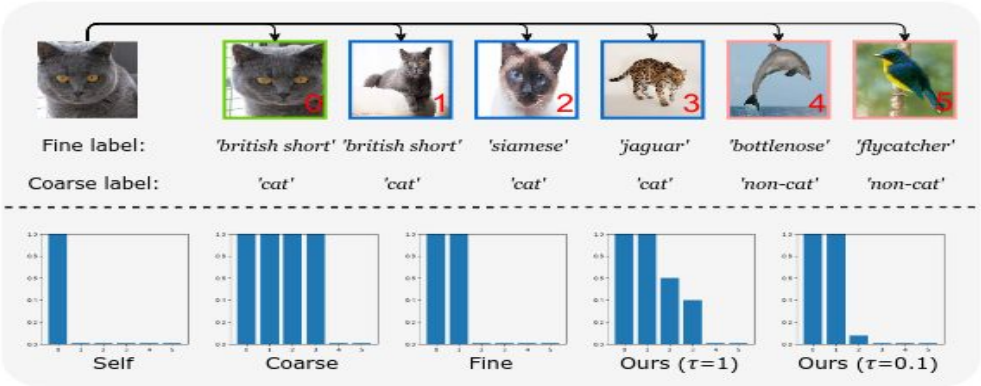


Fig: MaskCon

Maskcon framework

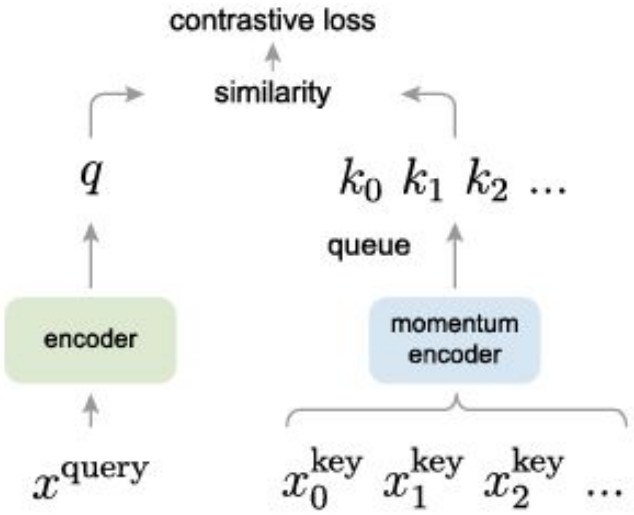


Fig: MOCO paper

$$\boldsymbol{d}_i = [\cos(\boldsymbol{h}_i, \boldsymbol{h}_1), \cos(\boldsymbol{h}_i, \boldsymbol{h}_2), ..., \cos(\boldsymbol{h}_i, \boldsymbol{h}_N)],$$

Finding similarity between query and key.

$$\boldsymbol{q}_i \triangleq \text{softmax}(\boldsymbol{d}_i/\tau_0),$$

Finding qi by applying softmax to di with temperature parameter.

$$R(f, h) = \sum_{i=1}^{N} L_{con}(\boldsymbol{x}_i, \boldsymbol{z}_i; f, h),$$

where the contrastive loss L_{con} is defined as follows:

$$L_{con}(\boldsymbol{x}_i, \boldsymbol{z}_i; f, h) = -\sum_{n=1}^{N} \boldsymbol{z}_i^n \log \boldsymbol{q}_i^n.$$

Z denotes the sample-wise normalized inter-sample relations.

Maskcon Framework

Here we defined inter sample relationship for different approaches

$$z_{ij}^{self} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Self - Supervised
Contrastive Learning

$$z_{ij}^{sup} = \begin{cases} 1, & \text{if } \boldsymbol{y}_i = \boldsymbol{y}_j \\ 0, & \text{if } \boldsymbol{y}_i \neq \boldsymbol{y}_j \end{cases}$$

Supervised
Contrastive Learning

Different loss function for the previous approaches

$$L_{Grafit} = wL_{supcon} + (1-w)L_{selfcon}$$
$$L_{CoIns} = wL_{ce} + (1-w)L_{selfcon}$$

w controls the relative weight of each loss.

$$L_{ce}(\boldsymbol{x}_i, \boldsymbol{y}_i; f, g) = -\sum_{m=1}^{M} \boldsymbol{y}_i^m \log \boldsymbol{p}_i^m$$

Cross entropy loss

Maskcon Framework

$$z'_{ij} = \frac{\mathbb{1}(\boldsymbol{y}_j = \boldsymbol{y}_i) \cdot \exp(d'_{ij}/\tau)}{\sum_{n=1,n\neq i}^{N} \mathbb{1}(\boldsymbol{y}_n = \boldsymbol{y}_i) \cdot \exp(d'_{in}/\tau)}, i \neq j,$$

where the similarity $\boldsymbol{d}'_i$ is given by

$$\begin{aligned}\boldsymbol{d}'_i = &[\cos(\boldsymbol{h}_i^k, \boldsymbol{h}_1), ..., \cos(\boldsymbol{h}_i^k, \boldsymbol{h}_{i-1}),\\ &\cos(\boldsymbol{h}_i^k, \boldsymbol{h}_{i+1}), ..., \cos(\boldsymbol{h}_i^k, \boldsymbol{h}_N)].\end{aligned}$$

$$z'_{ij} = z'_{ij} / \max(\boldsymbol{z}'_i),$$

$$z_{ij}^{mask} = \begin{cases} 1, & \text{if } i = j \\ z'_{ij}, & \text{if } i \neq j \end{cases}$$

$$L = wL_{maskcon} + (1-w)L_{selfcon}$$

Here we defined inter sample relationship for MaskCon approach :
Here for other coarse label it is zero and for same coarse label it is probability between 0 to 1

It is between the keys of the images

For the same image it is 1 but for other it is same

This is the final loss function which is convex combination of the Maskcon loss and self contrastive loss

Maskcon Framework

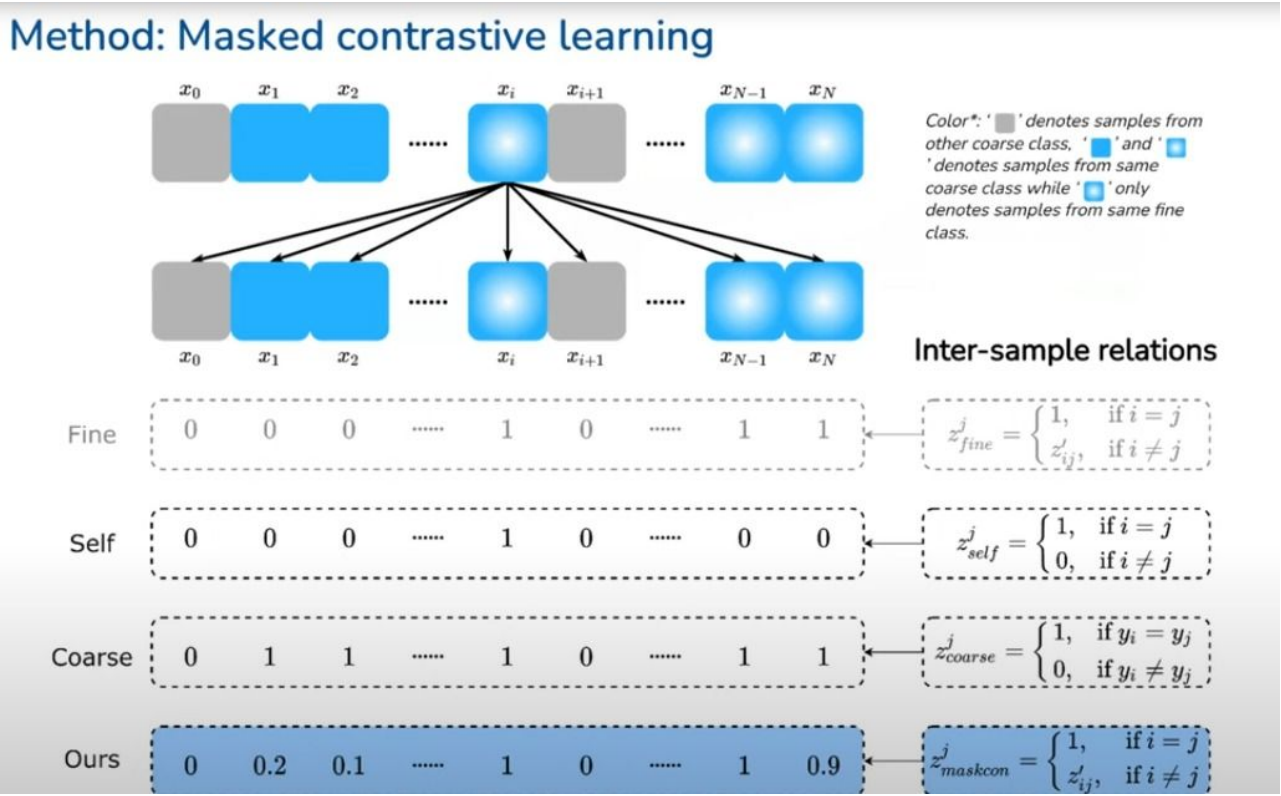


Fig: MaskCon

Maskcon framework

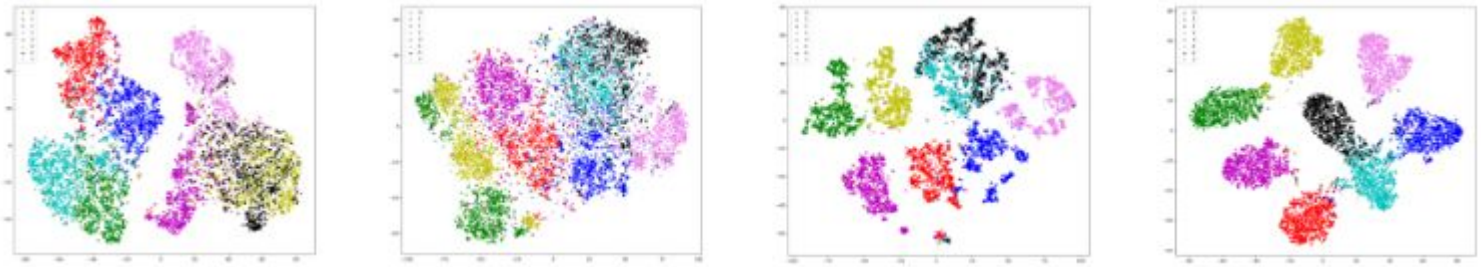


Fig: MaskCon

Can the gap between completely supervised setup and unsupervised setup be reduced by introducing similarity between data points?

- Here ‘SupFine’ is applying same algorithm with fine labels

Results from the paper

Method	Recall@1	Recall@2	Recall@5	Recall@10
SelfCon	10.28	14.15	22.36	30.34
Grafit	18.13	25.46	37.19	46.64
SupCon	13.36	19.40	29.77	39.38
CoIns	18.36	25.54	37.09	46.89
SupCE	12.23	14.15	27.76	37.03
SupFINE	33.97	44.55	57.23	65.77
MaskCon (Ours)	**19.08** (6.86↑)	**26.21**	**38.17**	**47.96**

Results on ImageNet -1K dataset

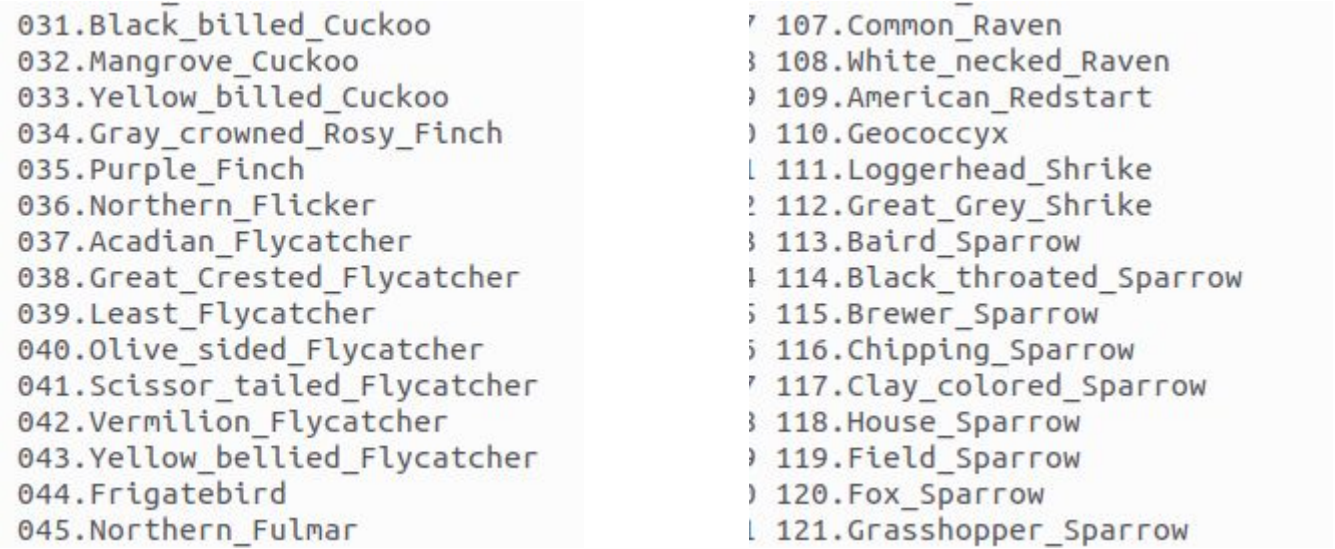
Method	Recall@1	Recall@2	Recall@5	Recall@10
SelfCon	70.36	75.57	81.53	85.13
Grafit	74.02	78.82	84.13	87.91
SupCon	53.69	59.55	67.12	72.78
CoIns	70.84	76.01	82.2	86.08
SupCE	36.35	42.39	50.30	56.52
SupFINE	83.94	88.04	91.95	94.00
MaskCon (Ours)	**74.05** (37.7↑)	**78.97**	**84.48**	**87.96**

Results on SOP Split-1 dataset

Table : MaskCon

Our proposed method

- Can we make the model learn better by leveraging the information associated with the binary vectors corresponding to every data point?



Caltech-UCSD Birds-200-2011

Caltech-UCSD Birds-200-2011 (CUB dataset)

- A collection of 11,788 images spanning 200 bird species.
- Each image is annotated with detailed attributes, including part locations and bounding boxes.
- Binary attributes for more precise identification and analysis.

```
301 has_crown_color::olive
302 has_crown_color::green
303 has_crown_color::pink
304 has_crown_color::orange
305 has_crown_color::black
306 has_crown_color::white
307 has_crown_color::red
308 has_crown_color::buff
309 has_wing_pattern::solid
310 has_wing_pattern::spotted
311 has_wing_pattern::striped
312 has_wing_pattern::multi-colored
```

```
1 309 0
1 310 0
1 311 0
1 312 0
2 1 0 1
2 2 0 1
2 3 0 1
2 4 0 1
2 5 0 1
2 6 0 1
2 7 0 1
2 8 0 1
```

312- dimensional binary vector for every data point.

Our proposed method

- Create vector representation for each image, given binary valued attributes.
- Find a similarity metric between each image attribute vector within the same coarse label.
- Obtain a probability distribution using softmax function.
- Take convex combination of the obtained probabilities with the vector Zij'.
- Using refined probabilities in the original MaskCon loss function to obtain fine labels.

Metrics to capture similarity

- **Cosine similarity** - Dot product between similar species should be higher.
- **Hamming distance** - The difference between dissimilar data points is higher.
- **Euclidean distance** - Similar points have less distance between them.

Adding the newly obtained information to the MaskCon loss

A convex combination of similarity measures Zij′ and Zij′_new are taken to augment it into the loss function :

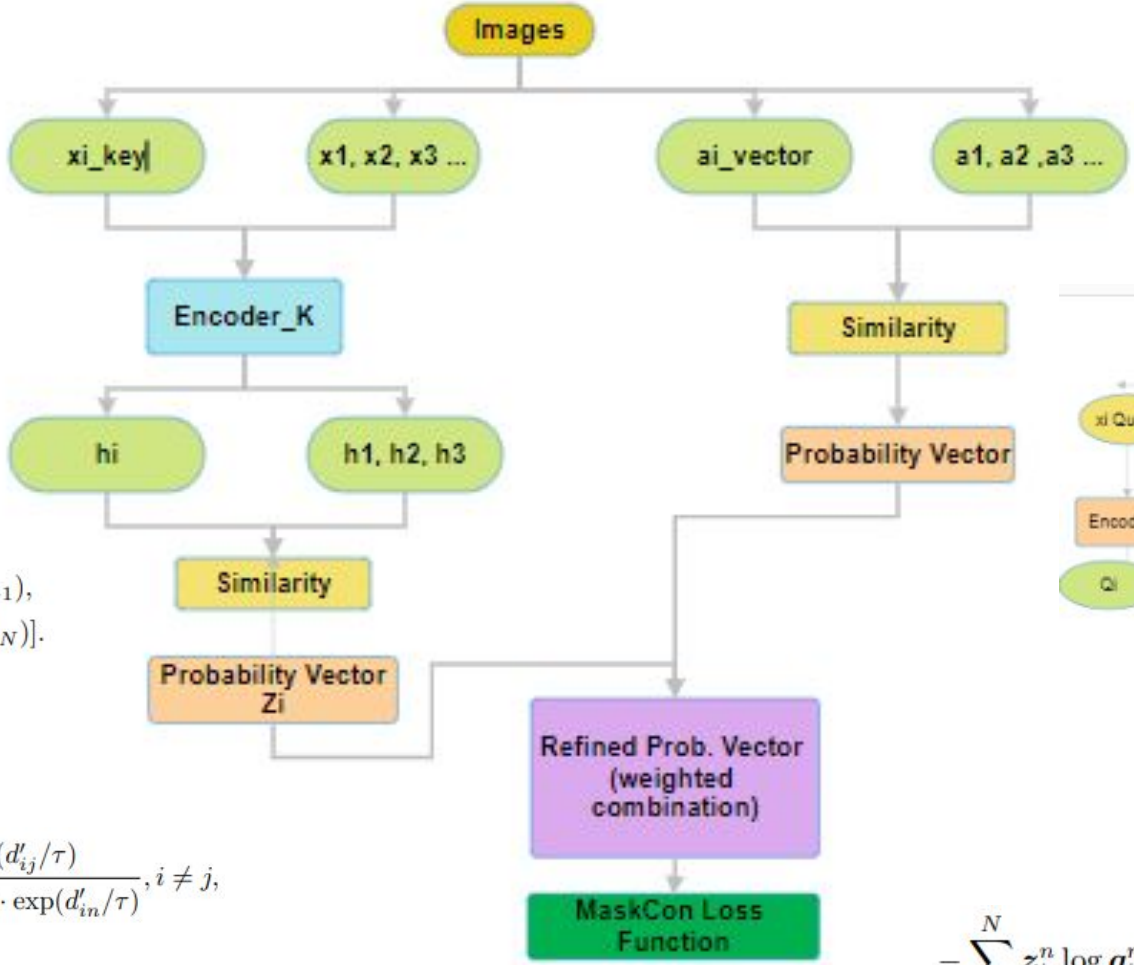
Zij′_new = w*(Zij′) + (1-w)*(Zij′_new) , 0< w<1.

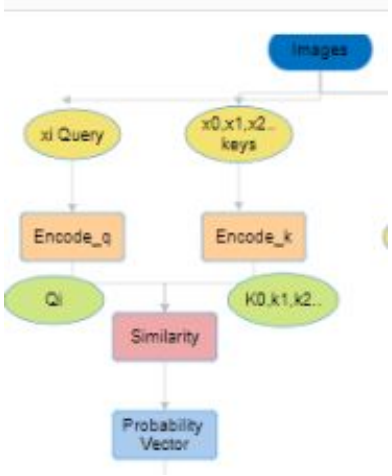
New loss : - Replace (Zij′) with Zij′_new in contrastive loss function

Why adding information in the loss function is more intuitive?

To guide initial probabilities of Z_ij with Z_new_ij which in turn will help the model better learn the similarities between the data points belonging to the same coarse labels.

END to END MASKCON WITH OUR ADDITION





$$\boldsymbol{d}'_i = [\cos(\boldsymbol{h}_i^k, \boldsymbol{h}_1), ..., \cos(\boldsymbol{h}_i^k, \boldsymbol{h}_{i-1}), \cos(\boldsymbol{h}_i^k, \boldsymbol{h}_{i+1}), ..., \cos(\boldsymbol{h}_i^k, \boldsymbol{h}_N)].$$

$$z'_{ij} = \frac{\mathbb{1}(\boldsymbol{y}_j = \boldsymbol{y}_i) \cdot \exp(d'_{ij}/\tau)}{\sum_{n=1,n\neq i}^{N} \mathbb{1}(\boldsymbol{y}_n = \boldsymbol{y}_i) \cdot \exp(d'_{in}/\tau)}, i \neq j,$$

$$z'_{ij} = z'_{ij} / \max(\boldsymbol{z}'_i),$$

$$-\sum_{n=1}^{N} \boldsymbol{z}_i^n \log \boldsymbol{q}_i^n.$$

Experiment results

	Recall@1	Recall@2	Recall@5	Recall@10	Recall@50	Recall@100
MaskCon	63.635	73.85	84.85	91.44	97.51	98.48
Euclidean similarity W1 = 0.3 & w2 = 0.7	60.73	71.28	83.37	89.73	97.65	98.63
SupFine	69.05	78.75	87.41	92.35	97.63	98.61

Fine classes = 200

Experiment results

	Recall@1	Recall@2	Recall@5	Recall@10	Recall@50	Recall@100
Euclidean Similarity w1= 0.5 &w2 =0.5	79.61	88.35	96.69	99.80	99.80	99.80
Euclidean similarity W1 = 0.3 & W2 = 0.7	80.38	88.93	96.50	98.64	99.80	99.80
Maskcon	77.86	87.57	94.36	97.86	99.41	99.80
Cosine similarity W1 = 0.3&w2= 0.7	75.72	84.66	93.20	97.08	99.80	1.00
SupFine	87.76	92.23	94.95	97.67	99.80	1.00

Fine classes = 20

Experiments results

	Recall@1	Recall@2	Recall@5	Recall@10	Recall@50	Recall@100
MaskCon	70.90	80.05	91.36	95.89	99.20	99.71
Euclidean similarity W1 = 0.3 & w2 =0.7	71.00	81.05	92.00	95.89	99.20	99.71
Euclidean similarity W1 = 0.5 & w2 =0.5	69.40	80.27	90.64	94.96	99.42	99.85
SupFine	76.53	84.73	92.08	95.39	99.42	99.78

Fine classes = 50

Conclusions

- The impact of adding additional information has shown some improvements in the better clustering of data points according to their fine classes.
- More metrics are to be explored in order to improve the Recall@k metrics in the datasets irrespective of their sizes.
- In annotation - deficit settings, such approaches can show promising results.

Thank You