A Comparative Analysis of Multi-Task Learning Approaches in the Context of Multi-Label Remote Sensing Image Retrieval

Jun Xiang

29 March 2021

Overview

Introduction

This project aims to compare the retrieval performance of multi-task metric learning approaches on remote sensing datasets.
The three methods we studied :

- Diverse Visual Feature Aggregation (Diva)
- Divide and conquer (Dac)
- Boosting Independent Embedding Robustly (Bier)

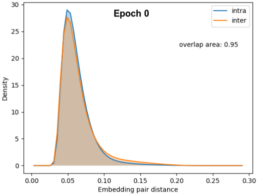
Dataset

BigEarthNet

- Patch info : $20 \times 20, 60 \times 60, 120 \times 120$, 12 channels
- Number of Patches : 519,284
- Class : 43
- Average labels per patch : 3
- Most common class : 24 (Mixed Forest), 22,729 samples (38%)

FIGURE – Statistic of embedding pairs on BigEarthNet val set

Type	Shared labels	BigEarthNet Count	Percent
Inter	0	502,076,436	44.7
Intra	1	372,328,100	33.1
	2	190,863,827	16.9
	3	52,255,466	4.68
	4	6,226,744	0.58
	>4	403,038	0.04
Total		1,124,153,611	100



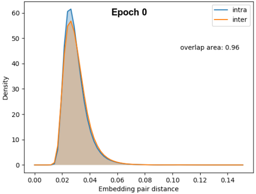
Dataset

MLRSNet

- Patch info : 256 $\times$ 256, 3 channels
- Number of Patches : 109,161
- Class : 60
- Average labels per patch : 4
- Most common class : 56 (Trees), 6410 samples (65%)

FIGURE – Statistic of embedding pairs on MLRSNet val set

		MLRSNet	
Type	**Shared labels**	**Count**	**Percent**
Inter	0	11,629,210	27.4
Intra	1	12,263,061	28.9
	2	7,505,954	17.7
	3	4,648,709	11
	4	3,404,608	8
	>4	2,921,763	7
Total		42,373,305	100



Dataset process

- Data split : ratio 50%/10%/40%

Data split	BigEarthNet	MLRSNet
Patch size	$120 \times 120 \times 12$	$256 \times 256 \times 3$
Crop size	$100 \times 100 \times 12$	$224 \times 224 \times 3$
Train	269,695	49,928
Val	60,824	9,816
Test	188,765	49,407

- Data process :

Data split	Data process
Train	random crop random flip standardized
Val Test	central crop standardized

Adversarial loss

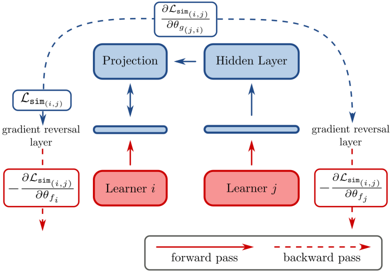


FIGURE – Adversarial loss

$$\zeta^{adv} = -\frac{1}{dim_i} \sum (g(i,j)(x_j) \odot x_i)^2 \tag{1}$$

- dim_i is the length of x_i
- g(i,j) : $R^{d_j} \rightarrow R^{d_i}$, projecting x_j into the space of x_i

Binomial loss

$$\zeta^{bin} = \sum log\left(1 + e^{-y_{ij}\beta_1(s_{ij}-\beta_2)C_{ij}}\right)$$

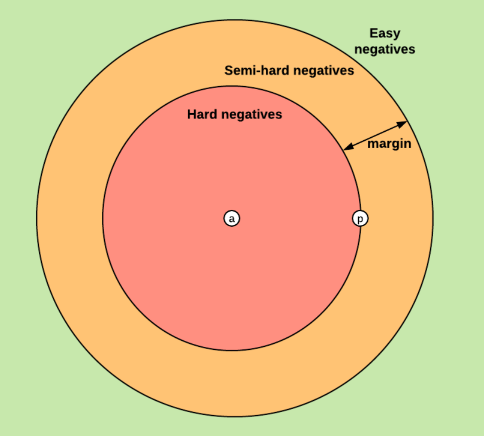
where:

$$s_{ij} = x_i {x_j}^T$$

$$C_{ij} = \begin{cases} 1, & \text{if } y_{ij} = 1 \\ 25, & \text{if } y_{ij} = -1 \end{cases}$$

β_1 : the scaling parameter and set to 2 by default
β_2 : the margin parameter with default initial value 0.5

Batch miner



- Semihard : online batch hard strategy
- Distance : online mining, more flexible (easy, semi-hard and hard)

Diverse Visual Feature Aggregation (Diva)

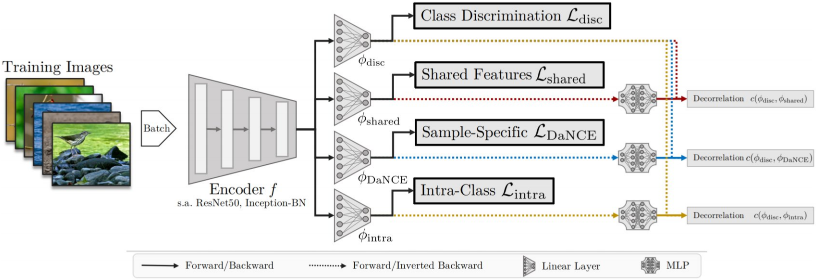


FIGURE – DiVA architecture (Milbich et al., 2020)

$$\begin{aligned}\zeta_{Diva} = \zeta_{disc} + \alpha_1\zeta_{shared} + \alpha_2\zeta_{intra} + \alpha_3\zeta_{DaNCE} \\ +\rho_1\zeta^{adv}(\phi_{disc}, \phi_{NCE}) \\ +\rho_2\zeta^{adv}(\phi_{disc}, \phi_{shared}) \\ +\rho_3\zeta^{adv}(\phi_{disc}, \phi_{intra})\end{aligned} \tag{2}$$

Divide and Conquer (D&C)

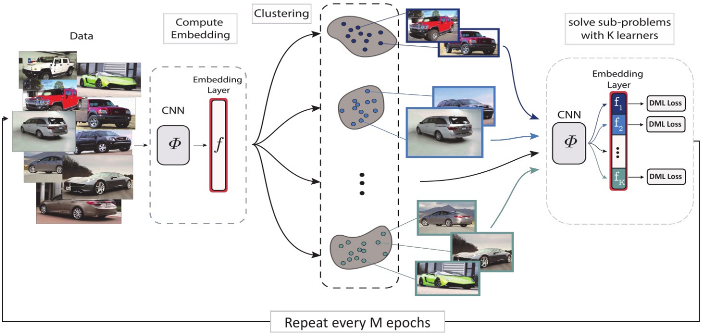


FIGURE – D&C architecture (Sanakoyeu et al., 2019)

$$\zeta_{D\&C} = \frac{1}{|T_k|} \sum_{t \in T_k} \zeta_k^{margin}(t) \tag{3}$$

Boosting Independent Embedding Robustly (Bier)

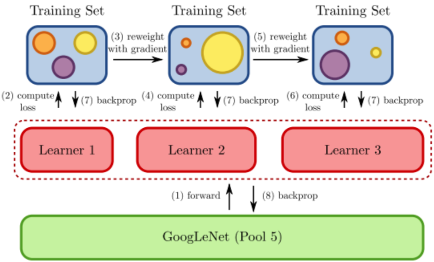


FIGURE – Bier architecture (Opitz et al., 2018)

$$\zeta_{Bier} = \zeta_{boosted} + \lambda_{div}(\lambda_{weight}\zeta_{weight} + \rho_1\zeta^{adv}(\phi_1, \phi_2) + \rho_2\zeta^{adv}(\phi_1, \phi_3) + \rho_3\zeta^{adv}(\phi_2, \phi_3)) \quad (4)$$

Metrics

- Recall
- R-Precision
- MAP

TABLE – Comparison of metrics (Musgrave et al., 2020)

Retrieved	Correct	R@1	R@10	R-P@10	MAP@10
10 results	1st	100	100	10	10
10 results	1st, 10th	100	100	20	12
10 results	1st, 2th	100	100	20	20
10 results	All	100	100	100	100

Training setup

Training Setup	
Number of samples per class	2 for MLRSNet 4 for BigEarthNet
Embedding size	512
Backbone	ResNet50
Weight decay of backbone	1e-4
Learning rate of backbone	1e-5
Weight decay of embedding layer	1e-4
Learning rate of embedding layer	1e-5
Learning rate of adversial layer	1e-5
initial β for Margin loss	1.2
initial β_2 for Binominal loss	0.5
Learning rate of β, β_2	5e-4
Learning rate scheduler	step gamma 0.3 tau [55]
Optimizer weight decay	4e-4
Optimizer	Adam

Experiment plan

Setting	Backbone	Margin (β)	Binomial (β_2)
1	frozen	fixed value 1.2	fixed value 0.5
2	frozen	trainable (class)	
3	frozen	trainable	trainable
4	unfrozen	fixed value 1.2	fixed value 0.5
5	unfrozen	trainable (class)	
6	unfrozen	trainable	trainable

Retrieval scores

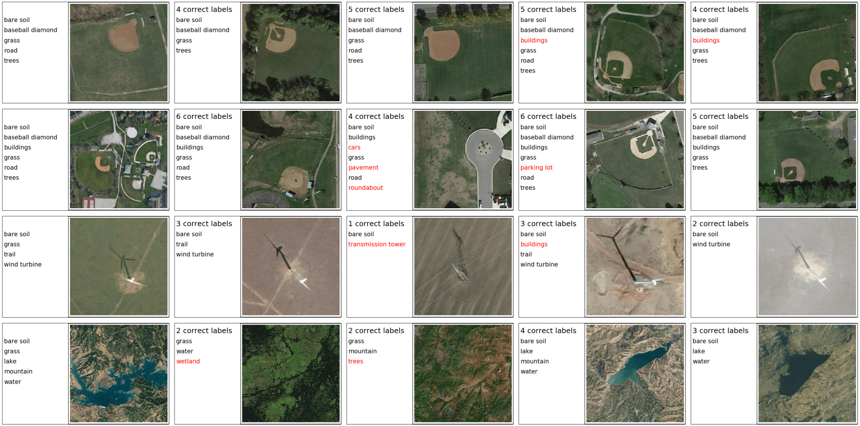
- BigEarthNet

Methods	Backbone	Margin	R@1	R-P@8	MAP@8
Baseline	unfrozen	fixed	80.5	**43.6**	36.0
Diva	unfrozen	trainable	68.7	35.0	27.1
D&C	unfrozen	fixed	76.5	40.4	32.6
Bier	unfrozen	trainable	**82.0**	43.4	**36.3**

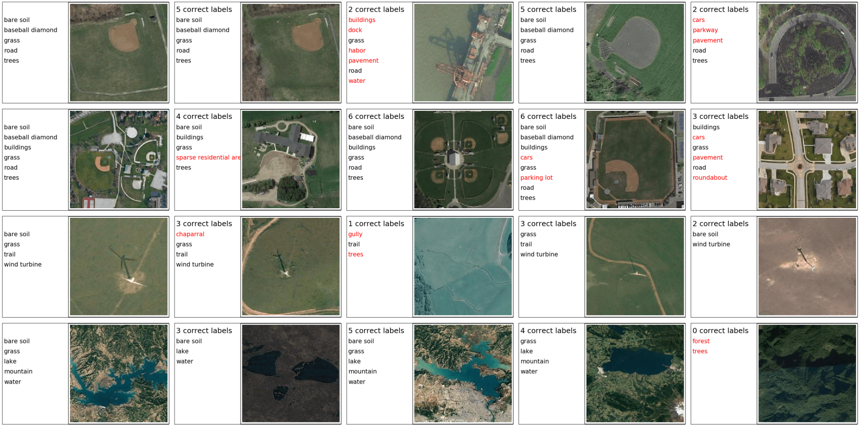
- MLRSNet

Methods	Backbone	Margin	R@1	R-P@8	MAP@8
Baseline	unfrozen	fixed	**74.9**	**39.6**	**31.6**
Diva	unfrozen	trainable	72.8	36.1	29.3
D&C	unfrozen	fixed	73.6	37.6	29.7
Bier	unfrozen	trainable	73.4	38.6	30.5

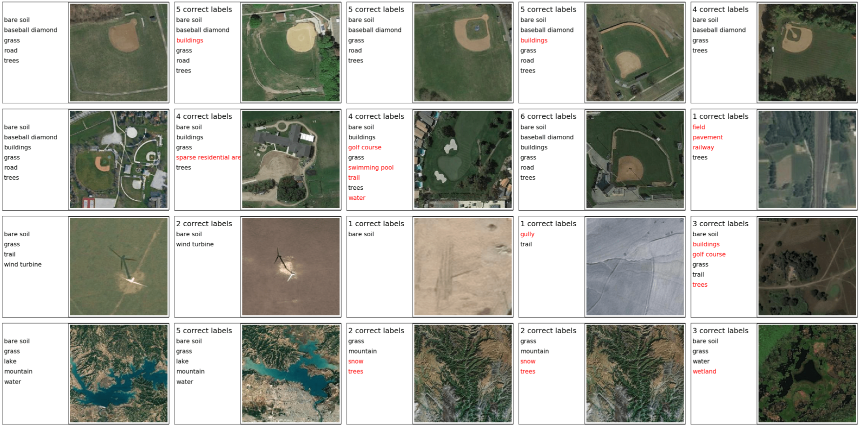
Retrieved images : Baseline



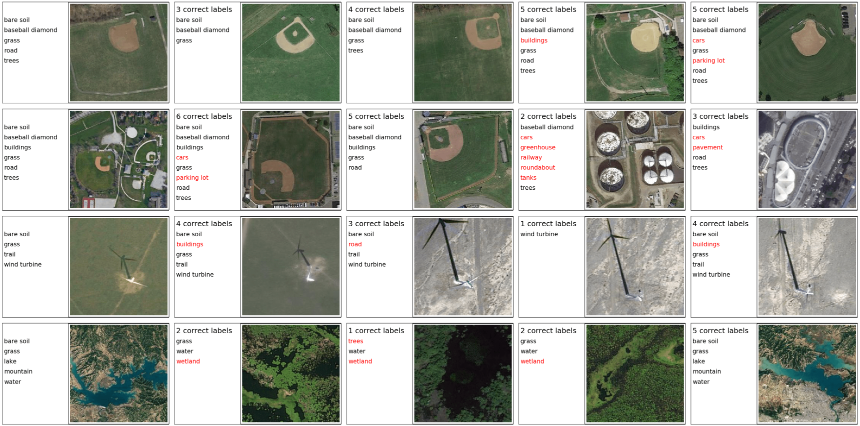
Retrieved images : Diva



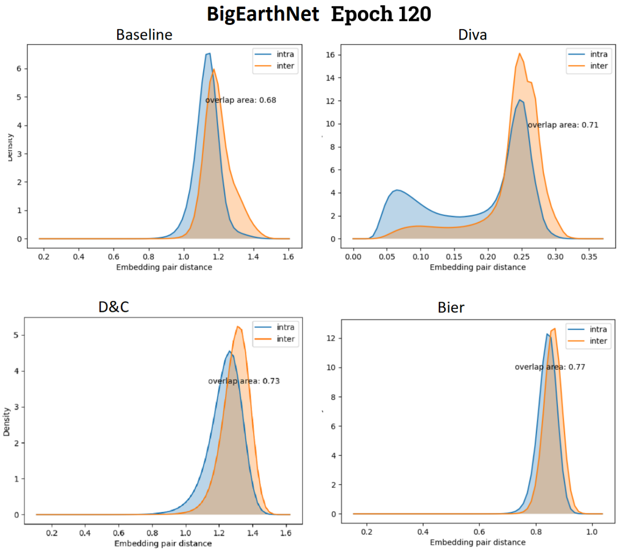
Retrieved images : D&C

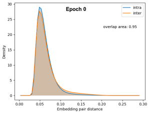


Retrieved images : Bier

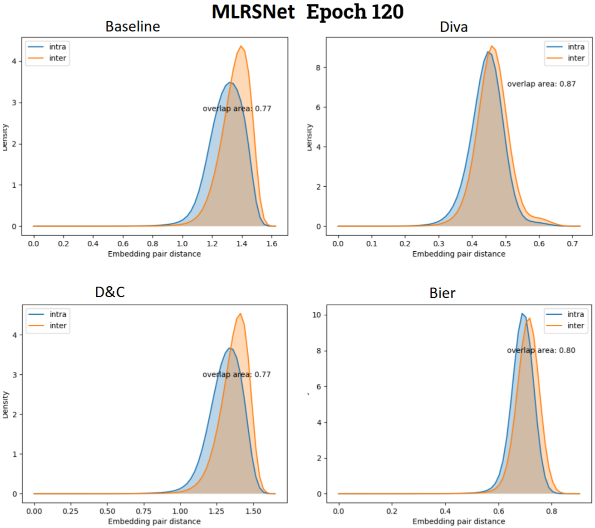


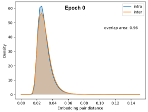
Embedding distance density on BigEarthNet



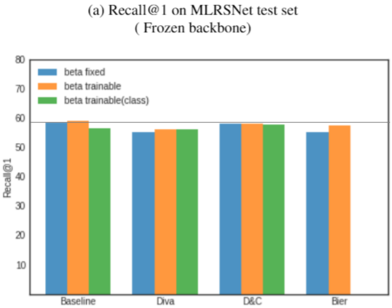


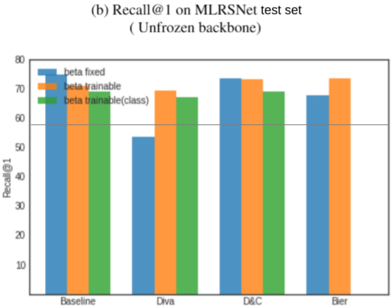
Embedding distance density on MLRSNet



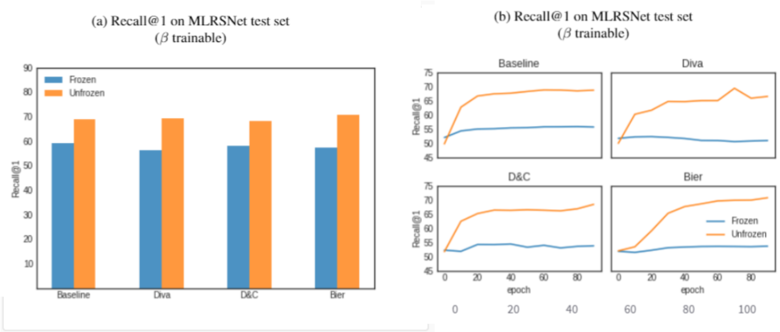


Margin β : Fixed VS Trainable



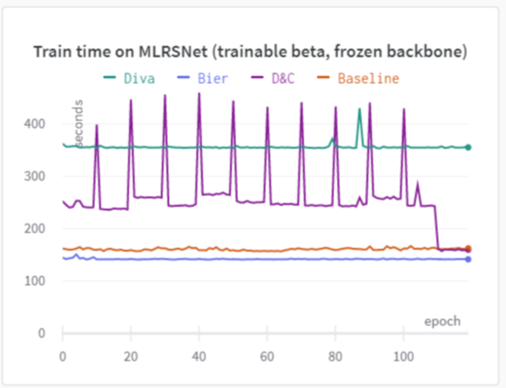


Back bone : Frozen VS unfrozen



Training cost

- PC configuration : GeForceGTX 1080 8G and RAM 32G.
- Batch size : 120



Conclusion

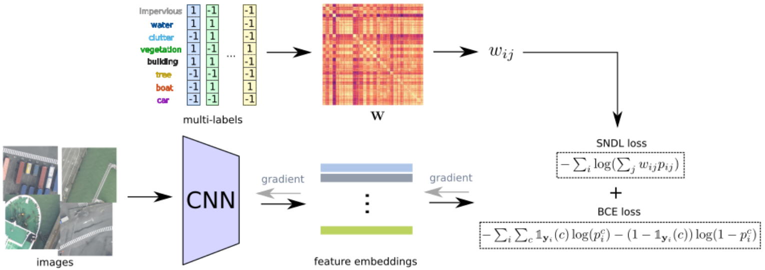
Conclusion :

- Multi-task approaches didn't outperform single-task that much
- Bier worked better than Diva and D&C
- Frozen pretrained Resnet50 is very limited to learn features on remote sensing dataset

Further work :

- Explore optimal initial margin β for each dataset
- Improve the correlation of semantic distance and embedding distance :
  - Explore other triplet mining methods which consider multiple labels jointly
  - Other loss function like SNCA (scalarable network component analysis)

SNCA



Impervious
water
clutter
vegetation
building
tree
boat
car
multi-labels
W
w_{ij}
SNDL loss
$-\sum_i \log(\sum_j w_{ij} p_{ij})$
+
BCE loss
$-\sum_i \sum_c \mathbb{1}_{\mathbf{y}_i}(c) \log(p_i^c) - (1 - \mathbb{1}_{\mathbf{y}_i}(c)) \log(1 - p_i^c)$
images
CNN
gradient
gradient
feature embeddings

Reference

Timo Milbich, Karsten Roth, Homanga Bharadhwaj, Samarth Sinha, Yoshua Bengio, Björn Ommer, and Joseph Paul Cohen. Diva : Diverse visual feature aggregation for deep metric learning. 2020.

Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check. 2020.

M. Opitz, G. Waltner, H. Possegger, and H. Bischof. Deep Metric Learning with BIER : Boosting Independent Embeddings Robustly. **arXiv :cs/1801.04815**, 2018.

Artsiom Sanakoyeu, Vadim Tschernezki, Uta Büchler, and Björn Ommer. Divide and conquer the embedding space for metric learning, 2019.