LaSO: Label-Set Operations networks for multi-label few-shot learning

Amit Alfassy*, Leonid Karlinsky*, Amit Aides*, Joseph Shtok, Sivan Harary, Rogerio Feris
IBM Research AI
Haifa, Israel

Raja Giryes
School of Electrical Engineering, Tel-Aviv University
Tel-Aviv, Israel

Alex M. Bronstein
Department of Computer Science, Technion
Haifa, Israel

CVPR 2019

Introduction

Multi-Label Learning

Person, Sports Ball, Tennis Racket

Person, Tie

Person, Ski

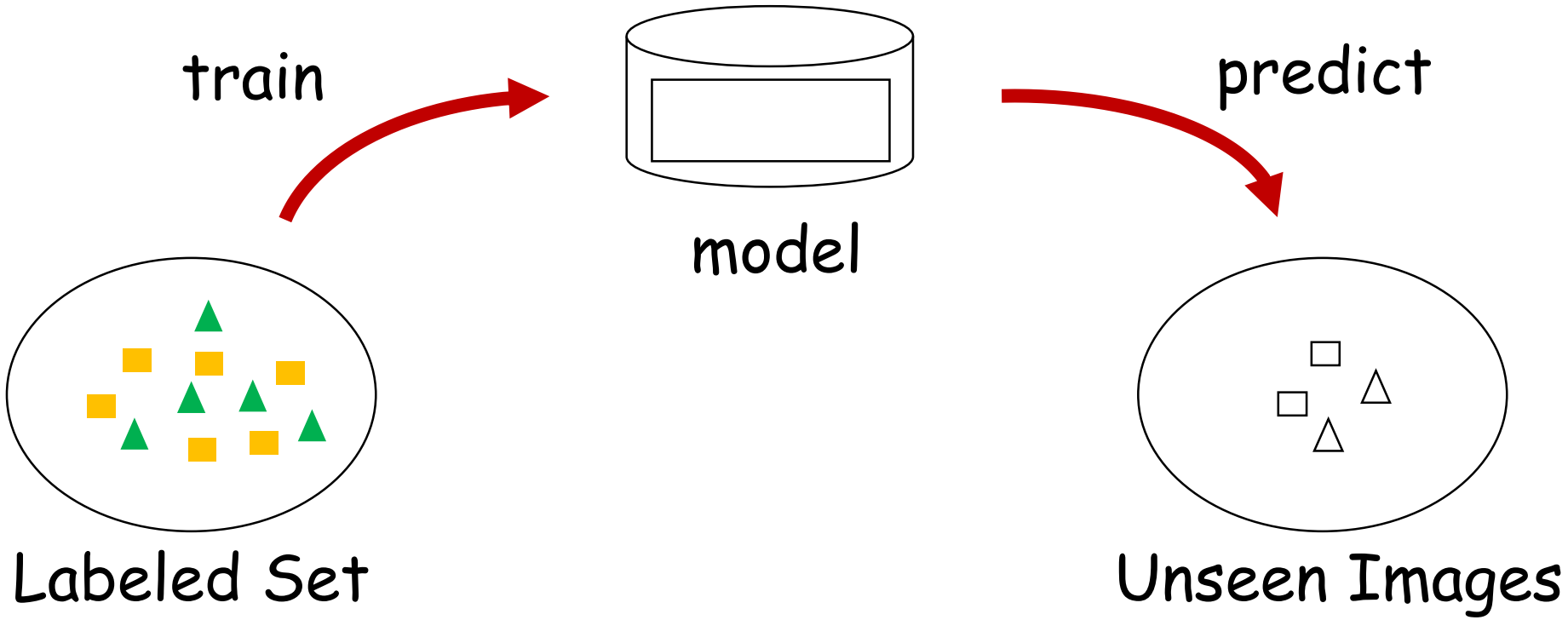
relevant

The goal is to

learn a classifier on training set that can predict all the relevant labels for unseen instances.

Introduction

If you have a multi-label learning task…



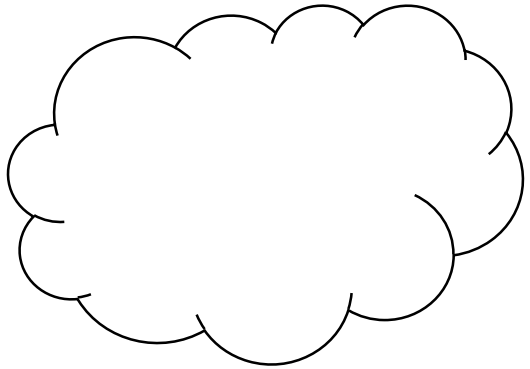
Does it work in case of few labeled data ?

Introduction

An example of few-shot Learning

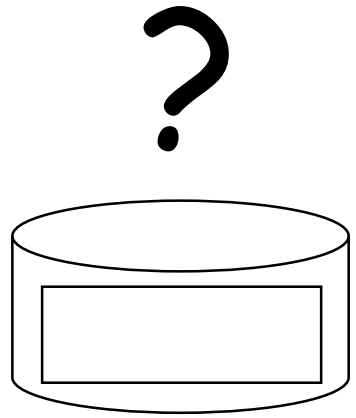
One shot

+



Images of other animals

→



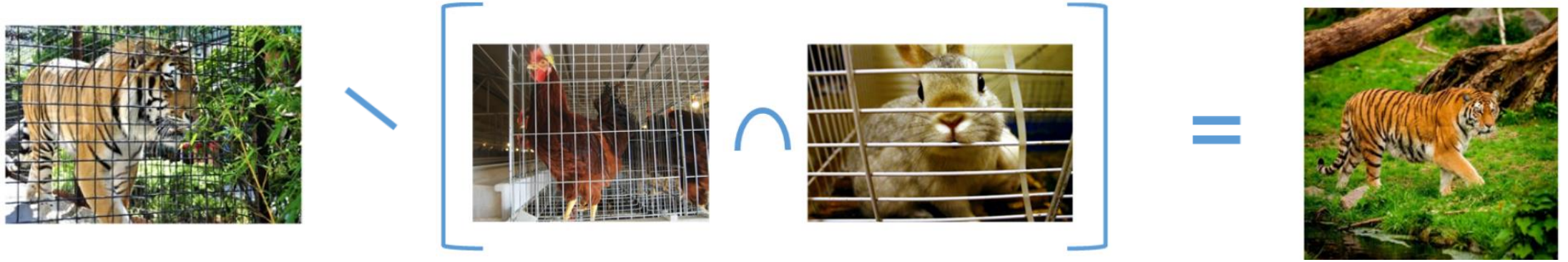
model

Methods

- Meta learning
- Data synthesis (main contribution)

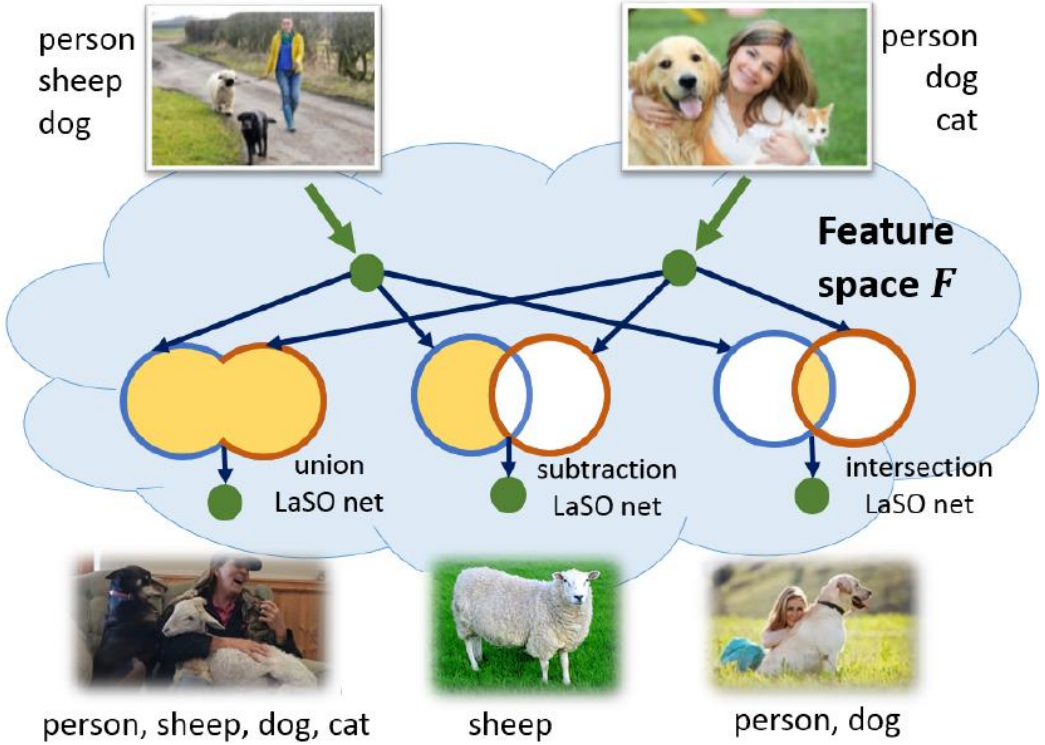
Introduction

A more challenging task…



Label-set operation (LaSO)

Manipulate the 'semantic content' of the samples in feature space 'by example' .



Method

Notation

X and Y : input images

$L(X), L(Y) \subseteq \mathcal{L}$: corresponding label sets

F_X and F_Y : corresponding feature vectors

M_{int} : a model that can accept two images in some feature space and produce a feature vector representing their common semantic content.

$$M_{int}\left(F_X, F_Y\right) = Z_{int} \in \mathcal{F}$$

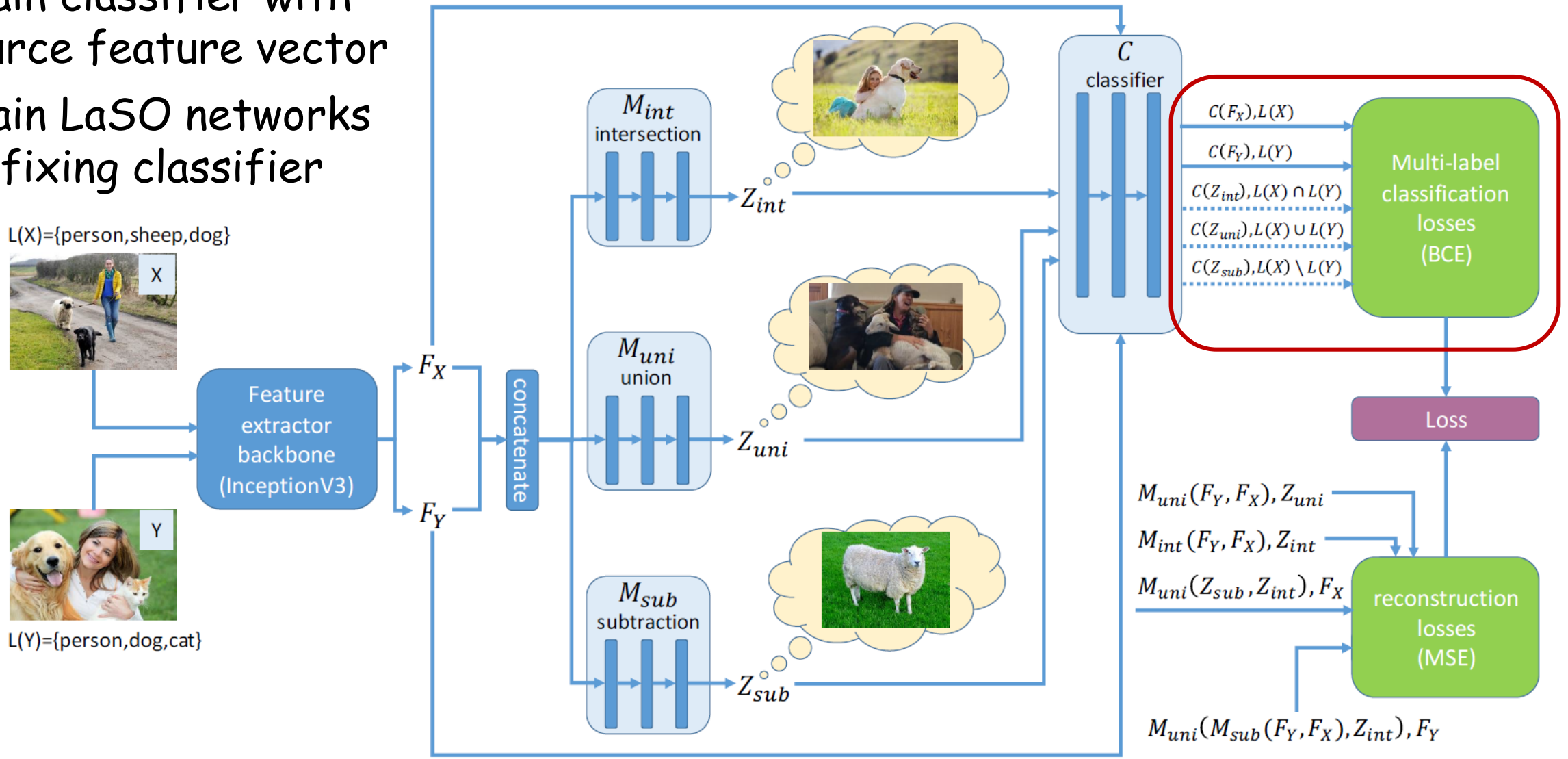
M_{sub} : a model that can implicitly remove concepts present in one sample from another sample.

M_{uni}

Framework

① Train classifier with source feature vector

② Train LaSO networks by fixing classifier



Implementation: Classification Loss

$$\sigma(x) = (1 + \exp(x))^{-1}$$

Use the Binary Cross-Entropy (BCE) multi-label classification loss to train the classifier C and the

$$BCE(s, l) = -\sum_i l_i \log \sigma(s_i) + (1 - l_i) \log(1 - \sigma(s_i))$$

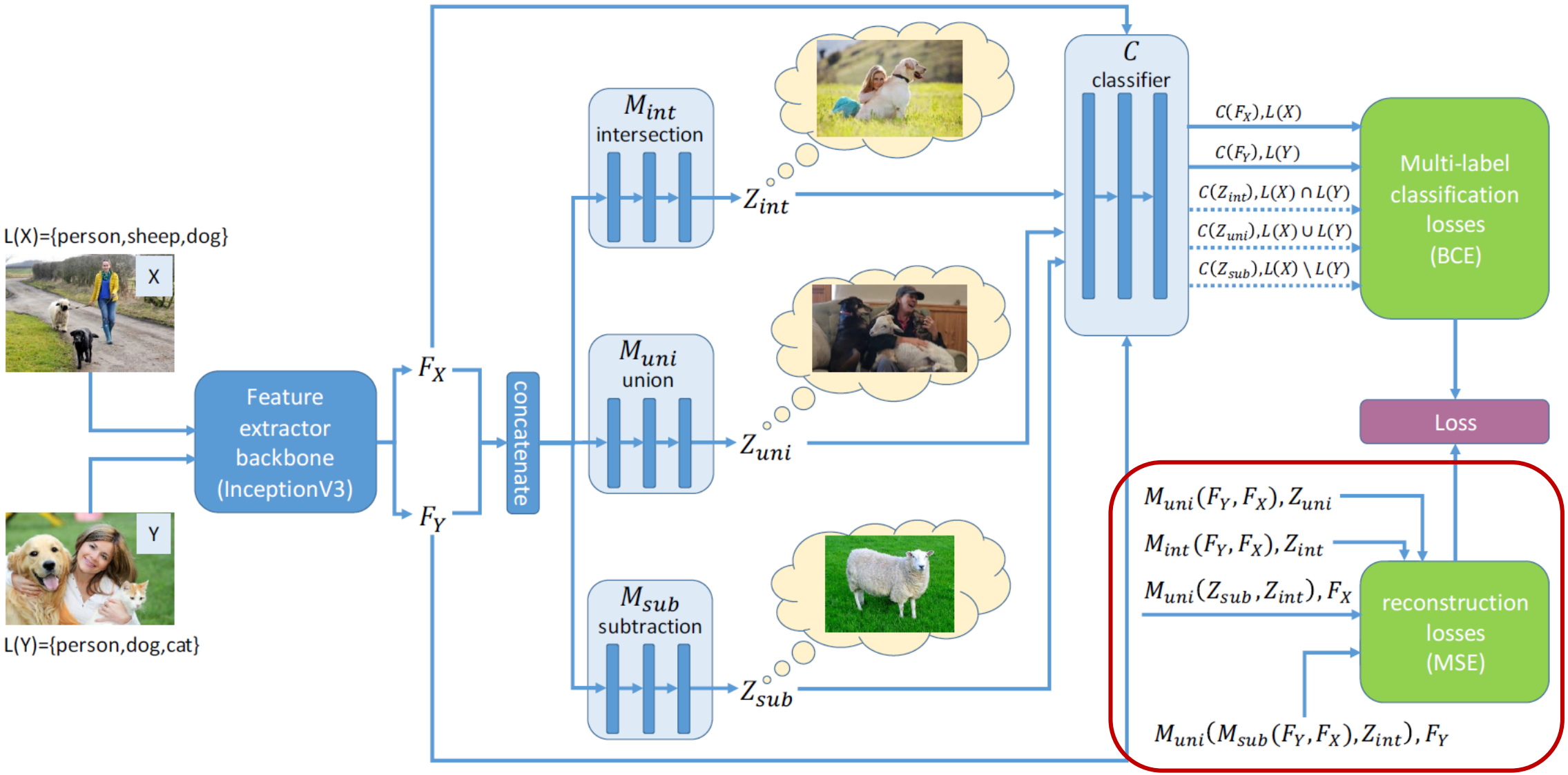
The classifier C is trained by using the combination of the losses from the source feature vectors

$$C_{loss} = BCE(C(F_X), L(X)) + BCE(C(F_Y), L(Y))$$

The LaSO networks are trained using the following loss by fixing the classifier

$$\begin{aligned} LaSO_{loss} = & BCE(C(Z_{int}), L(X) \cap L(Y)) + \\ & BCE(C(Z_{uni}), L(X) \cup L(Y)) + \\ & BCE(C(Z_{sub}), L(X) \setminus L(Y)) \end{aligned}$$

Framework


L(X)={person,sheep,dog}
X
Y
L(Y)={person,dog,cat}
Feature extractor backbone (InceptionV3)
F_X
F_Y
concatenate
M_{int} intersection
M_{uni} union
M_{sub} subtraction
Z_{int}
Z_{uni}
Z_{sub}
C classifier
$C(F_X), L(X)$
$C(F_Y), L(Y)$
$C(Z_{int}), L(X) \cap L(Y)$
$C(Z_{uni}), L(X) \cup L(Y)$
$C(Z_{sub}), L(X) \setminus L(Y)$
Multi-label classification losses (BCE)
Loss
$M_{uni}(F_Y, F_X), Z_{uni}$
$M_{int}(F_Y, F_X), Z_{int}$
$M_{uni}(Z_{sub}, Z_{int}), F_X$
$M_{uni}(M_{sub}(F_Y, F_X), Z_{int}), F_Y$
reconstruction losses (MSE)

Implementation: Reconstruction Loss

The loss is used to enforce symmetry for the symmetric *intersection* and *union* operations

$$R_{loss}^{sym} = \frac{1}{n}\| Z_{int} - M_{int}\left(F_Y, F_X\right)\|_2 + \frac{1}{n}\| Z_{uni} - M_{uni}\left(F_Y, F_X\right)\|_2$$

Use simple expression

$$R_{loss}^{mc} = \frac{1}{n}\| F_X - M_{uni}\left(Z_{sub}, Z_{int}\right)\|_2^2 + \frac{1}{n}\| F_Y - M_{uni}\left(M_{sub}(F_Y, F_X), Z_{int}\right)\|_2^2$$

Experiment

MS-COCO

- Training and validation sets
- 64 'seen' and 16 'unseen' classes
- Training

① Filter the training set leaving only images that did not contain any of the 16 unseen class

② Pre-train the feature extractor backbone separately as a multi-label classifier on filtered sets.

③ Use the filtered set to train feature extractor backbone and the LaSO models.

Experiment: Evaluating Performance of LaSO

Use a pre-trained classifier to test the LaSO networks.

	64 seen classes	16 unseen classes
intersection	77	48
union	80	61
subtraction	43	14
original (non-manipulated) feature vectors	75	79

Table 1. Evaluating feature vectors synthesized by the LaSO networks using the *classification* performance on the 64 *seen* and on the 16 *unseen* MS-COCO categories. Classification is performed w.r.t. the expected label set after each type of operation, and on the original feature vectors for reference. All tests are performed on the MS-COCO validation set, not used for training. Numbers are in mAP %.

Experiment: Evaluating Performance of LaSO

Use retrieval tests to evaluate the synthesized feature vectors directly without any classifier.

	64 seen classes			16 unseen classes		
	top-1	top-3	top-5	top-1	top-3	top-5
intersection	0.7	0.79	0.82	0.47	0.71	0.78
union	0.61	0.71	0.74	0.44	0.64	0.71
subtraction	0.19	0.32	0.4	0.21	0.4	0.51
original (non-manipulated) feature vectors	0.56	0.72	0.76	0.56	0.75	0.81

Table 2. Evaluating feature vectors synthesized by the LaSO networks using the *retrieval* performance on the 64 *seen* and on the 16 *unseen* MS-COCO categories (Sec. 4.1.1). Retrieval quality is measured w.r.t. the expected label set after each type of operation. All tests are performed on the MS-COCO validation set, not used for training. Numbers are *mean Intersection over Union* (mIoU) between the label sets of the retrieved samples and the expected label set, the mean is taken over the different queries. The top-k averages the maximum IoU obtained among closest k retrieved samples. In order to assess the expected range of retrieval performance in feature space $\mathcal{F}$, we also provide a reference of the same quality measurement for retrieval using the the original non-manipulated feature vectors.

Experiment: Evaluating Performance of LaSO

Experiment: Analytic approximations to set operations

Operator	Expression 1	Expression 2
Union	$F_X + F_Y$	$\max(F_X, F_Y)$
Intersection	$F_X \cdot F_Y$	$\min(F_X, F_Y)$
Subtraction	$F_X - F_Y$	$\text{ReLU}(F_X - F_Y)$

dataset	method	subtraction	intersection	union
MS-COCO	analytic	29.0	74.7	76.5
	learned	**43.0**	**77.0**	**80.0**
CelebA	analytic	37.0	**52.0**	47.0
	learned	**69.0**	48.0	**75**

Table 3. **Ablation study:** comparing the learned operators with analytic alternatives. All numbers are in mAP %.

Experiment: Multi-label few-shot classification

Baselines:

1. Training directly on the small labeled set
2. Using standard image augmentation while training on the small labeled set
3. Using the mixUp augmentation technique

	1-shot	5-shot
B1: no augmentation	39.2	49.4
B2: basic aug.	39.2	52.7
B3: mixUP aug.	40.2	54.0
analytic intersection aug.	40.7	55.4
analytic union aug.	44.5	55.6
learned intersection aug.	40.5	57.2
learned union aug.	**45.3**	**58.1**

Table 4. Multi-label few-shot mAP (in %) on 16 unseen categories from MS-COCO. The feature extractor and the LaSO networks are trained on the remaining 64 MS-COCO categories. Average of 10 runs are reported, tested on the entire MS-COCO test set. MixUP baseline uses the original code of [40].

Random pairs of examples from the small training set (1 or 5-shot x 16 classes) were used for label-set manipulations.

Thanks