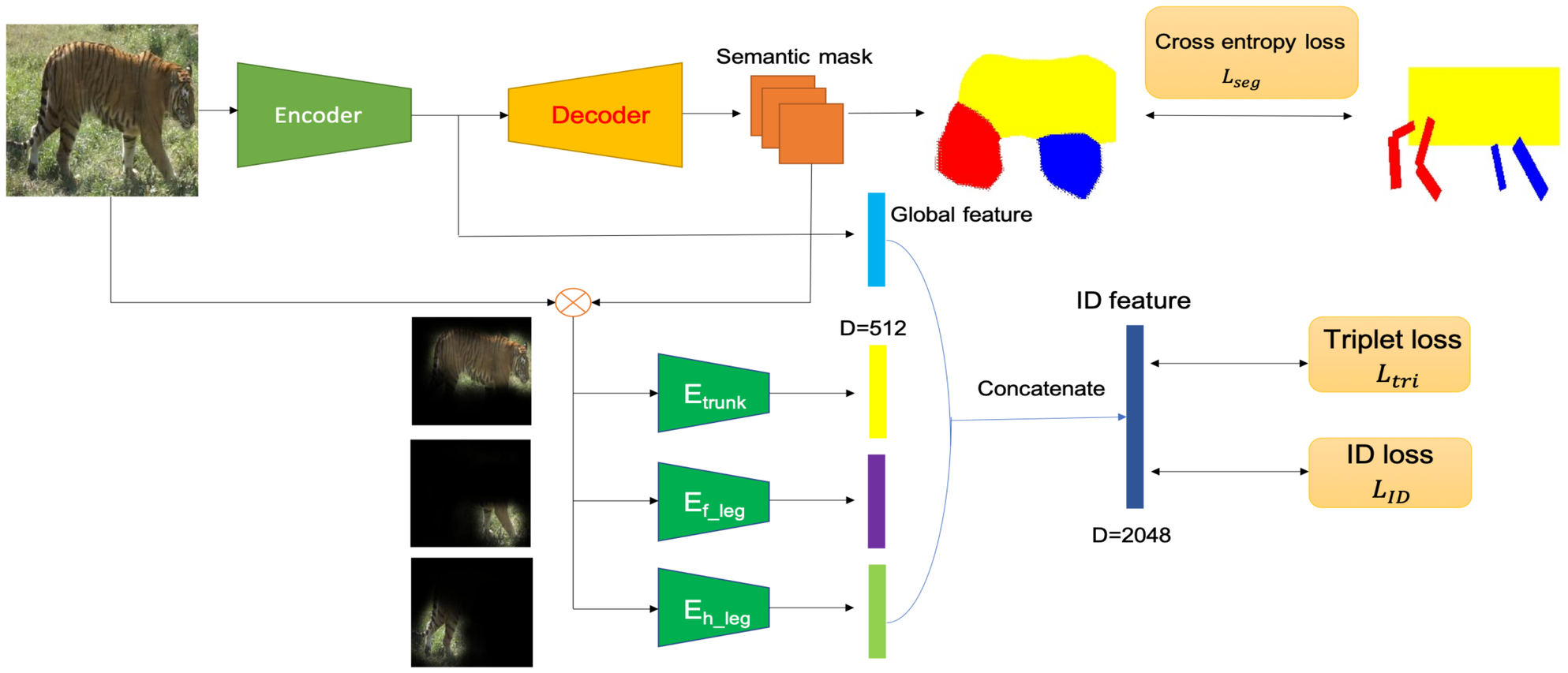
DLCV final challenge-Tiger Re-ID
Team 1

Member: 詹書愷、林奕廷、陳泓均、潘彥銘



Introduction

a. Motivation: For Re-ID, information of "tiger" regions is more important than that of background.

b. Idea: Semantic maps can be used as attention maps.

c. Approach: An end-to-end architecture that utilizes poses to generate coarse segmentation maps, then uses the maps as attention guides for Re-ID training.

Proposed Model

a. Semantic Mask Generation

- Use key-points to generate box-shaped masks.
- 4 classes: background, trunk, h_leg, f_leg

b. Architecture

i. Pose-Guided Segmentation

- *L_seg:* cross entropy loss for segmentation
- Extra penalty for leg class mis-prediction

ii. Semantic Attention Re-ID

- Segmentation model returns distributions for n classes
- Attention for each part = original image x semantic distribution of the corresponding class
- *L_id:* cross entropy loss for classification
- *L_tri:* hard triplet loss (L2)

iii. $L = L\_seg + \alpha \times L\_tri + ß \times L\_id$

Implementation Details

a. Dataset - ATRW Plain Re-ID

- Contains key-points and id labels for each image
- We generate semantic mask ground truths for each image.

b. Training Setting

We use ResNet18 as our model's backbone. Each time we use batch size = 64 and epochs=50 to train our model. And our optimizer's hyper-parameters are as follow: learning rate=2e-5, decay with "poly" policy; α: 0.1; ß: 0.05.

Experiments and results

a. Different ways to concat global & local fetures:

I. Three pathways: global, trunk and the legs. (For the leg part, extract features from the 6 leg parts and concat. the features)

II. Three pathways: global, trunk and the legs. (For the leg part we consider all leg parts as the same label in the segmentation map)

III. Four pathways: global, trunk, front legs, hind legs. (Four labels in segmentation map)

IV. Five pathways: global, trunk, front legs, hind thigh, hind shank. (Five labels in segmentation map)

And the performance of the four different structure is listed below:

Structure	I	II	III	IV
Val. Acc.	0.6190	0.6054	0.6611	0.6416

b. Different backbone

First, we use different layer of ResNet to be our backbone (with batch size = 32):

Layers	18	34	50
Val. Acc.	0.6069	0.6446	0.6355

Second, we choose two kinds of network whose parameters are similar to be our backbone (with batch size = 32):

Network	ResNet50	ResNeXt50
Val. Acc.	0.6355	0.6295

c. Different definition of label

We use two kinds of definition of label:

- Soft label: Our segmentation model will output 4 heatmaps for 4 classes, which include background, trunk, front legs and hind legs. Each heatmap represents the probability distribution of the class.
- Hard label: Based on the soft label, for each pixel , we let the highest value (probability) among 4 classes be 1, others be 0.

Definition	Soft	Hard
Val. Acc.	0.6069	0.4955

Reference

1. Amur tiger re-identification in the wild
2. Pose-Guided Complementary Features Learning for Amur Tiger Re-Identification
3.Human Semantic Parsing for Person Re-identification