DLCV Final Challenge 1: Amur Tiger Re-ID

Tzu-Yuan Lin, Ywuan-Chai Chong

Chih-Chun Yang, Meng-Lin Huang

National Taiwan University

Introduction

Tiger Re-IDentification (Re-ID) aims to retrieve all the images containing the same tiger as the query in the database. We surveyed several papers and implemented a model which is proposed in Human Re-ID field and looks fit for this challenge [1]. Also, we refer to the method of the first place [2] in the original ICCV workshop and that is the best of all models we've implemented. In addition, we proposed a new model architecture (FRNet) to solve this task and expected that it can learn disentanglement feature. Successfully, the new model passed the baseline.

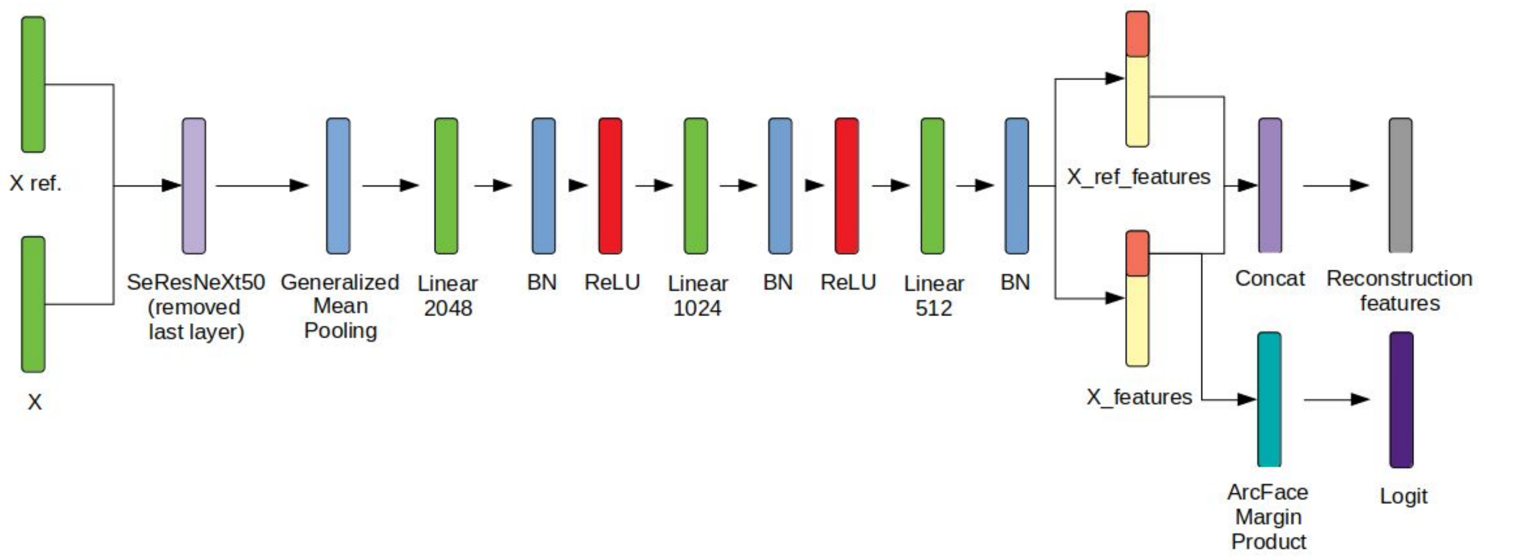
Methods

Data Augmentation

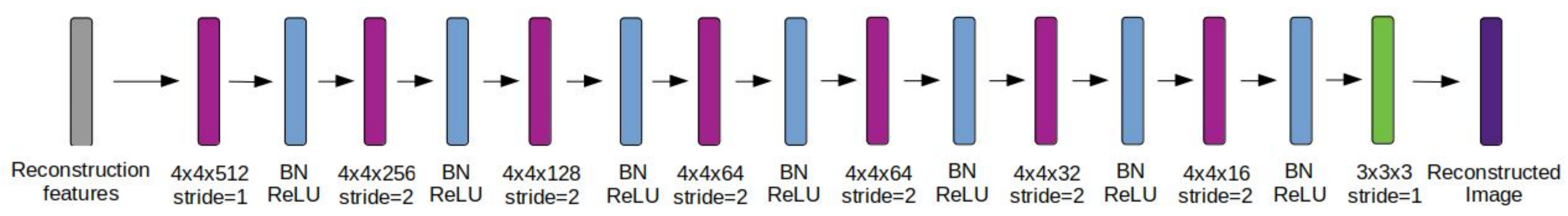
Augmentation	Details
Rotation	range = [-10, 10] (degree)
Color Jitter	brightness = 0.2, saturation = 0.2, contrast = 0.2, hue = 0.2
Random Erasing [3]	probability = 0.5, s_l = 0.02, s_h = 0.4, r_1 = 0.3
Random Horizontal Flip	with new ID given
Random Crop	size = [512, 512]
Pseudo Label	semi-supervised, label unlabeled data during model training

Table 1. Augmentation technique used.

Model Architecture



(a) Encoder of our model. X and X_ref are two different input images with the **same** label. X_features and X_ref_features are extracted features by encoder. We assume the extracted features includes the id information (red part) and other abstract information (yellow part). Then, we concatenated the id information of X_features and abstract information of X_ref_features to reconstruct image which is expected to be as similar to X ref. as possible. We also calculate Additive Angular Margin Loss (ArcFace) [7] by using id information of X_feaures to ensure the model is able to assign correct label to X.



(b) Decoder of ours model. This decoder is used for reconstructing image while giving a concatenated features.

Figure 1. FRNet architecture

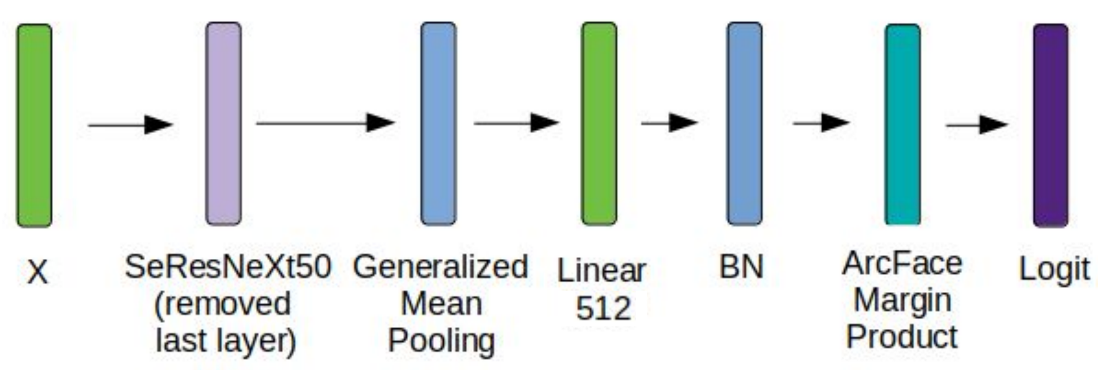


Figure 2. Our best model architecture.

Figure 2 is the architecture of our best model that we referred to the method of the first place model [2] in original competition and made some changes to fit our dataset. First, we used SeResNeXt50 as backbone to extract features of image and passed the features to a 512 dim. linear layer to do dimension reduction and transformation. Finally, we optimized the model with CE loss, ArcFace loss [7] and Triplet loss [6]. We also used Label Smoothing [5] to regularize the model and ID Re-Ranking [4] to get highest rank 1 score.

Experiments

Result

Model	Rank 1 (%)
ResNet152	71.39
SeResNet50	71.24
SeResNeXt50	71.63
ABD-Net [1] w/o ID Re-Ranking [4]	67.92
ABD-Net	73.49
FRNet (ours)	69.88
Our Best model	**76.81**

Table 2. Comparison of different model Rank 1 performance on queries provided by TA. FRNet refers to **Fig. 1**, Our best model refers to **Fig. 2**

Ablation Test

Best model w/o	Rank 1 (%)
- Label Smoothing [5]	73.99
- Triplet Loss [6]	73.27
- ArcFace [7]	69.05
- ID Re-Ranking [4]	67.39

Table 3. Ablation study of best model.

Conclusion

We have tried several models and tricks, and following is our observations:

1. ArcFace and ID Re-Ranking are the most significant tricks according to the task.
2. Different backbone only has little differences.
3. Pseudo label and Random erasing would exacerbate model performance.
4. Our proposed FRNet model is expected to achieve feature disentanglement.

References

[1] *Tianlong Chen, et al.* ABD-Net: Attentive but Diverse Person Re-Identification. ICCV 2019.

[2] *Cen Liu, et al.* Part-Pose Guided Amur Tiger Re-Identification. ICCV Workshop 2019.

[3] *Hao Luo, et al.* Bag of Tricks and A Strong Baseline for Deep Person Re-identification. CVPR Workshop 2019.

[4] *Zhun Zhong, et al.* Re-ranking Person Re-identification with k-reciprocal Encoding. CVPR 2017.

[5] *Szegedy, et al.* Rethinking the Inception Architecture for Computer Vision. CVPR 2016.

[6] *Florian Schroff, et al.* FaceNet: A Unified Embedding for Face Recognition and Clustering. CVPR 2015.

[7] *Jiankang Deng, et al.* ArcFace: Additive Angular Margin Loss for Deep Face Recognition. CVPR 2019.