



ClothingID for Clothing Image Classification and Retrieval

Student: ZHAN, Huangying Supervisor: Prof. WANG, Xiaogang
The Chinese University of Hong Kong

Introduction

Clothing plays an important role in our daily life. There is rich information carried by clothing (e.g. gender, age group, nationality...). There are also many important applications related to clothing (e.g. online shopping, human recognition). In this project, we proposed a new training strategy in deep learning. We introduced a large clothing class identification task (ClothingID) as a pre-training stage in deep convolutional neural network (DCNN), in order to improve performance of clothing image related tasks. Clothing image classification and clothing retrieval are used to evaluate our ClothingID model.

Methodology

A pre-training stage allows the DCNN model to learn some good features before training on the target dataset. Conventionally, people use ImageNet model as the pre-trained model for their target task. After pre-training stage, the pre-trained model is trained again on target dataset. This technique is called fine-tuning. Fine-tuning ImageNet typically can result in an acceptable performance. ImageNet model has an objective to classify different objects into 1000 categories. There are many useful and good features learnt by ImageNet. However, ImageNet is not specifically designed for clothing image. ClothingID is introduced to allow model to learn clothing related features such that a better initialization is provided to our target tasks.

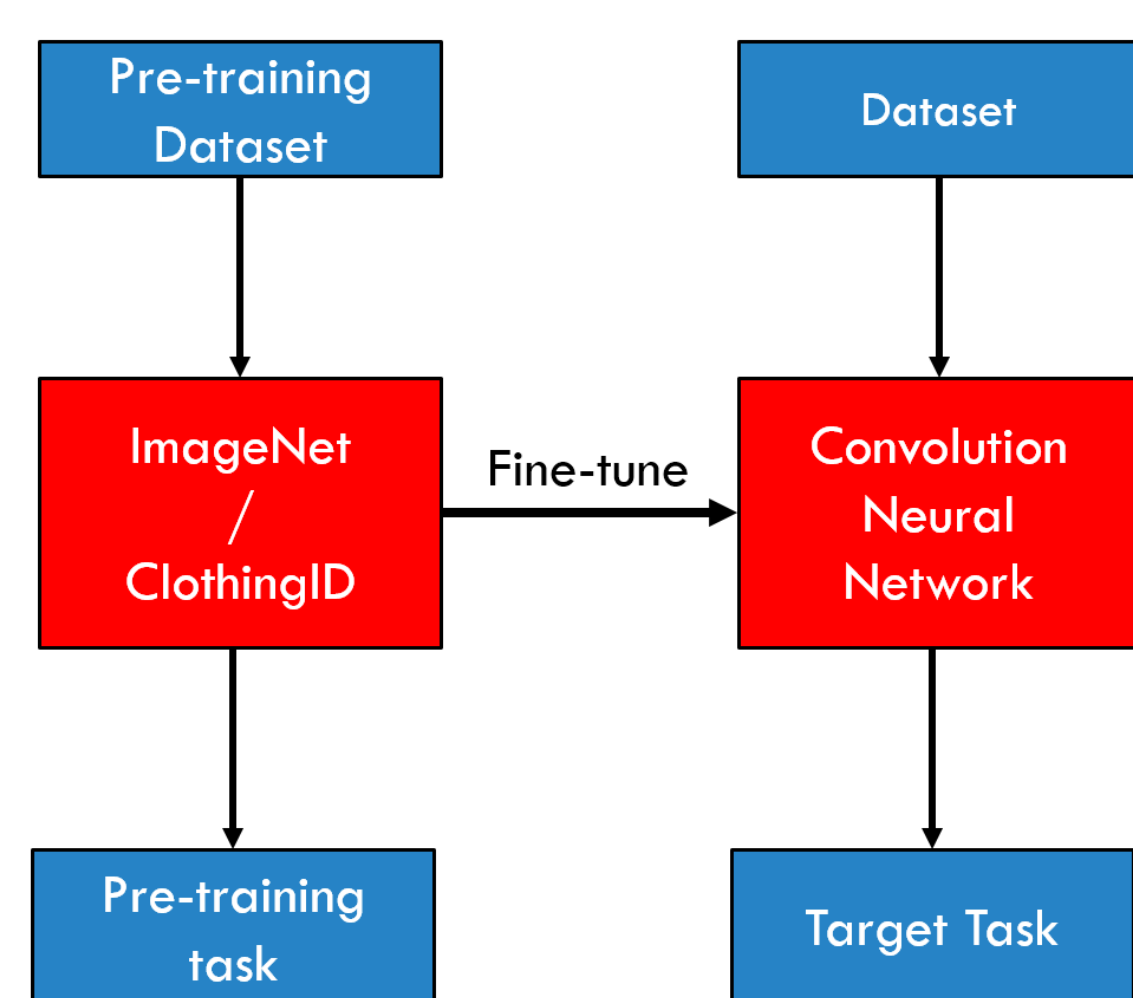


Figure 1: Fine-tuning & General training strategy

Basically, the objective of this project is to show that, fine-tuning ClothingID to clothing image related tasks can perform better than fin-tuning ImageNet to those tasks.

Experiment

DCNN Architecture

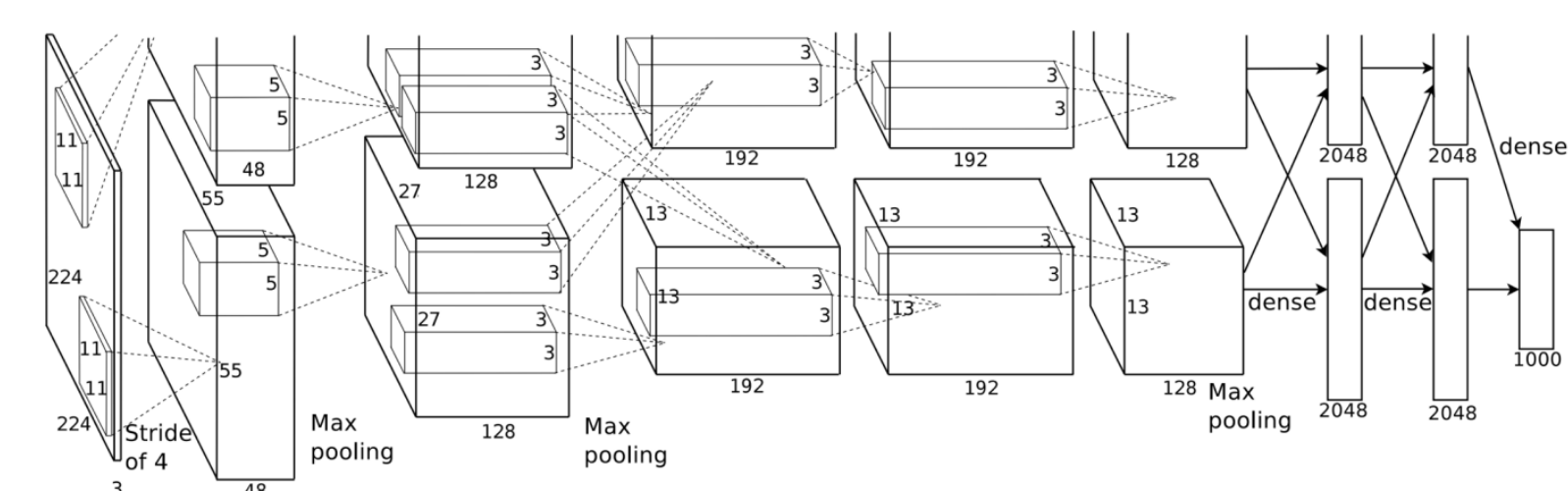


Figure 2: AlexNet Architecture

Dataset

	ClothingID	Classification-14	Classification-26	Retrieval
Categories #	22,000	14	26	/
Training images #	526,647	52,431	41,097	19,903
Test images #	44,000	7,000	2,600	20,053
Total	570,467	59,431	43,697	39,956

Table 1: Dataset statistics

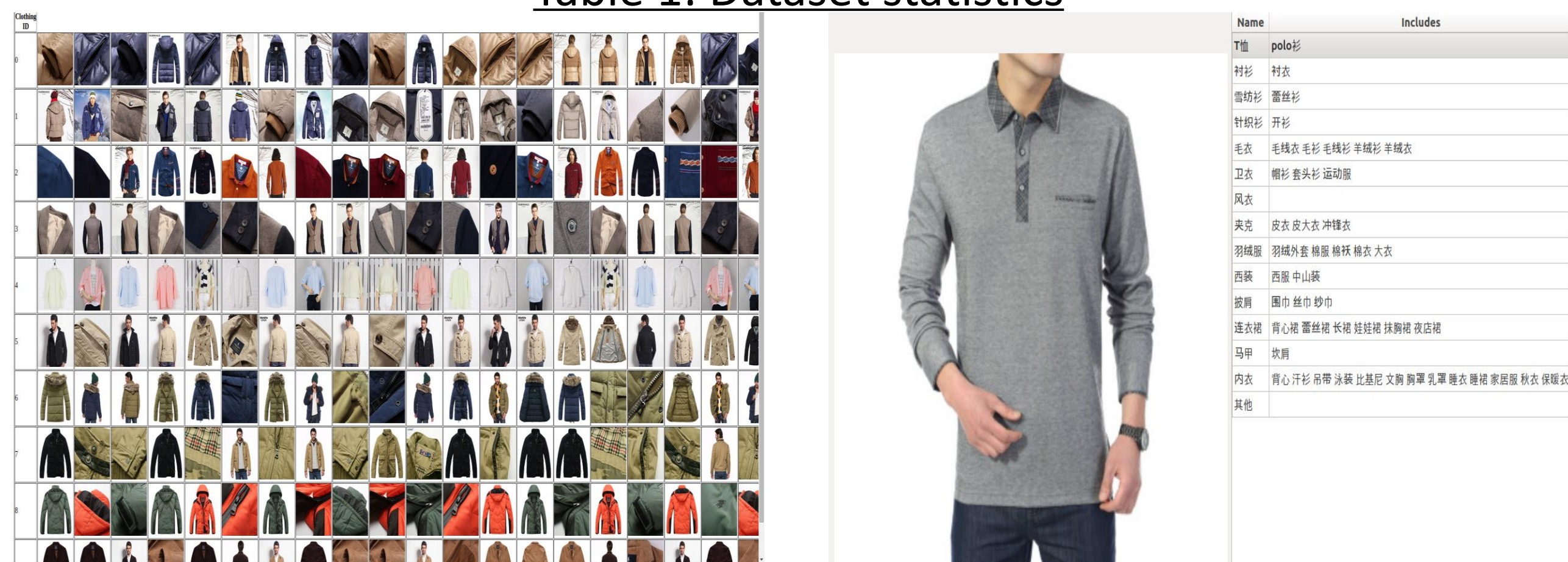


Figure 3,4: Dataset example (Left: ClothingID; Right: Classification-14)

Experiment

Dataset

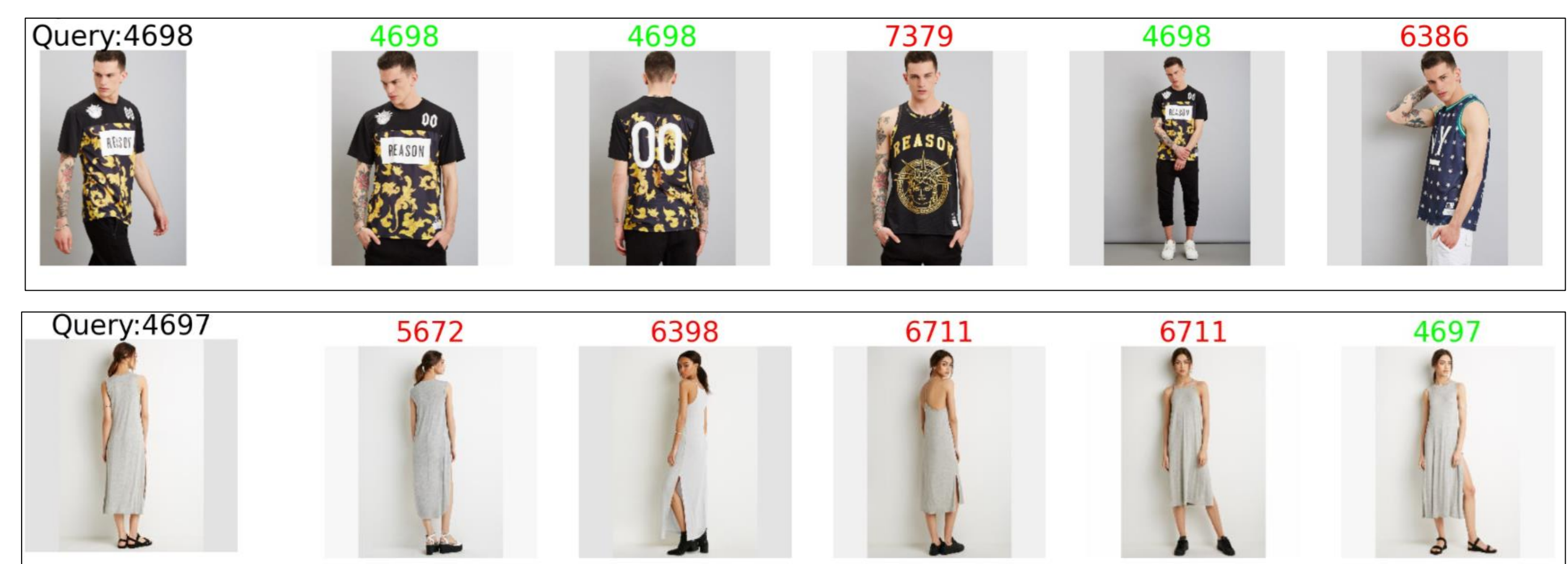


Figure 5: Dataset example (Retrieval)

Result

Classification

	Classification-14	Classification-26
Fine-tune ImageNet	0.668	0.795
Fine-tune ClothingID	0.670	0.799

Table 2: Classification result

Retrieval

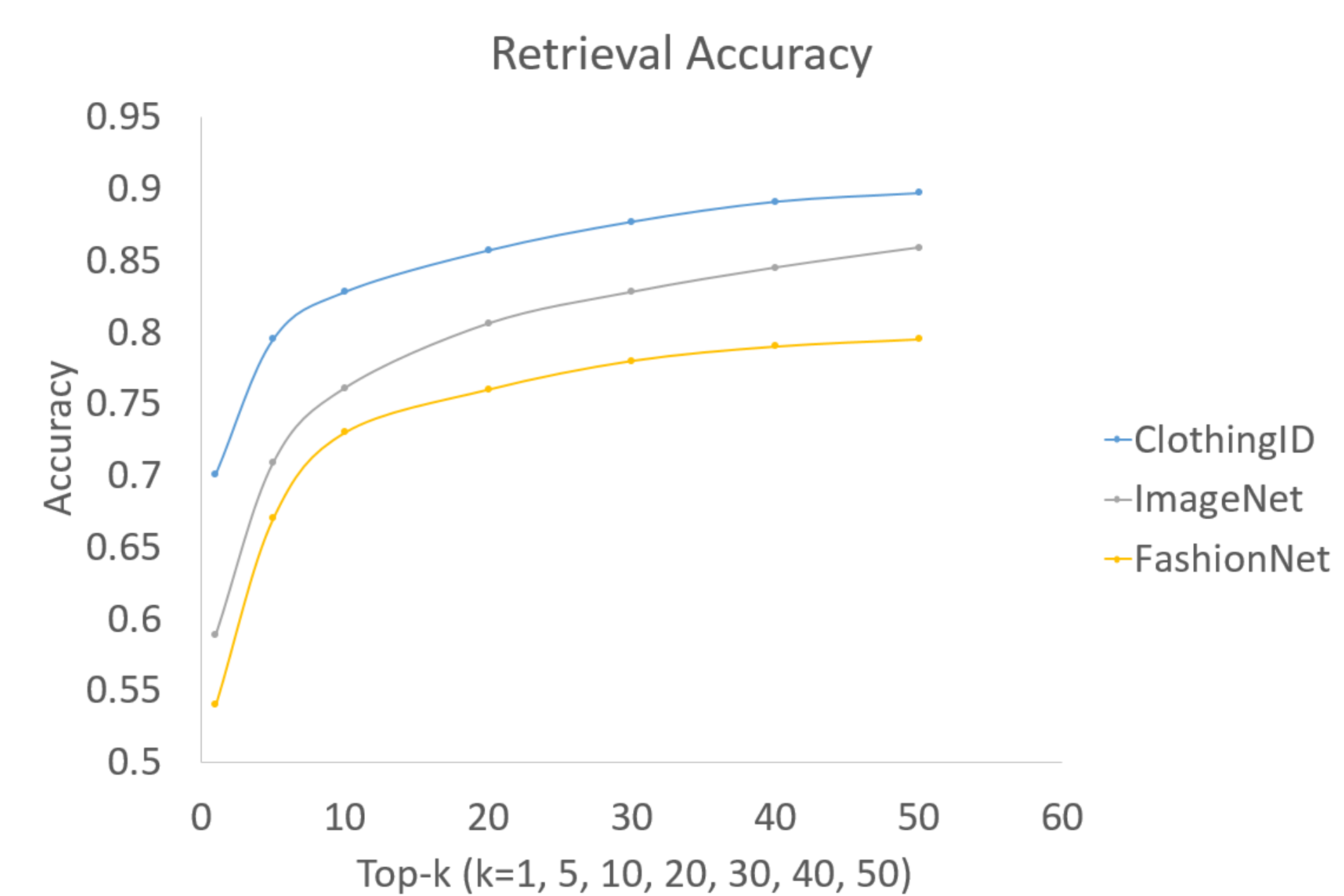


Figure 6: Retrieval result

Discussion

Future work

- Different network architecture (i.e. complex/deeper network)
- Test model on more difficult retrieval task (Customer-to-shop cloth retrieval)
- Model improvement (e.g. localization of main object; fine-graining)

Conclusions

Dataset construction

- A large scale clothing identification dataset (ClothingID)
- A clothing classification dataset (14 categories)

Model training

- ClothingID model as a good pre-trained model for clothing image tasks
- Classification models
- Retrieval model

Method

In this project, we proposed a possible pre-training strategy. Our result shows that introducing a large class identification as pre-training stage allows the model to learn some useful features related to a specific category (clothing) and then to improve performance of related tasks. This method can be generalized to other objects rather than clothing.

Acknowledgement

I would like to thank my supervisor Prof. Wang for his guidance and support in this project. He also provided some very good ideas during discussion. Also, I would like to thank Xiao Tong and Shen Yantao. They gave me a lot of supports in this project.

References

1. Xiaogang, Wang, "Deep Learning in Computer Vision", 2015
2. I. S. G. E. H. Alex Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS Proceeding*, 2012.
3. A. Z. K. Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *ILSVRC-2014*, 2014.
4. L. Fei-Fei, "CS231n: Convolutional Neural Networks for Visual Recognition," Stanford University, Jan 2015. [Online]. Available: <http://cs231n.stanford.edu/index.html>. [Accessed July 2015].
5. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," arXiv preprint arXiv:1408.5093, 2014.
6. Z. Liu and X. Wang, "CompFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," 2015.