2018 DLCV Final Challenge: Few-Shot Object Recognition

Yen-Ting Liu, Tsung-Shan Yang, Guan-Shiuan Kuo, Po-Chun Hsu
劉彥廷、楊宗山、郭冠軒、許博竣
National Taiwan University

Group 15

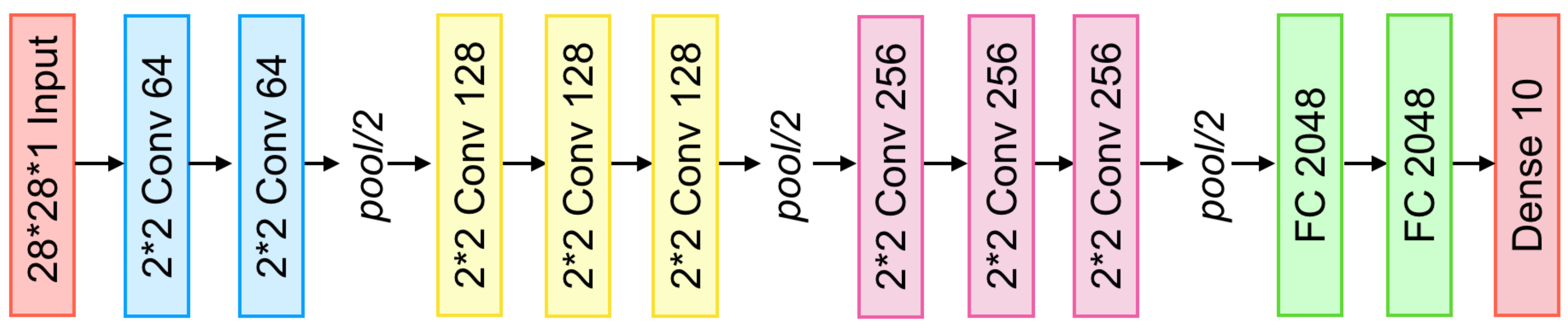
(/ ‵Д′) / ~╧╧

Introduction

Few-shot learning aims to recognize novel visual categories from very few labelled examples. In this project, we are required to solve two tasks, including small data supervised learning and one / few-shot learning. We will demonstrate several methods which are commonly used to solve the few-shot classification problems.

Little Data Supervised Learning

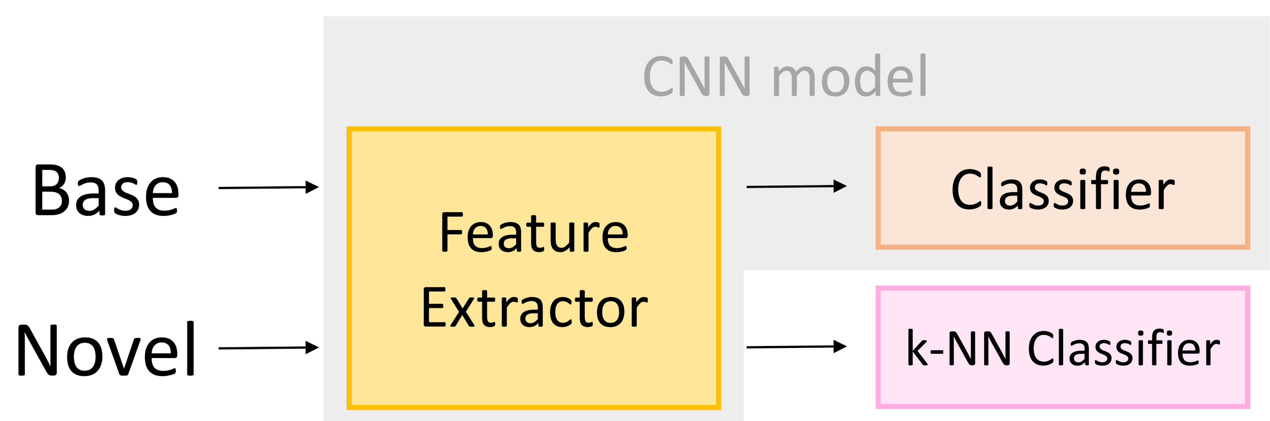
We adopt a VGG16-like model on the 10-class classification task, with 200 labelled images for each class.



* Validation / Test Accuracy: 90.00 % / 85.19 %
* Baseline Accuracy: 79.90 %

CNN & k-Nearest Neighbors

We train a classification model on base classes data first, and generate the features of novel classes data by the feature extractor. Then, we calculate the average features of each class and train the k-NN classifier with them.

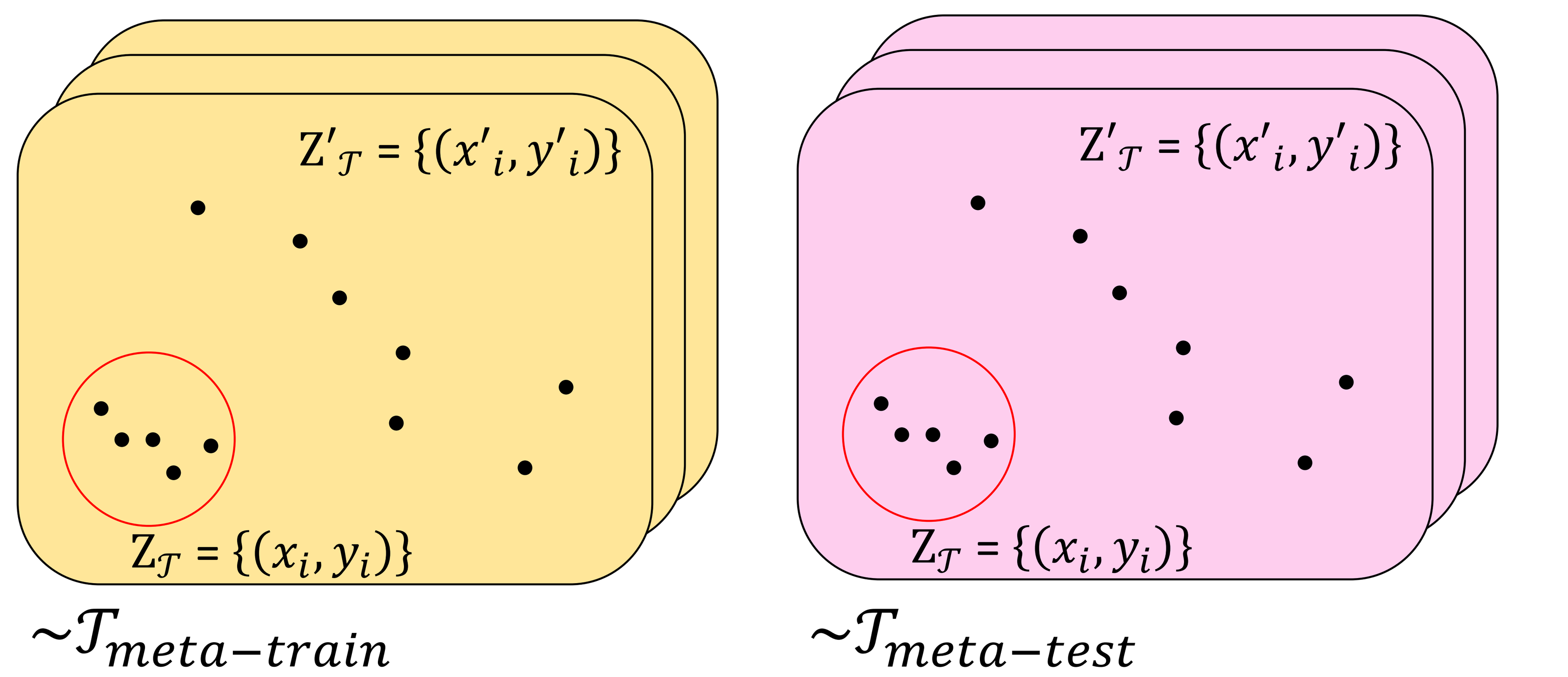


Meta-Learning

The goal of meta-learning is to enable a base learning algorithm to adapt to new tasks efficiently, by generalizing from a set of training tasks $\mathcal{T} \in \mathbb{T}$.

Training on a given task (base class), we are given access to training samples $Z_{\mathcal{T}} = \{(x_i, y_i)\} \sim \mathcal{T}_{meta-train}$ and test samples $Z'_{\mathcal{T}} = \{(x'_i, y'_i)\} \sim \mathcal{T}_{meta-train}$.

Adapting to another task (novel class), we can use $Z_{\mathcal{T}} = \{(x_i, y_i)\} \sim \mathcal{T}_{meta-test}$ to predict the y'_i from $Z'_{\mathcal{T}} = \{(x'_i, y'_i)\} \sim \mathcal{T}_{meta-test}$.



The following are meta-learning models:

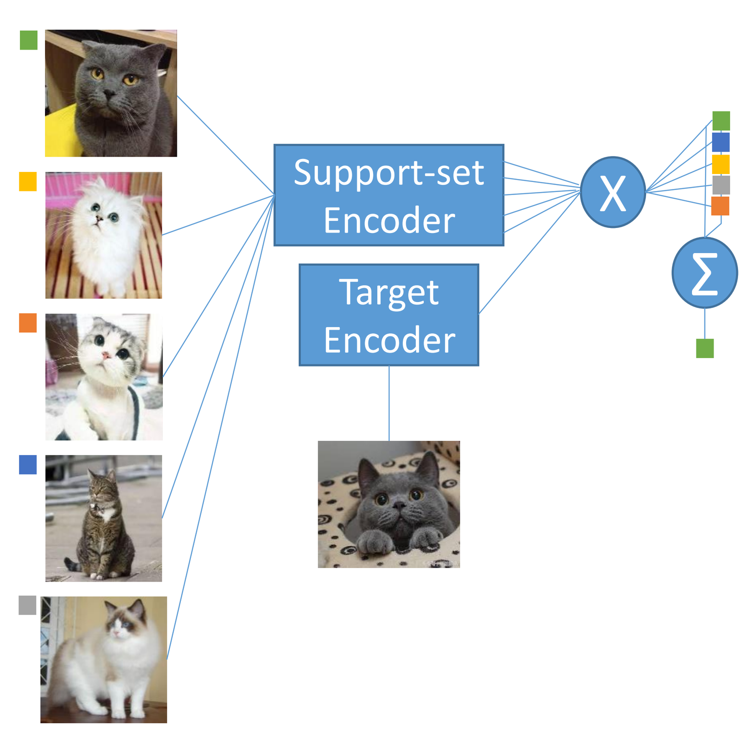
- Matching Network
- Relation Network

Matching Network

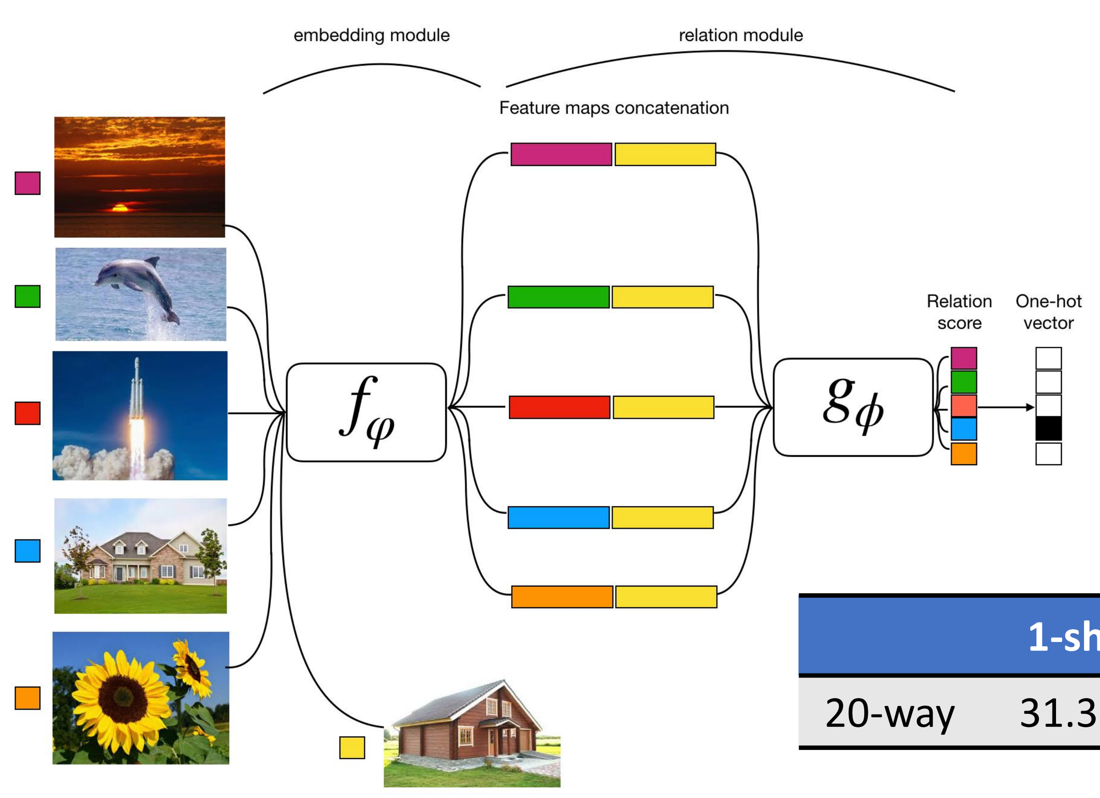
Matching Network calculate cosine similarities between the encoded target image and each encoded image in the support set.

At last, the target image can be classified to one of the class in the support set according to these similarities.

	1-shot	5-shot	10-shot
5-way	37.37%	47.32%	45.26%
10-way	24.31%	27.83%	24.05%
20-way	12.08%	13.87%	13.19%



Relation Network

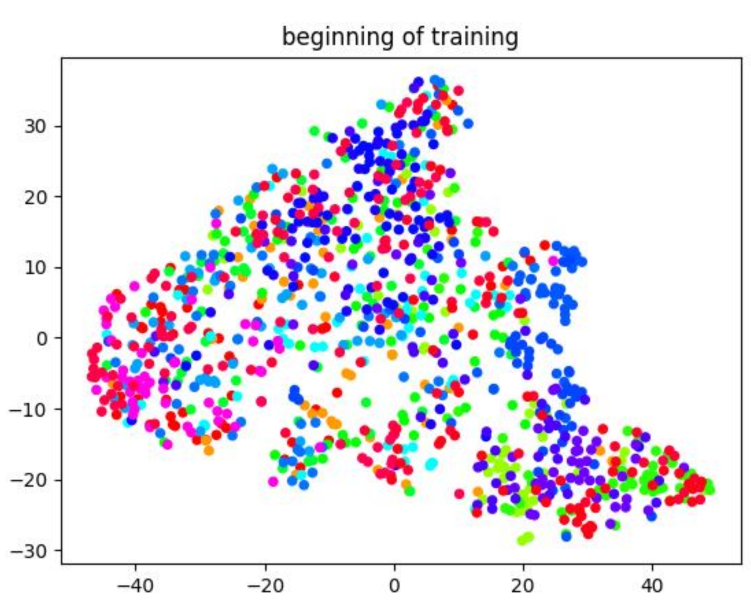


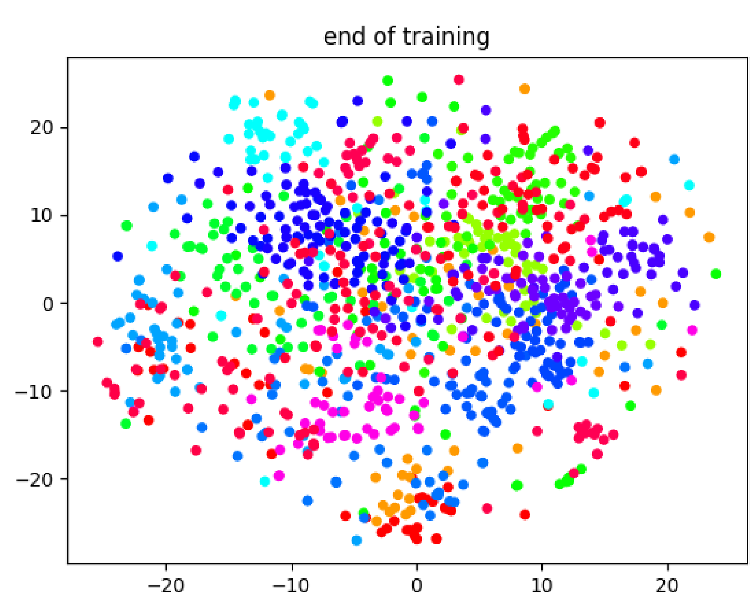
Relation Network learns an embedding module and deep non-linear distance metric through end-to-end training.

	1-shot	5-shot	10-shot
20-way	31.31%	48.44%	52.66%

Experiments

Visualization of features extracted from CNN model





20-way classification accuracies for each model

Model	Accuracy 1-shot	5-shot	10-shot
CNN & k-NN	**34.60%**	**49.85%**	**59.05%**
Matching Network	12.08%	13.87%	13.19%
Relation Network	31.31%	48.44%	52.66%