1-012 学生発表

Image synthesis for One-shot Classification with Triplet Network

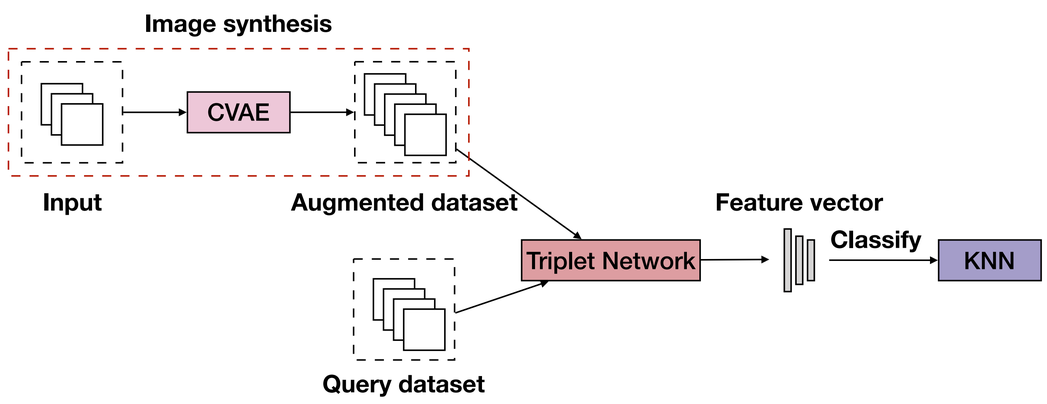
周 牧 (筑波大)　谷村 勇輔 (産総研)　中田 秀基 (産総研)

Abstract

- One-shot learning aims to learn information from one, or only a few training dataset, our goal is to predict the query images by giving only single example of each class, where we do data augmentation to synthesis more images.
- We use **Conditional VAE** to synthesize more images to augment the scarce dataset.
- This approach improves the one-shot classification problem.

Method

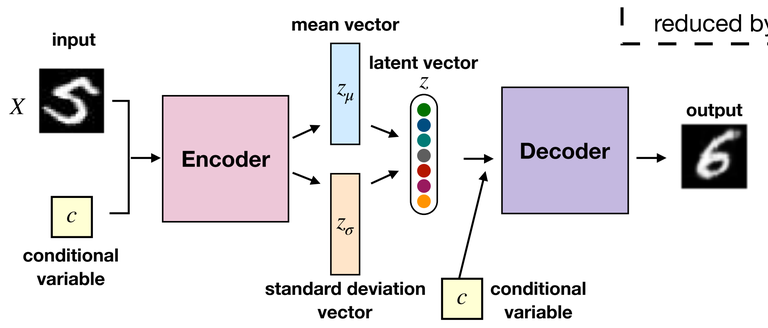
Overview



Prediction procedure

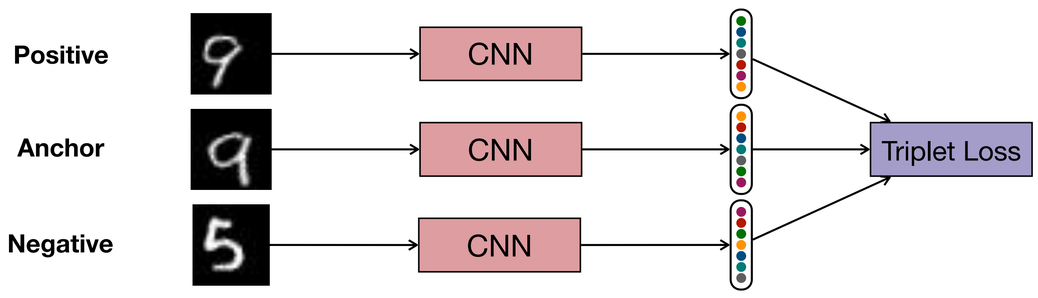
2.1 Conditional Variational autoencoder

2 types of input as conditional variable:
- Label of digits.
- Feature of digit images reduced by PCA.



Conditional variational autoencoder

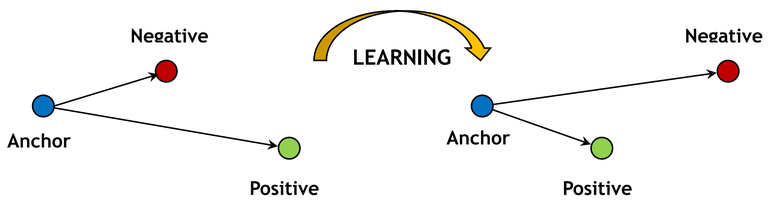
2.2 Triplet network



Triplet Network Model

2.3 Triplet loss function

- $L(A,P,N) = max(||f(A)-f(P)||^2 - ||f(A)-f(N)||^2 + \alpha, 0)$



The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and the maximizes the distance between the *anchor* and a *negative* of a different identity[1].

Image Synthesis

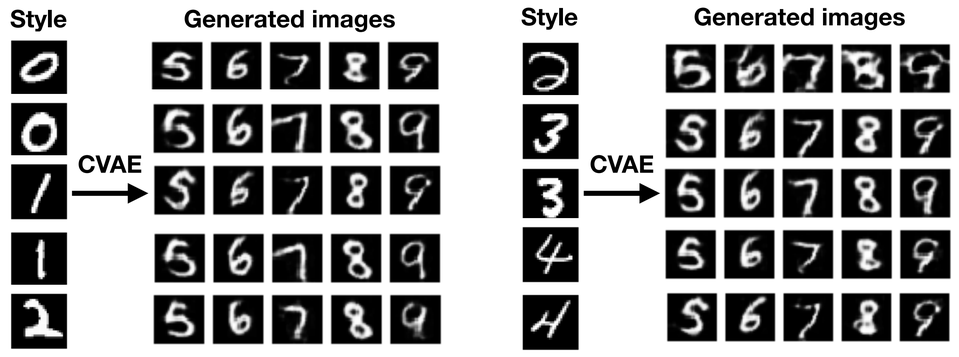
Initial Dataset

- For digit 0-4: choose all the data.
- For digit 5-9: choose **single** sample.

Augmented Dataset - Image synthesis with Conditional VAE

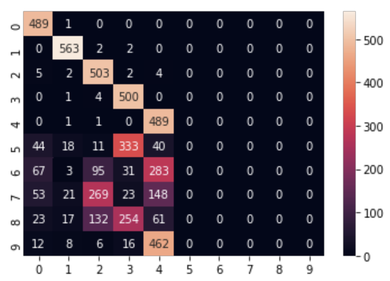
1. **PCA** based CVAE
   - Training: All digits 0-4.
   - Synthesize more 5-9 with encoded sample 5-9.

2. **Label** based CVAE
   - Training: All digits 0-4 + augmented digit 5-9 (rotation, flip, etc.).
   - Synthesize new 5-9 with style of 0-4.

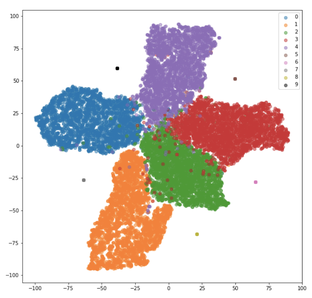


Result

1. **PCA** based CVAE

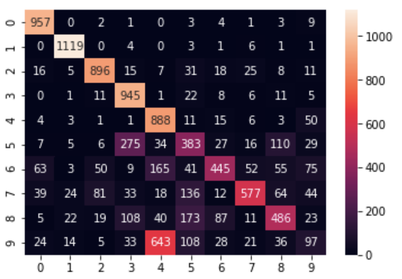


(a) Confusion matrix (quite bad).

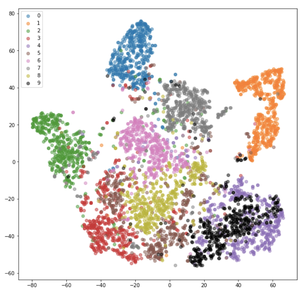


(b) T-SNE visualization

2. **Label** based CVAE



(c) Confusion matrix



(d) T-SNE visualization

Method (dataset)	Accuracy					
	5	6	7	8	9	Average
TripletNN (not Augmented)	14%	18%	11%	6%	0%	9.8%
CNN (Augmented)	25%	26%	16%	13%	13%	20.8%
TripletNN (Augmented)	42%	56%	66%	14%	14%	46.8%
TripletNN (Label-CVAE)	50%	74%	69%	10%	10%	56.4%

Result for one-shot classification.

Discussion

- Label-based CVAE improve the classification accuracy.
- Try to improve PCA-based CVAE model to generate more reliable and multiform images with variation using single sample.
- Limitation: classifier cannot separate digit 4 and digit 9.

Reference

[1] F. Schroff, D. Kalenichenko, J. Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering.

Acknowledgement

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).
This work was supported by JSPS KAKENHI Grant Number 19K11994.