# VA-NN: A Proposed Machine Learning Model with Limited Labelled Constraint
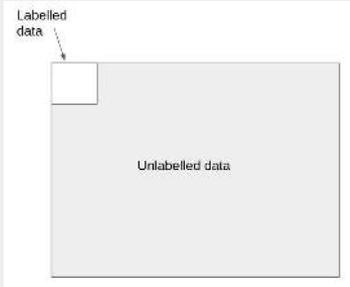
Thanh Son Phung, Chris Armen
Department of Computer Science, Trinity College

## 1) Problem Statement

- Need to classify/label data, e.g., cat images with label 'cat'.

- Labelling data by hand takes a long time.
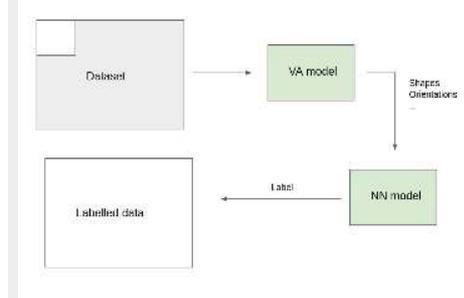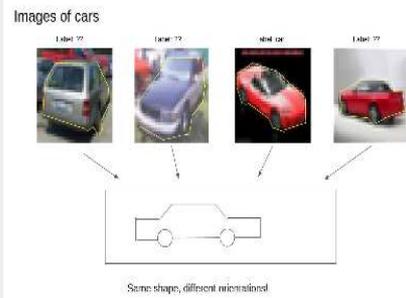
- Goal is to label all data.



## 2) Observations and Underlying Idea

- Most ML applications only use labelled data when training model -> waste unlabelled data!

- Solution: utilize unlabelled data when training model.

- Benefit: - easy to train.
        - able to capture details of objects in images without using labels.
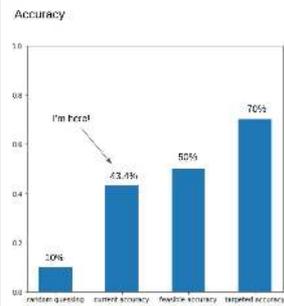                - utilize unlabelled data.



## 3) Model Architecture

- Train VA to extract features by reconstructing images.

### Variational Autoencoder



- Train NN to classify/label images based on encoded vectors.

### Neural Network



## 4) Experiment and Results

- Model run on CIFAR-10 dataset: 50,000 images to train VA, 3,000 to train NN, 7,000 tor validate and test entire model.

- **Accuracy: 43.4%**

- Problems:
  - low resolution images
  - lack of data

- Solutions:
  - change dataset
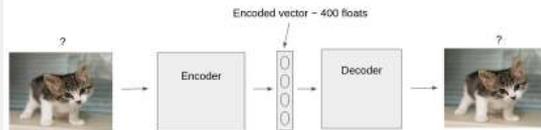  - rotate/crop images for more data points.