

## IN3060/INM460 Computer Vision Coursework report

- **Student name, ID and cohort:** Kai Sun Anson Lam [REDACTED]
- **Google Drive folder:** [https://drive.google.com/drive/folders/1cCBrdruSE-jTwrwGJtt0-Zxpvk1vTrQT?usp=share\\_link](https://drive.google.com/drive/folders/1cCBrdruSE-jTwrwGJtt0-Zxpvk1vTrQT?usp=share_link)

### Data

A provided dataset (with a training and testing set) and personal dataset are used to train and test various machine learning models to detect face covering. Each of them includes images and corresponding labels. There are 3 types of labels (0: no mask worn, 1: mask worn, and 2: mask worn improperly).

In the provided dataset, while the training set includes 2394 images and labels, the testing set has 458 images and labels. Both sets are extremely imbalanced, given that more than half of the images are labelled as 1, which may affect the application effectiveness of the trained models to real world data.

The personal dataset, on the other hand, is completely balanced. There are totally 24 images and labels, with exactly 8 images for each type of label.

### Implemented methods

#### Common Data Pre-processing and Training Process

Some general data pre-processing are performed on original images initially. Unsharp masking is used to sharpen images with emphasised details. Also, images are resized to standard inputs to models. Changing images from 3-colour channel to single grey channel is implemented within models, if necessary.

#### Main Method 1: SVM with SIFT

SVM with SIFT as feature descriptor is implemented in a bag of visual words approach. Since key points identified by SIFT might be high-dimensional, SVM with high capability to process them is ideal. 2 variants are developed with different kernels, which are 'rbf' and 'linear'. While C, gamma value, other hyperparameters can be tuned, default values are used.

#### Main Method 2: MLP with SIFT

MLP is similarly implemented with SIFT feature descriptor in a bag of visual words approach. Compared to SVM, more hyperparameters are tuned. For instance, for solver, 'adam' is chosen instead of 'sgd'; for hidden layer sizes, 3 (500, 250, 100) are used instead other combinations of 1 and 2 layers; for learning rate, 0.001 is preferred over 0.1 and 0.01.

#### Main Method 3: CNN

CNN is implemented as feature extractor with ResNet-18 and ResNet-50 pre-trained on ImageNet as 2 variants. With the same weights IMAGENET1K\_V1, both only have final layers being trained and updated. Similar to MLP, some hyperparameter have to be tuned. For example, for optimiser, 'AdamW' is selected rather than 'Adam' and 'SGD'; for criterion, cross entropy loss is used; for learning rate, 0.001 is chosen.

#### Implementation on Personal Dataset

Implementation of the MaskDetection function on personal dataset is similar to that to provided dataset, since images in personal dataset is cropped in advance and structure of both datasets are the same.

### Results

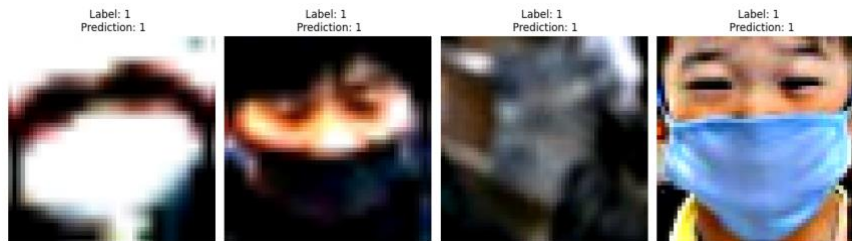
#### Results Overview

	Linear SVM with SIFT	RBF SVM with SIFT	MLP with SIFT	ResNet-18 CNN	ResNet-50 CNN
Accuracy (Provided Dataset)	high (0.846)	high (0.859)	high (0.805)	high (0.853)	very high (0.890)

Accuracy (Personal Dataset)	low (0.304)	low (0.304)	low (0.304)	low (0.391)	medium (0.565)
Speed	medium	medium	medium	low	very low
Model Size	medium	medium	medium	large	very large

- Overall accuracy on provided dataset is higher than that on personal dataset, possibly due different data distribution in datasets (provided dataset for both training and testing is very imbalanced, while the personal dataset is perfectly balanced)
- ResNet-50 CNN is the most accurate, but at the expense of very low speed and very large model size
- ResNet-50 CNN is the more accurate than ResNet-18 CNN in light more hidden layers to capture different level of abstraction in images
- Linear SVM with SIFT, RBF SVM with SIFT, and MLP with SIFT all have similar accuracy (for both datasets), speed, and model size

#### **Provided Dataset Result 1.1: Linear SVM with SIFT**



#### **Provided Dataset Result 1.2: RBF SVM with SIFT**



#### **Provided Dataset Result 2: MLP with SIFT**



#### **Provided Dataset Result 3.1: ResNet-18 CNN**



#### **Provided Dataset Result 3.2: ResNet-50 CNN**



### **Personal Dataset Result 1.1: Linear SVM with SIFT**



### **Personal Dataset Result 1.2: RBF SVM with SIFT**



### **Personal Dataset Result 2: MLP with SIFT**



### **Personal Dataset Result 3.1: ResNet-18 CNN**



### **Personal Dataset Result 3.2: ResNet-50 CNN**



## **References**

Lab 5, 6, 7, 8, 9