Person Image Re-Identification

Review 2

18BCE0715 - Sanjit C K S

What I'll be talking about

1. Overview
2. Dataset Description
3. Convolution Neural Networks and Image Processing
4. General Approach of Image Re-Identification
5. Proposed Methodology and Architecture
6. Algorithm
7. Environment Set Up
8. Implementation Progress

Overview

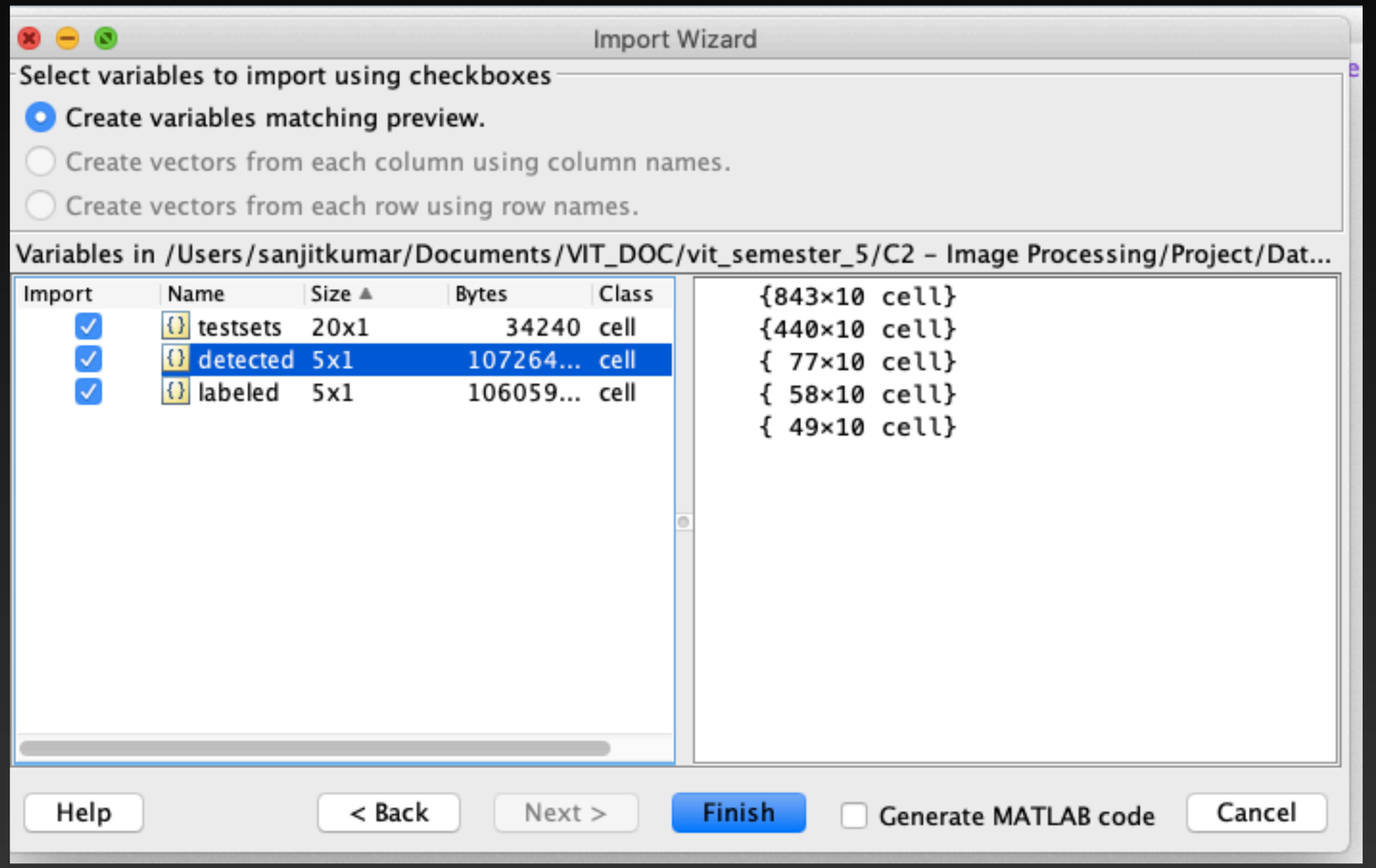
- Image re-identification with multiple cameras has been a major area of interest in the past 5 years.
- A network of cameras - different angles - monitor the same geographical area.
- In such a case, it is of great functional interest to identify/ label the same object/person in from all different cameras.
- The best way to do object detection with images is via a Convolutional Neural Network.

Dataset Description - CUHK03

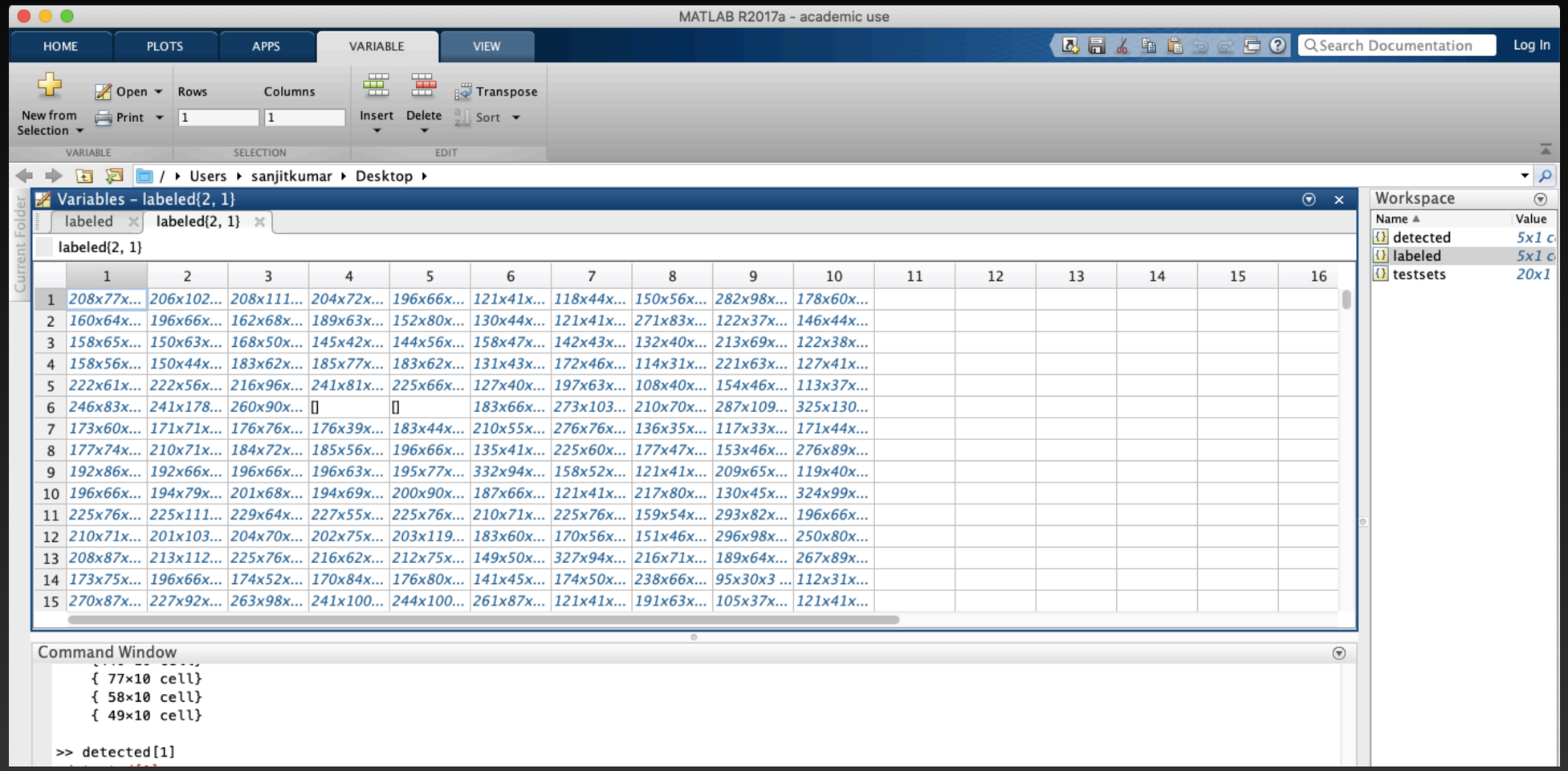
- Chinese University of Honk Kong
- CUHK03 is the first person re-identification dataset that is large enough for deep learning.
- **CUHK03 - 1,360 identities, 13,164 images, manually cropped + automatically detected**
- It provides the bounding boxes detected from DPM and manually labelling.
- 2014
- CUHK01 - 971 identities, 3884 images, manually cropped - 2012
- CUHK02 - 1816 identities, 7264 images, manually cropped - 2013

Dataset Description

Dataset Description



Dataset Description



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	208x77x...	206x102...	208x111...	204x72x...	196x66x...	121x41x...	118x44x...	150x56x...	282x98x...	178x60x...						
2	160x64x...	196x66x...	162x68x...	189x63x...	152x80x...	130x44x...	121x41x...	271x83x...	122x37x...	146x44x...						
3	158x65x...	150x63x...	168x50x...	145x42x...	144x56x...	158x47x...	142x43x...	132x40x...	213x69x...	122x38x...						
4	158x56x...	150x44x...	183x62x...	185x77x...	183x62x...	131x43x...	172x46x...	114x31x...	221x63x...	127x41x...						
5	222x61x...	222x56x...	216x96x...	241x81x...	225x66x...	127x40x...	197x63x...	108x40x...	154x46x...	113x37x...						
6	246x83x...	241x178...	260x90x...	[]	[]	183x66x...	273x103...	210x70x...	287x109...	325x130...						
7	173x60x...	171x71x...	176x76x...	176x39x...	183x44x...	210x55x...	276x76x...	136x35x...	117x33x...	171x44x...						
8	177x74x...	210x71x...	184x72x...	185x56x...	196x66x...	135x41x...	225x60x...	177x47x...	153x46x...	276x89x...						
9	192x86x...	192x66x...	196x66x...	196x63x...	195x77x...	332x94x...	158x52x...	121x41x...	209x65x...	119x40x...						
10	196x66x...	194x79x...	201x68x...	194x69x...	200x90x...	187x66x...	121x41x...	217x80x...	130x45x...	324x99x...						
11	225x76x...	225x111...	229x64x...	227x55x...	225x76x...	210x71x...	225x76x...	159x54x...	293x82x...	196x66x...						
12	210x71x...	201x103...	204x70x...	202x75x...	203x119...	183x60x...	170x56x...	151x46x...	296x98x...	250x80x...						
13	208x87x...	213x112...	225x76x...	216x62x...	212x75x...	149x50x...	327x94x...	216x71x...	189x64x...	267x89x...						
14	173x75x...	196x66x...	174x52x...	170x84x...	176x80x...	141x45x...	174x50x...	238x66x...	95x30x3 ...	112x31x...						
15	270x87x...	227x92x...	263x98x...	241x100...	244x100...	261x87x...	121x41x...	191x63x...	105x37x...	121x41x...						

Convolution Neural Networks

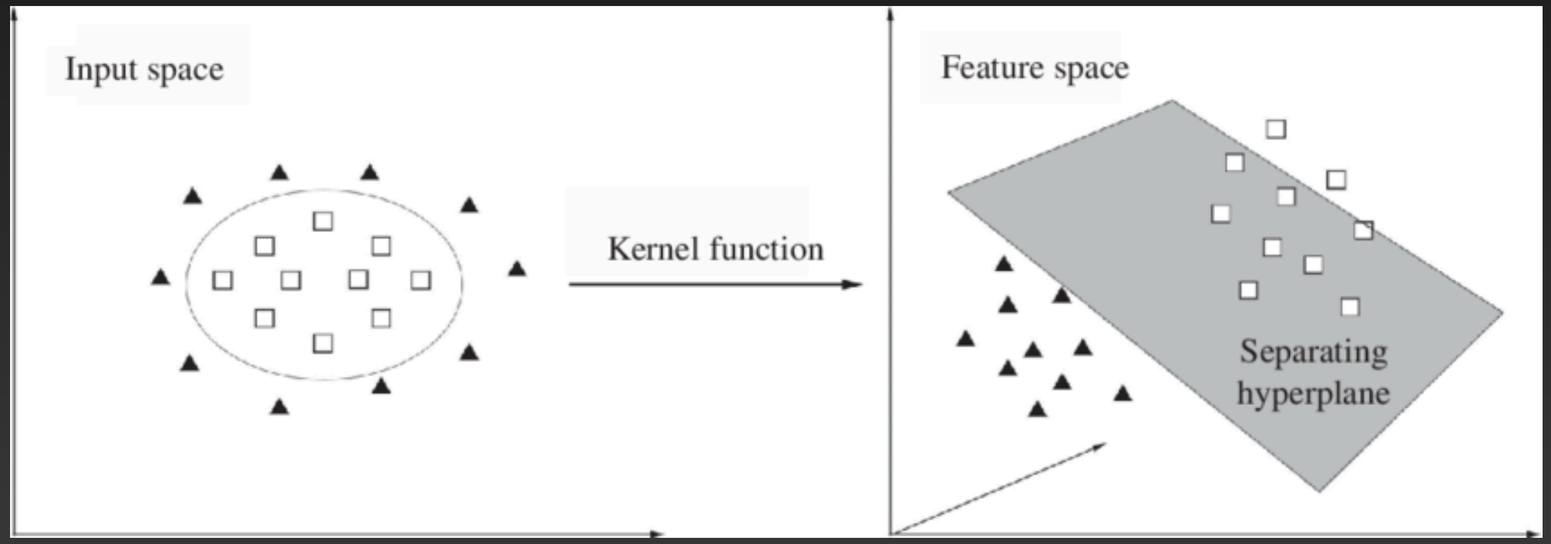
- Convolutional Neural Networks work take the approach of machine learning to learn the best convolutions that highlight the correct features to improve image identification
- Objects, in this case people have defining features that can be highlighted with the correct mask/filter/kernel.
- By using convolutional neural networks these can be learned. Convolution neural networks' efficiency are highly dependent on architecture.

General Approach

- The process of re-identification, in a multi CCTV camera surveillance network can be explained as follows,
    1. A Person walks into the area of coverage of single camera (that is a part of a network of cameras).
    2. The images of the person are processed for feature extraction and object detection
    3. The person leave the area of coverage of the first CCTV camera and enters the coverage area of another camera
    4. The Neural Network now knows enough features about the original person’s image (object) to re-identify it from the 2nd camera’s feed/pictures

Proposed Methodology

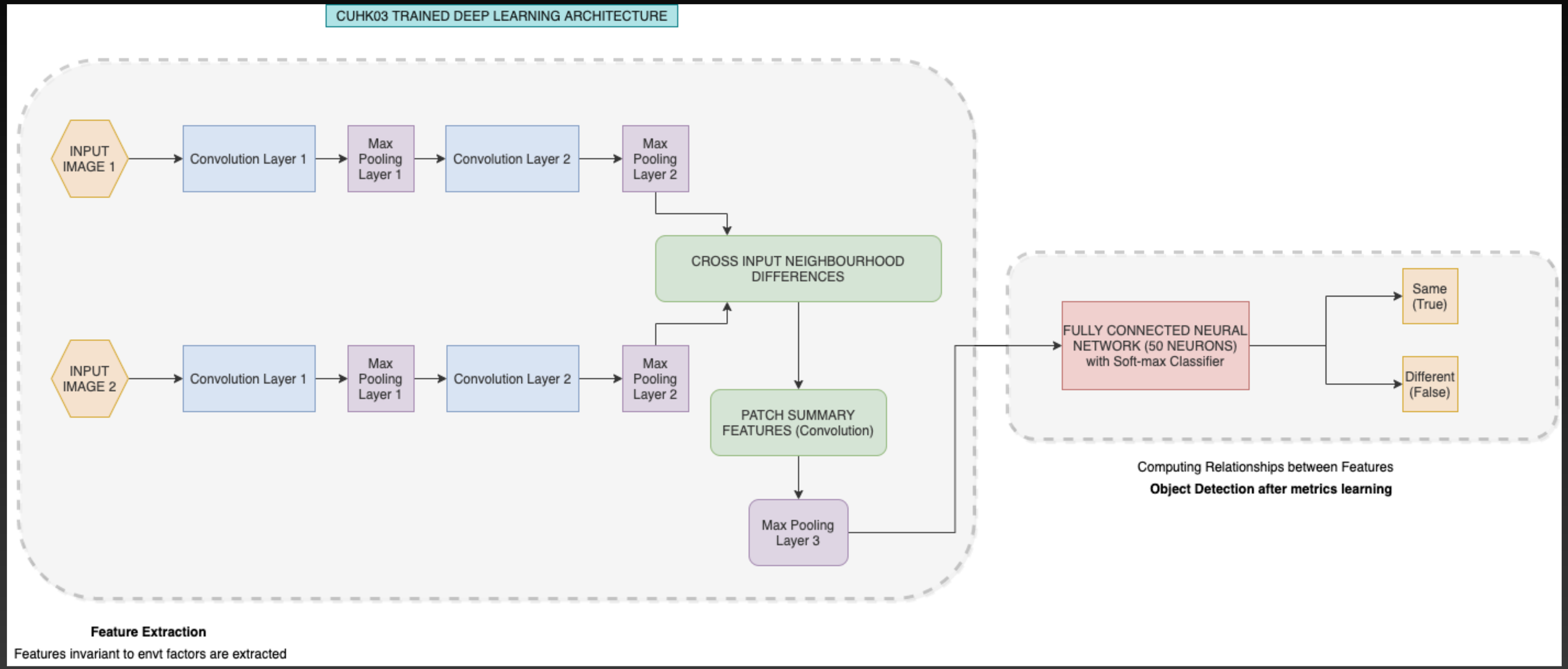
- Typically image re-identification methodology is of 2 main components -
    A. A method for extracting features from input images
    B. A metric for comparing the features
- Finding better features - invariant to light, pose and view-point changes
- Metric learning - learning approaches - mapping from feature space to a new space - feature vectors from same image are closer



Proposed Methodology

- This implementation is based on the paper - An Improved Deep Learning Architecture for Person Re-Identification by Ejaz Ahmed, Michael Jones and Tim K. Marks.
- One of the first few papers that attempted - **deep learning for Person Re-identification as binary classification**.
  A. Inputting 2 images both of which contain a person's full body
  B. Classification of the pair of images as *same* or *different* (based on whether or not its the same person in the 2 pictures).

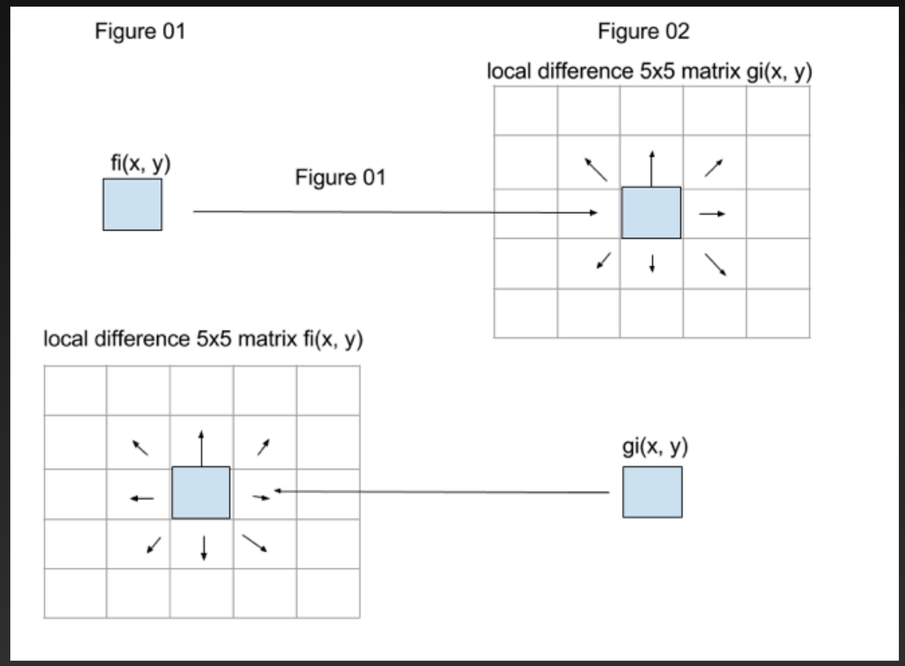
Proposed Architecture



Proposed Methodology

Architecture

1. 2 Tied Convolution Layers - For feature extraction
2. Each of this is followed by a Max Pooling layer (1 & 2) - reduces the size of the image by a factor of 2.
3. Cross input neighbourhood difference - rough relationship among features from the two input images - neighbourhood maps - positional differences - invariance



Proposed Methodology
Architecture

4. Patch summary features - convolution layers summarise these neighbourhood difference maps by producing a summary representation of the differences
5. Max Pooling Layer 3 - Final - reduces the dimensions to 18x5 pixels
6. Fully connected neural network - This is where the relationship is found with ReLu activation and Softmax loss function.

Algorithm

- The approach/algorithm used is Convolution Neural Networks with Cross Input Neighbourhood Difference.
    1. A neural network with the above architecture is created with Tensor-flow.
    2. The CUHK03 dataset is used to train the model (with the cuhk03.mat file) against 13,164 images and 1,360 identities.
    3. Images are convolved and max pooled to simplify and reduce the un-wanted features and retain the core features that help identify the person.
    4. This knowledge about ‘what makes the feature map from pictures of the same person’ is learned during the training process.
    5. 2 Input images are inputted in the program.
    6. They are convolved and max-pooled like the training images. The 2 images are passed into cross input neighbourhood difference layer - position invariance is improved.
    7. The fully connected ANN detects the image and if the features are identical then the classifier returns true. Else it return false.

Environment Set Up

- Main Environment and Dependencies
  - Mac OSX 10.15.6 - 4 Cores CPU - i5 Processor - 1600 MHz DDR3
  - Python 3.6.4
  - TensorFlow 1.8
  - opencv-python 4.4.0.42
  - numpy 1.19.2
  - h5py 2.10.0

- TesorFlow Co-dependencies
  - sabsl-py 0.10.0
  - astor 0.8.1
  - bleach 1.5.0
  - gast 0.4.0
  - grpcio 1.32.0
  - • protobuf3.13.0
  - six1.15.0 • termcolor1.1.0
  - Werkzeug1.0.1
  - zipp3.1.0

Implementation

- PreProcessing - Some image augmentation - Invariance
- Main Network Architecture
- Training
- Hardware Computing Power
- Command-line Arguments