# Data-Augmentation for Reducing Dataset Bias in Person Re-Identification

**Niall McLaughlin***, Jesus Martinez Del Rincon, Paul Miller

n.mclaughlin@qub.ac.uk

**AMMDS 2015**

# Data-Augmentation for Reducing Dataset Bias in Person Re-Identification

Agenda for this talk:

- Explain what dataset bias is and how it applies to person re-identification
- Propose possible solutions based on:

**Increasing dataset diversity using data augmentation**

**Generating synthetic training data**

- Evaluate our solutions on cross-dataset re-identification

CENTRE FOR SECURE INFORMATION TECHNOLOGIES

# What is Person Re-identification?

**Training**

- Given image-pairs from different cameras
- Learn a matching function

**Testing**

?

- Given: Image of a person seen in camera A
  Gallery of persons seen in camera B
- Match the person with their image in camera B
- Changes in pose, illumination, angle of view etc.

# What is Person Re-identification?

**Training**

**Testing**


- Given image-pairs from different cameras
- Learn a matching function

- Given: Image of a person seen in camera A
  Gallery of persons seen in camera B
- Match the person with their image in camera B
- Changes in pose, illumination, angle of view etc.

# What is dataset bias?

- **Dataset bias**

  When a classifier fails to generalise outside its training dataset

  Can happen in all areas of computer vision (classification, detection etc…)

  Main cause: the training dataset is a biased sample from the visual world

- **Effect of Dataset Bias**

  1. Develop re-identification system using dataset A
  2. Test system on dataset A -> it’s good! ☺
  3. Test system on dataset B -> it’s much worse! ☹

Dataset A　　　　　　　　Dataset B

# Name that Dataset!

1.

2.

3.

4.

5.

6.

Which dataset do these images belong to?

# Name that Dataset!

1. **Viper**

2. **CAVIAR**

3. **iLids**

4. **ETHZ**

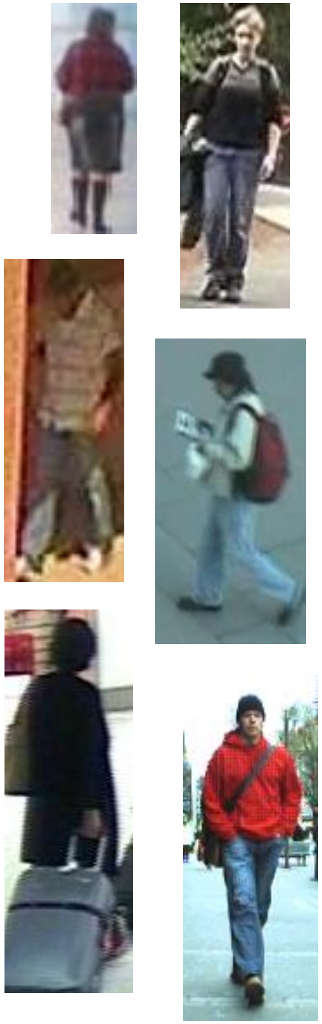
5. **CUHK**

6. **PRID 450**

Which dataset do these images belong to?

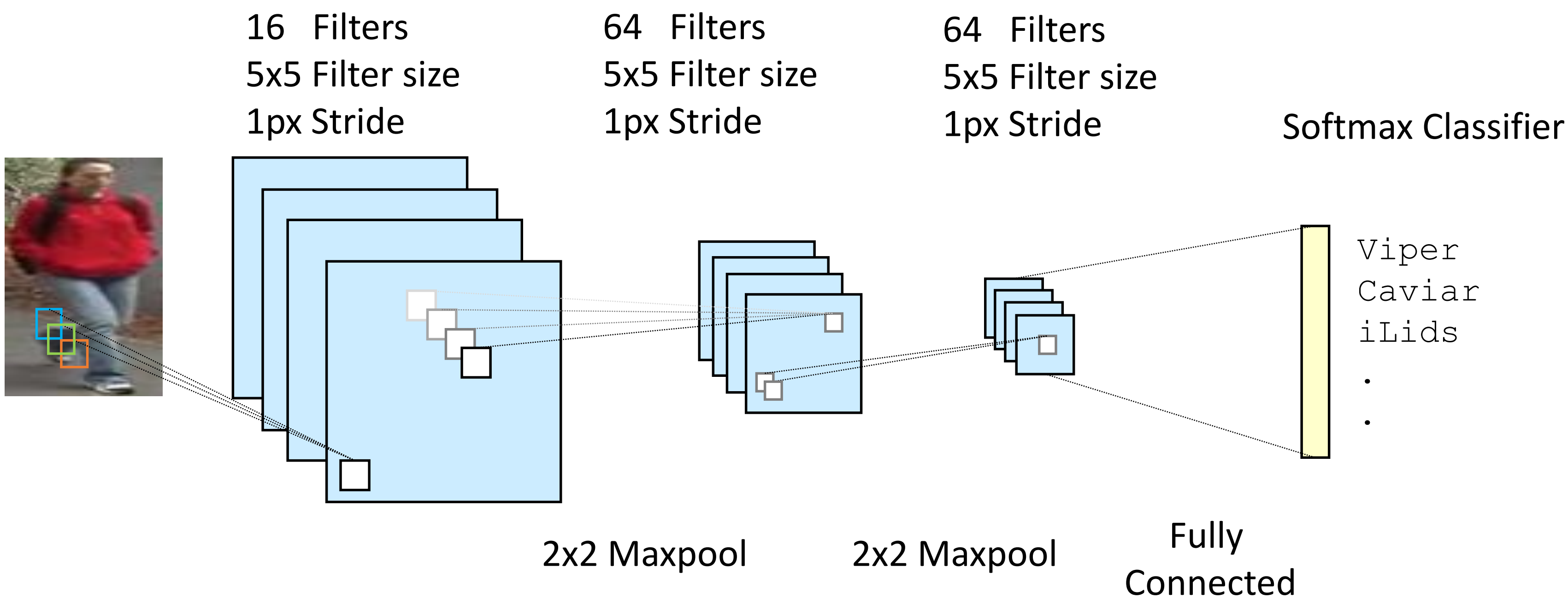
# Measuring Dataset Bias

- Can we find a way to detect dataset bias?
- Train a classifier to play *"Name that Dataset!"*
    - ***Objective:*** Given an image, predict its parent dataset
    - Create test set by combining multiple re-identification datasets
        - We use: Viper, iLids, Caviar, 3DPES
    - Label each image with its parent dataset
    - Split into 80% training, 20% testing
        - Note - No overlap between the *persons* in the training & testing sets to stop the classifier just recognising individual people!
    - Train a convolutional neural network (CNN) classifier
        - We use an existing network architecture successful on other problems



# Measuring Dataset Bias

Architecture of our CNN classifier




CNN classifier predicts the parent dataset of a given image

# Measuring Dataset Bias

- *Name that Dataset!* Classification results - confusion matrix

| Source Image Dataset \ Classifier's Prediction | Viper | iLids | Caviar | 3DPES |
|---|---|---|---|---|
| Viper | **0.96** | **0.02** | **0.02** | **0** |
| iLids | **0.03** | **0.91** | **0** | **0.06** |
| Caviar | **0.02** | **0** | **0.98** | **0** |
| 3DPES | **0** | **0.01** | **0** | **0.99** |

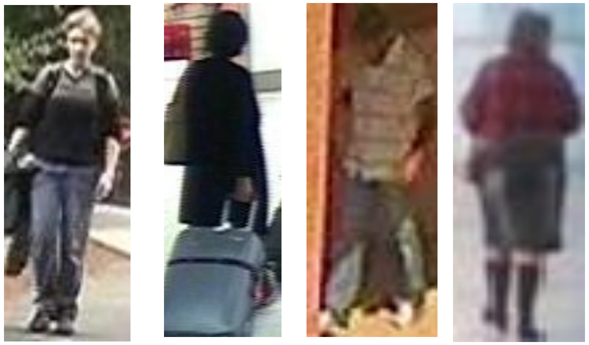
- If datasets contained only unbiased random samples from the visual world, the classifier would predict each class at approx. chance levels

# Measuring Dataset Bias

- *Name that Dataset!* game should be difficult…
  - Each dataset contains thousands of images
  - The datasets were collected to show large variability between persons viewed from different cameras
- Our results show that datasets are in fact easily distinguishable – therefore exhibit dataset bias
- Dataset Bias reduces real-world usefulness
  - What should happen:
    - Develop system using e.g. Viper
    - Deploy system to any real-world camera network(!)
  - Dataset bias means this may not work…

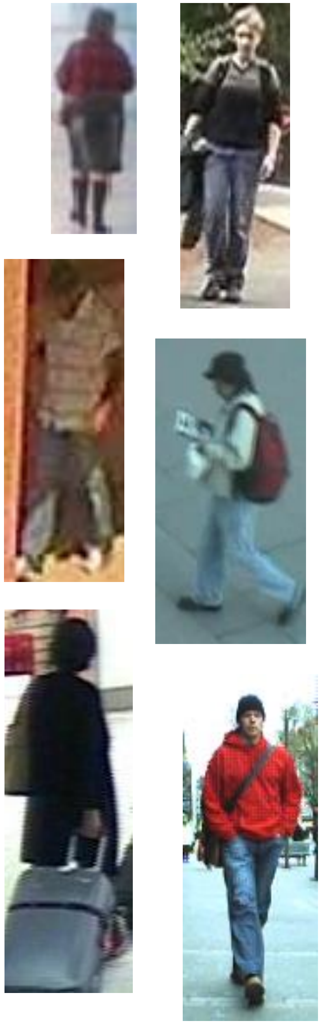
**Classifier's Prediction** (columns) / **Source Image Dataset** (rows)

| | Viper | iLids | Caviar | 3DPES |
|---|---|---|---|---|
| Viper | 0.96 | 0.02 | 0.02 | 0 |
| iLids | 0.03 | 0.91 | 0 | 0.06 |
| Caviar | 0.02 | 0 | 0.98 | 0 |
| 3DPES | 0 | 0.01 | 0 | 0.99 |



# Why does dataset bias happen?

- T. Antonio, and A Efros. "Unbiased look at dataset bias" *CVPR, 2011*
- **Selection bias**
  Dataset creators prefer certain kinds of images e.g. internet images
- **Capture Bias**
  People tend to take photos of similar things in similar ways
  Very expensive to capture the variability present in the real world
- **Label bias**
  Different people may label the same things in different ways
- **Negative set bias**
  As well as the object of interest, we must define what is not of interest
  Small negative set reduces classifier discrimination power
  A given dataset captures only a small fraction of the visual world



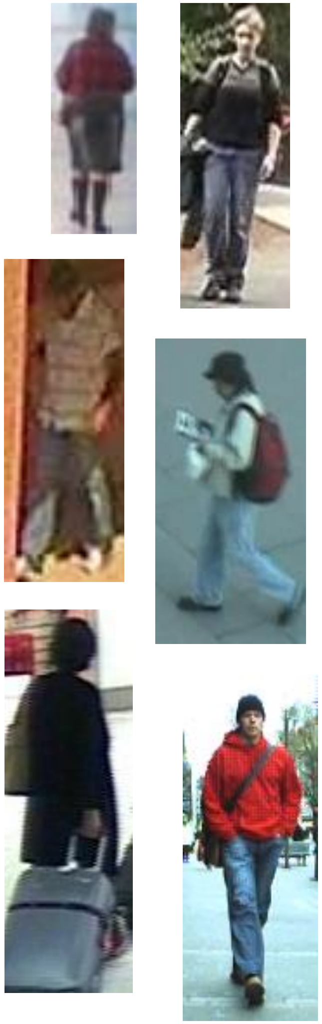
# Why does dataset bias happen?

T. Antonio, and A Efros. "Unbiased look at dataset bias." *CVPR, 2011*

- Dataset bias may be prevalent due to the way we usually develop and test algorithms
- Most computer vision tasks have several datasets available
- Train, validation, and test sets usually drawn from the same dataset
  - We know each dataset does not reflect the variability of the real world
  - Especially a problem for re-identification due to the small size of most datasets, and high correlation in appearance of most images
- Test results don't reflect true generalisation performance i.e. *how well does the system do outside the training dataset?*

# Addressing dataset bias

- **Problems**
  - Most re-identification dataset are too small
    - Lack of appearance variability
    - Images highly correlated
- **Possible Solutions**
    - We could just collect more training data, however…
    - Expensive and time consuming
    - More data from a different dataset often harms performance!
- **Our Contribution**
  - Artificially increase the variability of existing datasets
  - Generate realistic synthetic training data



# Data Augmentation

**Problem** - Re-identification dataset are small, and therefore lack variability

**Proposed Solution** - *Artificially increase training set size*

- During training, randomly modify each training image
- Network gets exposed to image variations not present in the training-set
- Combining data augmentation methods can further increase variability

# Simulated Backgrounds

- **Problems with existing datasets**
  - Contain many images that are highly correlated (even between diff. persons)
  - Small number of images per-person
  - Difficult (for a computer) to tell which image-parts are important

- **Possible Solution**
  - Generate synthetic images with artificial background
  - PRID 450 dataset includes ground-truth segmentations (Thanks!)

**Easier for re-identification system to learn which image-parts are most salient**

Original

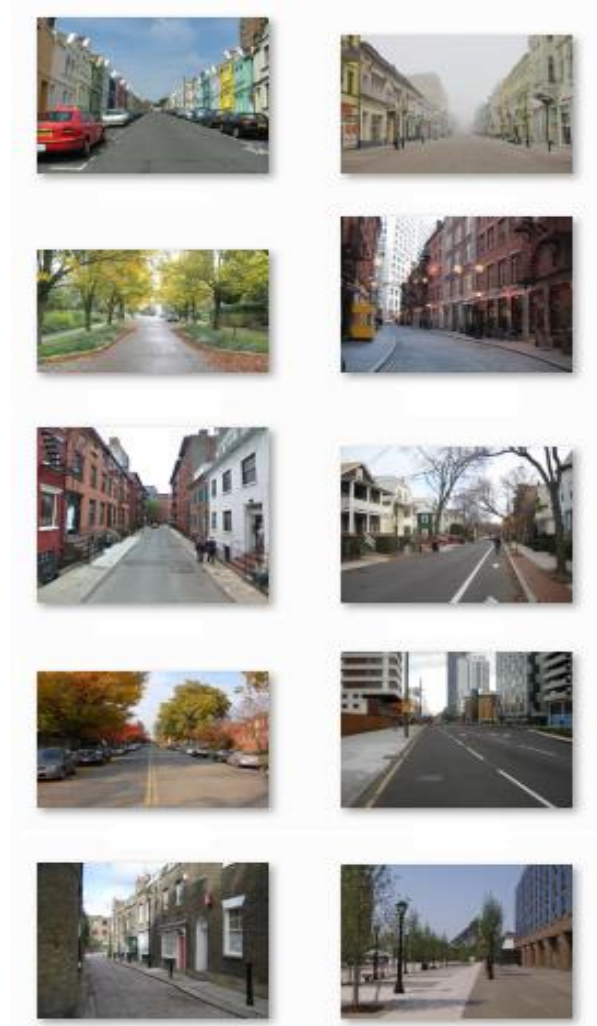
Mask

# Simulated Backgrounds

Original

Mask

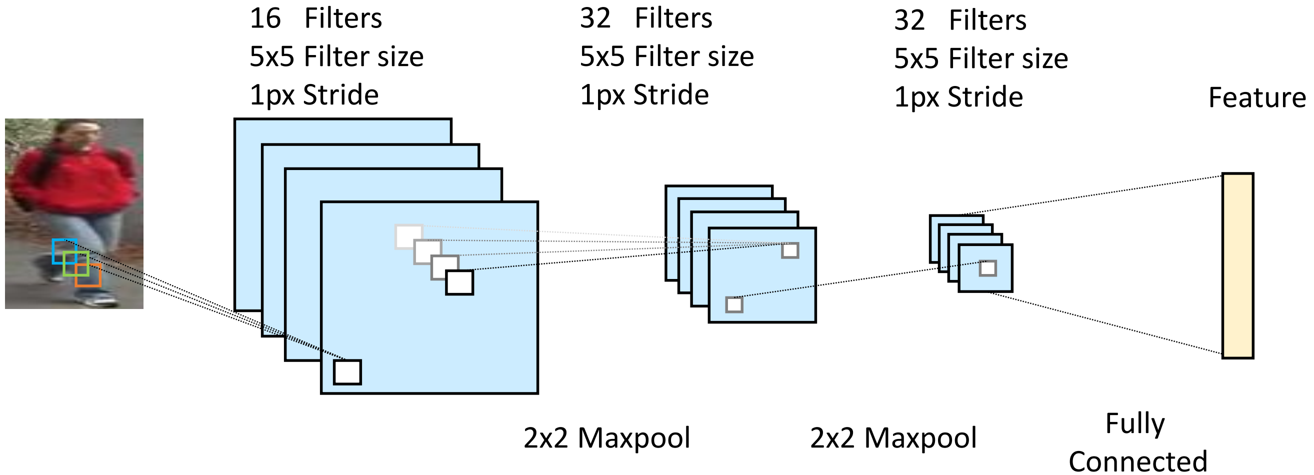
Background photos from internet



With Simulated Background

# Re-identification System

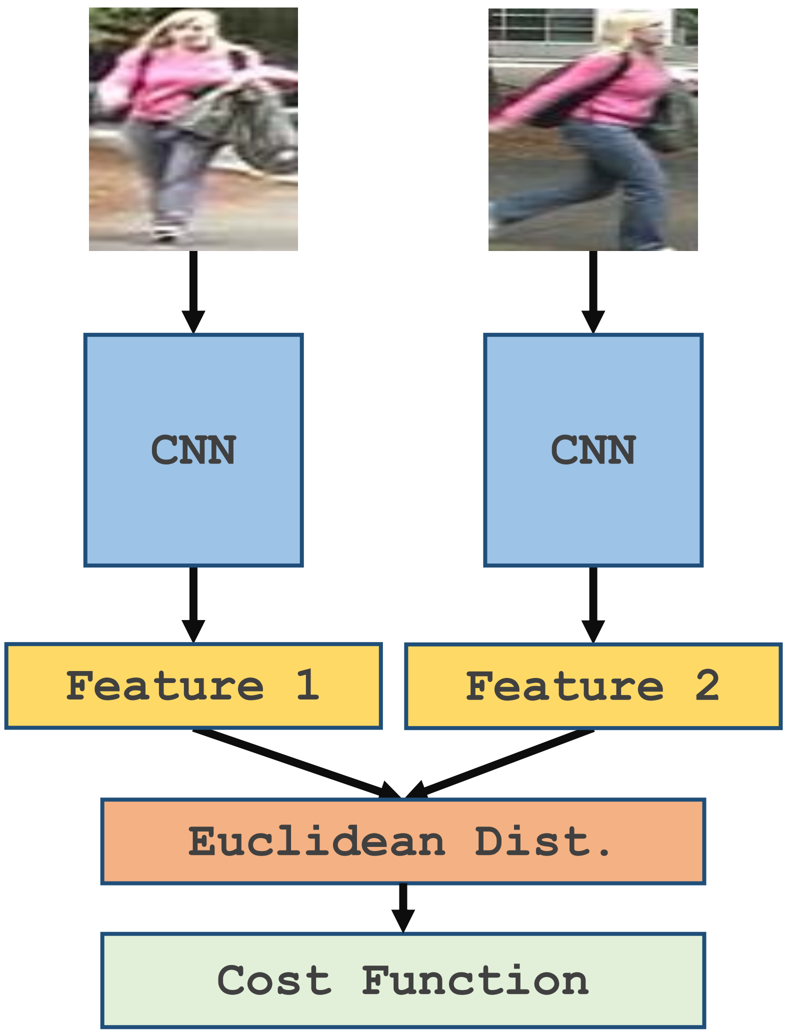
Use a convolutional neural network (CNN) to learn feature representation for re-identification images




Training Objective

- Map Images of the same person same vectors to similar vectors
- Map images of different persons to different vectors

# Re-identification System




We train a CNN to act as a feature extractor using a '*Siamese*' or embedding network architecture

In a *Siamese* network, both CNNs have identical weights

The image feature is the activations of the final layer neurons (just a vector of real numbers)

**During Training:**

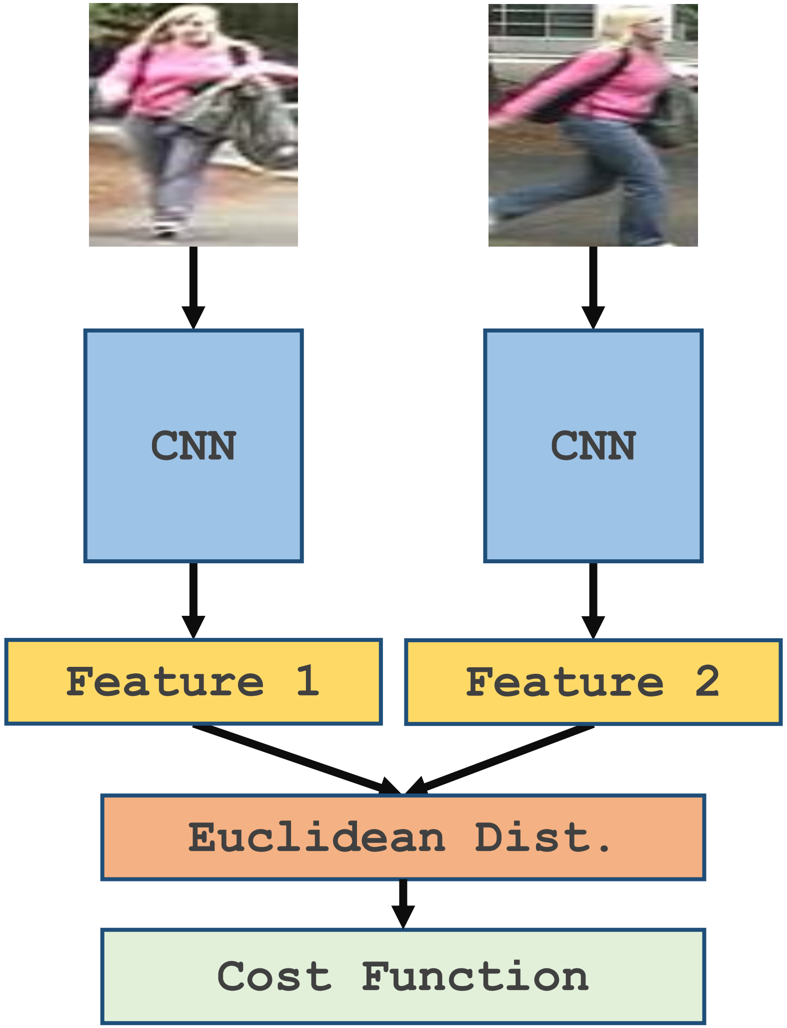
Show network pairs of similar/dissimilar images

**Training Objective:**

Images of the same person should have similar final layer neuron activations

Images of different persons should have different final layer neuron activations

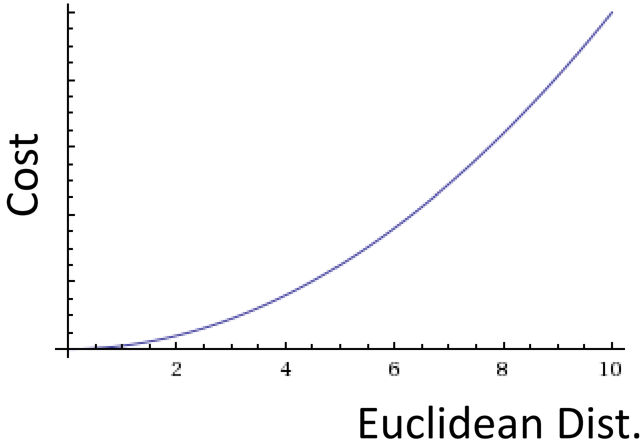
# Re-identification System



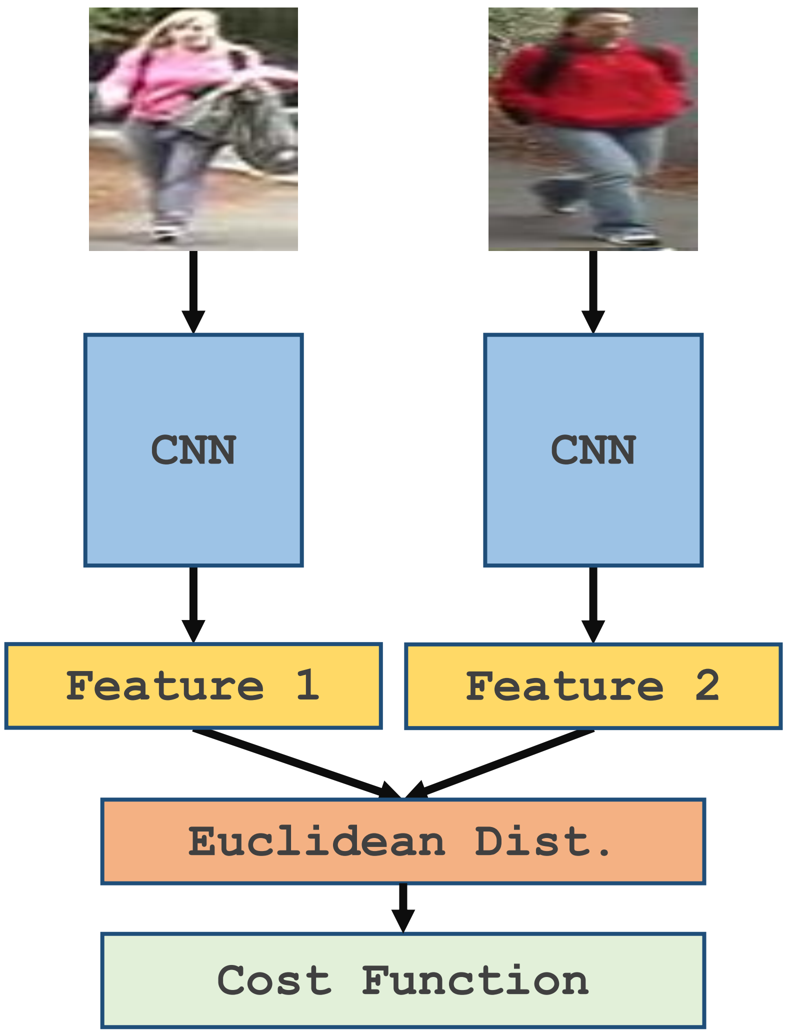

Train the CNN using a '*Siamese*' or embedding network architecture

Show network pairs of similar/dissimilar images

**Cost Function for similar images**
- Encourages feature vectors to be close



# Re-identification System

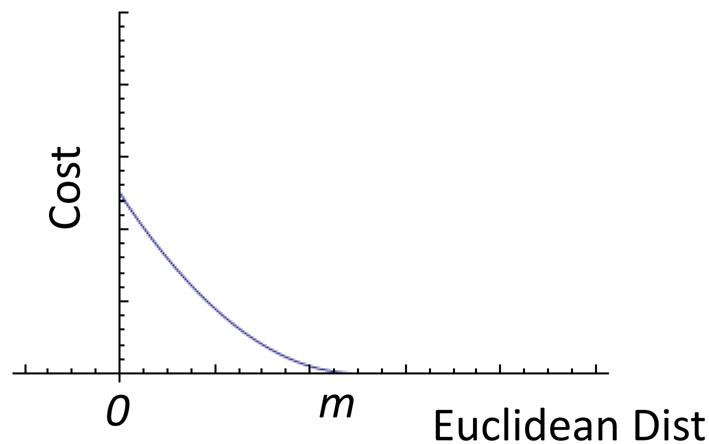



Train the CNN using a ‘*Siamese*’ or embedding network architecture

Show network pairs of similar/dissimilar images

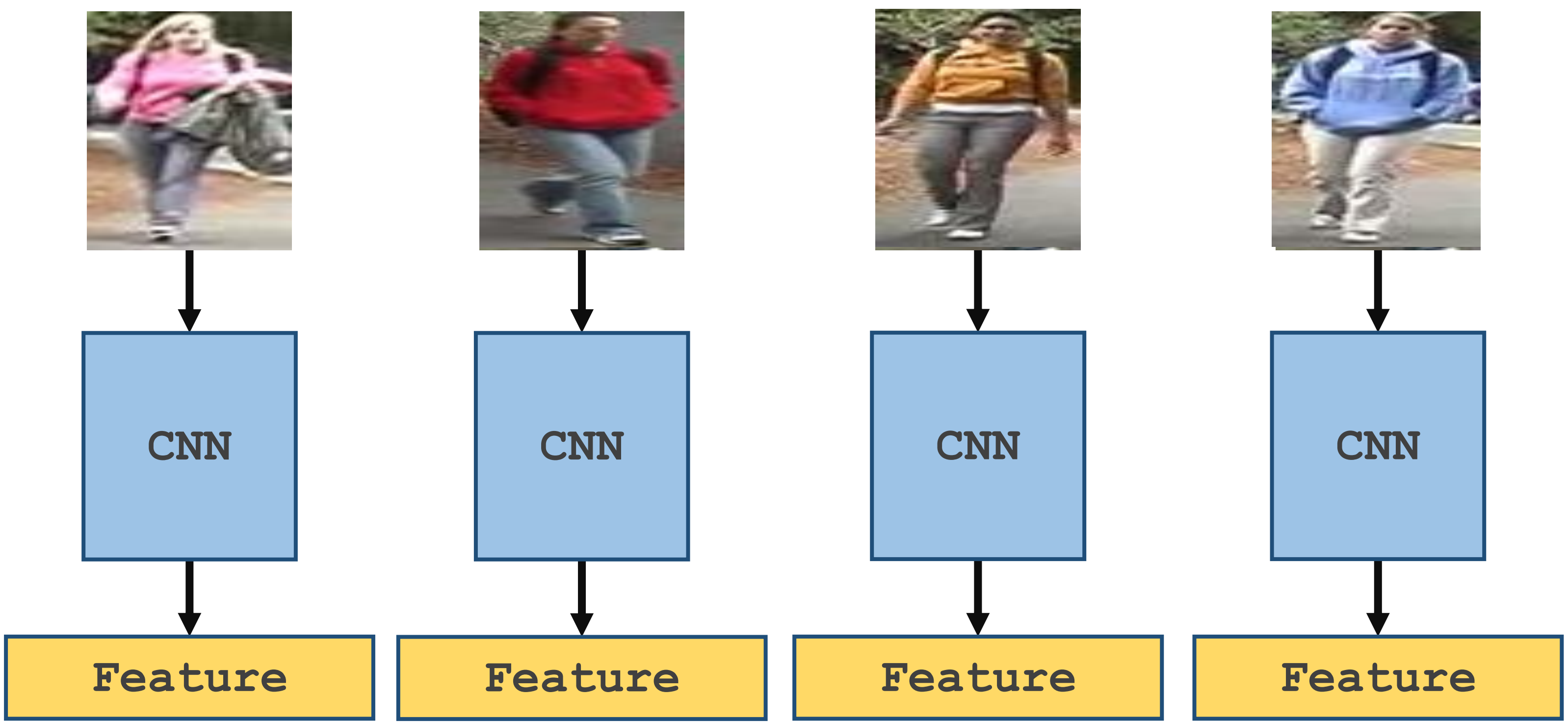
**Cost Function for dissimilar images**

- Encourages feature vectors to be separated by a margin *m*



# Re-identification System

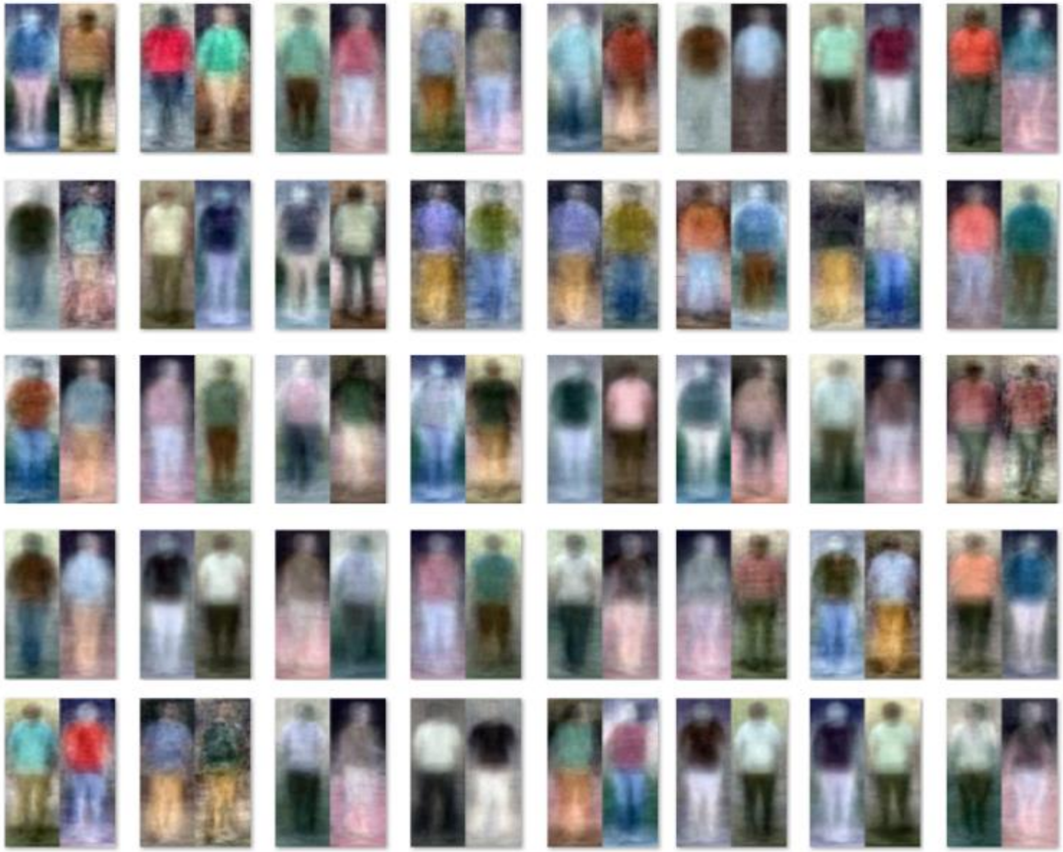
- During training the CNN has learned to act as a feature extractor
- During testing we show an image to the CNN to produce its feature representation
- Images are compared by taking the Euclidean distance between their features



# Visualise Learned Features

**What features does the neural network learn?**

For each output layer neuron visualise the image features that cause neuron to activate strongly positive (left of each pair), or strongly negative (right of each pair)



# Experimental Procedure

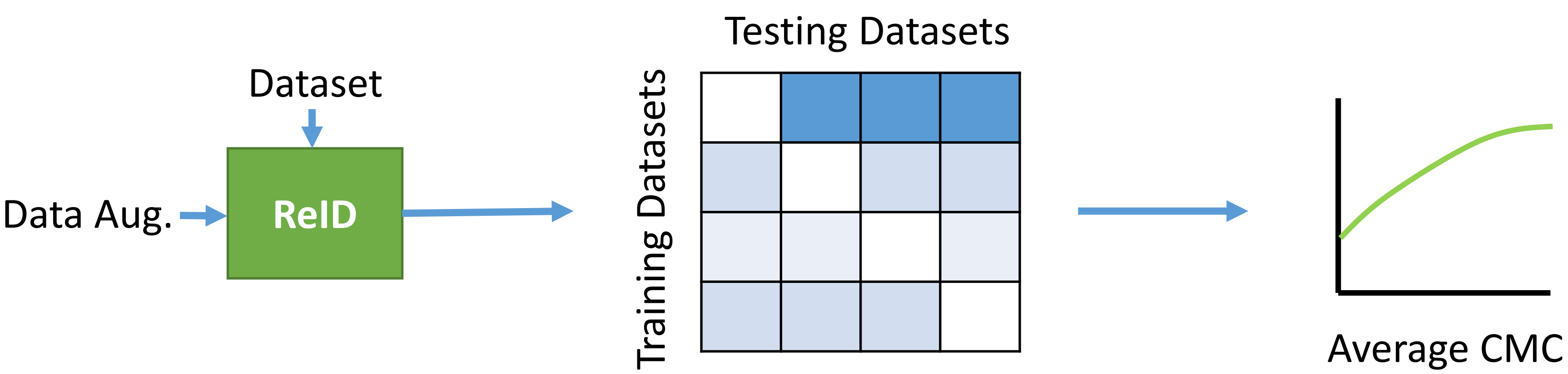
- **Question -** How does a given data augmentation method affect cross-dataset re-identification performance?

For all Data Augmentation Methods **Y**

    For all Datasets **X**

        Train re-identification system on **Dataset X** using **Augmentation Y**

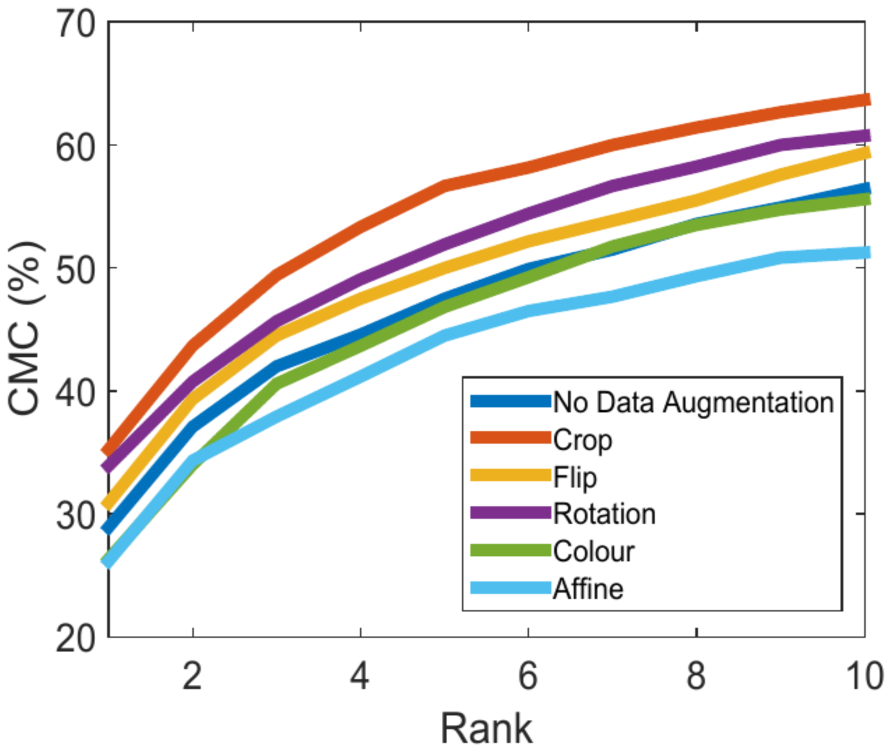
    Compute average CMC over all Datasets != **X**



# Cross Dataset Re-Identification

**Data augmentation methods used alone**

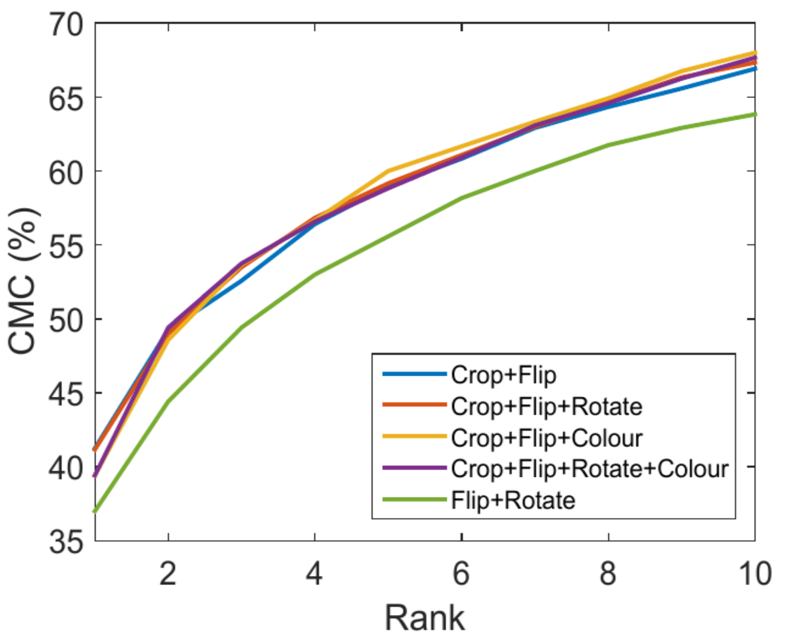
- Cropping gives biggest increase, followed by rotation and flipping
- Affine transformation and colour are worse than using no data augmentation



# Cross Dataset Re-Identification

**Top Combinations of data augmentation methods**

- Too many combination to show all results
- The top methods all perform similarly
- Combination of Cropping + Flip is essential to good results
- Adding more augmentation methods give only small change in results

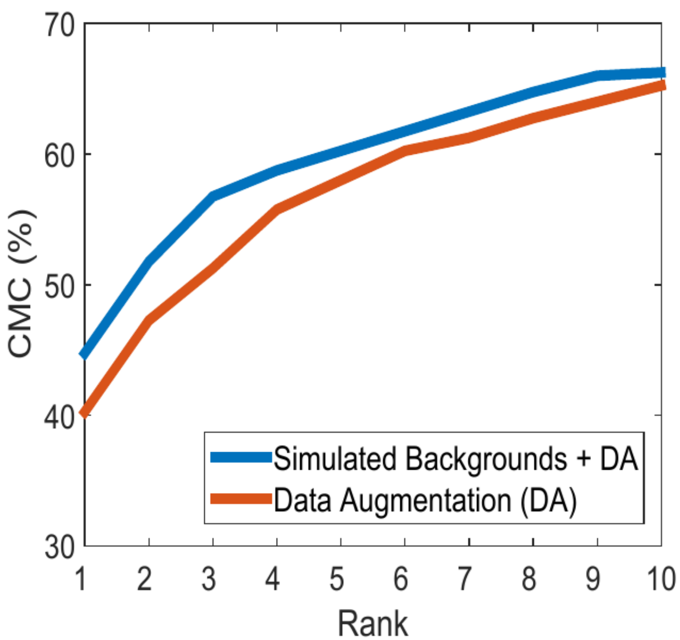


# Cross Dataset Re-Identification

**Simulated Background Data Augmentation**

- Train on PRID 450 with / without simulated backgrounds
  Note - Always using Cropping + Flipping data augmentation

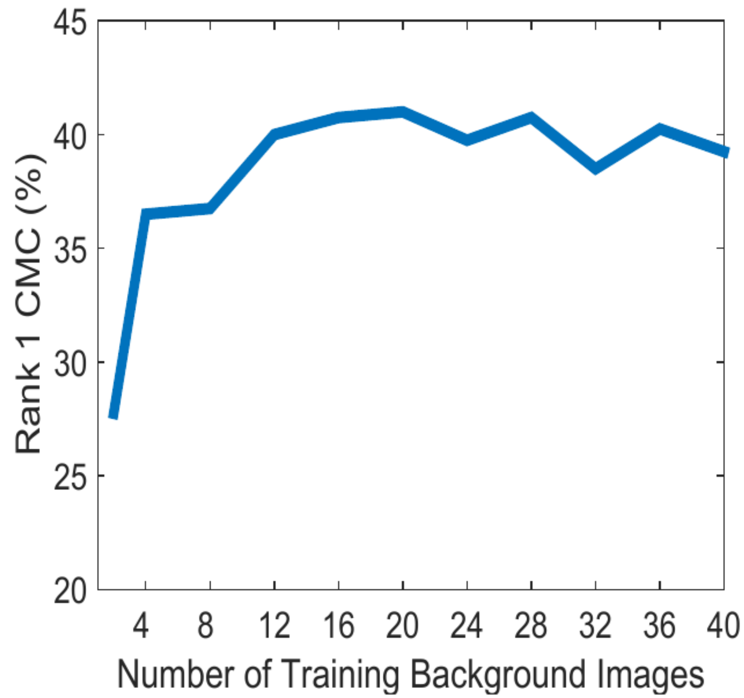
- Test on all other datasets - calculate average CMC curve



# Cross Dataset Re-Identification

**Changing Background Data Augmentation**

- How does number of training images with simulated backgrounds affect testing accuracy?




# Conclusions

- **Synthetic training data can improve performance**
  - Possible to extend this line of thinking much further!
- **Need for larger datasets with greater variability**
  - Collect new very large dataset
  - Combine existing datasets
- **Cross dataset benchmarking should be emphasised**
  - May better reflect real world performance
  - Reduces risk of over-fitting to a single dataset
- **Any Questions?**

n.mclaughlin@qub.ac.uk