# Multi-Modal Person Re-Identification using Lightweight Convolutional Neural Network


1st Reza Fuad Rachmadi
*Department of Computer Engineering*
*Faculty of Intelligent Electrical and Informatics Technology*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia 60111
fuad@its.ac.id

2nd I Ketut Eddy Purnama
*Department of Computer Engineering*
*Faculty of Intelligent Electrical and Informatics Technology*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia 60111
ketut@ee.its.ac.id

3rd Charles Chang
*Department of Computer Engineering*
*Faculty of Intelligent Electrical and Informatics Technology*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia 60111
changcenliang@gmail.com



***Abstract*—As a complement to security systems, CCTVs are increasingly used to monitor and analyze criminal acts done at a given location. However, the manual search for criminals is still prone to human error. One of the solutions to make the process more effective and efficient is with the use of re-identification. Re-identification is a computer vision and deep learning technique in which an anonymized identity of an image is matched with its owner. In this paper, we will study the method of re-identifying people with multi-modal images where the query is in the form of a body sketch drawn by several different artists. The highest Rank-1 precision achieved in this paper with the Lightweight Convolutional Neural Network is 12%.**



***Index Terms*—CCTV, Re-Identification, Multi-modal, Criminal**


## I. Introduction

OVER the years, technology has improved and revolutionized our daily lives. More and more technology created to alleviate human life, such as CCTV cameras, are increasingly used to monitor public spaces. CCTV cameras have been a long-standing security measure used in public and commercial settings. The recordings from CCTV cameras provide vital visual information and can act as a witness to a crime scene. These recordings have a prominent role to play in providing evidence in criminal investigations and disputes.

According to data taken from the Central Bureau of Statistics, there is a rising crime wave throughout Indonesia. Polda Metro Jaya alone recorded the highest number of violations, namely 31,934 incidents [1]–[4]. These facts encourage research on reducing the crime rate in various ways. Automation that can reduce the costs and workloads of the police force is needed.

Person re-identification is a computer vision and deep learning technique in which an anonymized identity of a person is matched with its owner. Person re-identification can simplify many activities that were previously done manually, by establishing a person re-identification system, the inspection of recordings conducted by the police force can be carried out faster and can reduce the costs of labor.

But a query photo of the target individual is not always readily available. Previous research of multi-modal person re-identification done by Lu Pang et al. [5] defined the problem of sketch re-identification, which uses a sketch instead of an image as the query of the model. While similar to facial sketch recognition, this problem is tackled using full-body sketches which add another dimension of complexity to the model. Furthermore, this problem is a challenging task due to the domain gap between sketch and photo. Sketches lack color information that is used as a differentiator between one individual and another. This study achieved Rank-1 precision of 34% using a state-of-the-art model and cross-domain adversarial learning.

In this paper, we investigated several lightweight Convolutional Neural Networks as a solution for sketch re-identification. We constructed the model based on the lightweight residual network used to solve the CIFAR dataset. We removed the fully connected layer of the original dataset and added two new fully connected layers.

## II. Design and Implementation

In this section, we will describe the design and implementation of lightweight Convolutional Neural Network for sketch re-identification. We will describe the experiment setting, the dataset used in the experiments, the training, and the testing processes.

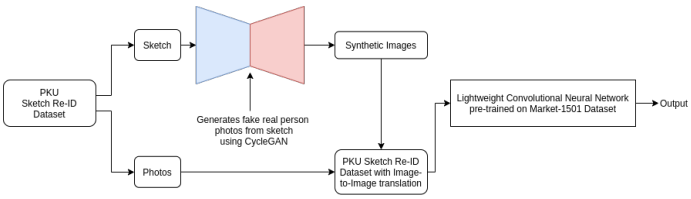


Fig. 1: Block Diagram of Working System

### A. Lightweight Convolutional Neural Network

Figure 1 shows the diagram for our proposed lightweight Convolutional Neural Network classifier based on the architecture used to solve the CIFAR dataset. However, we removed the final fully connected layer and added two new ones at the end of the classifier to ensure the model learns

good discriminatory features. Furthermore, unlike the original ResNet model, ours use an input resolution of 32x64x3 instead of 32x32x3.

In our experiments, we used two different lightweight residual networks, which are ResNet56 and ResNet110. Although the number of layers on these models is very deep, the number of parameters of the deepest classifier is still 1.7 million parameters which are less than the 23 million parameters that Resnet50 has.

| Name | Parameters |
|---|---|
| ResNet56 | 0.85M |
| ResNet110 | 1.7M |
| GoogleNet | 7M |
| DenseNet121 | 8.6M |
| ResNet50 | 23M |

TABLE I: Number of Parameters for Popular Models.

### B. PKU Sketch Re-ID Dataset

Fig. 2: Examples of PKU Sketch ReID Data

To evaluate the performance of our lightweight Convolutional Neural Network, we opted to use the PKU Sketch Re-ID dataset created by Lu Pang et al.. The dataset consists of 200 unique identities captured using two different cameras and one sketch corresponding to each individual, totaling 600 images. Figure 2 shows some examples of the dataset. The dataset was then divided into 150 identities for the training set and 50 identities for the testing set, as shown in figure 3. The images are manually cropped to ensure every photo contains one specific individual. As for the sketches, there are a total of five different artists to draw each identities sketches. Each artist has his/her own art style.


Fig. 3: Training-Testing Distribution

### C. Cross Domain Image-to-Image Translation

To help our model compensate for the lack of features in the query images, we decided to use image-to-image translation, specifically CycleGAN. CycleGAN allowed to mutually learn distributions of images, given two domains of the problems. GAN mimics the distribution of the generator given to the model which is photos and translates the style to generate synthetic photos from sketches. Essentially filling the gaps between the sketch and photos. Since the CycleGAN model is designed for unpaired translation, we created a paired testing set to ensure the style transfer is done between each unique individual.

For the CycleGAN, we used the model created by Jun-Yan Zhu et al. with the facades_label2photo pre-trained weights. The images used to train the models are the 600 images from the PKU Sketch Re-ID. The model is trained for 200 epochs with the learning rate initialized at 0.0002 and decaying to 0.0001 after 150 epochs.

### D. Training and Testing Process

Fig. 4: Examples of Random Erasing

To ensure the performance analysis is reliable, we perform the training and testing process ten times and take the average as the final evaluation metrics. All methods are evaluated using the Rank-1 accuracy of the model, following the baseline set by Lu Pang et al..

In our experiments, we handled the lack of training data by pre-training on existing person re-id datasets. We used two different datasets which are the DukeMTMC and the Market-1501 dataset. The training process is done for 100 epochs with the learning rate initialized at 0.1. The learning rate is set to decay by a factor of 0.1 every 40 epochs. Furthermore, we used random erasing and random crop to add more training data. Figure 4 shows some examples of data augmentation using those two methods.

## III. Results

Based on the results of our experiments we decided to use a droprate of 0.5, a random erasing portion of 0.5, and only choose to continue with the ResNet56 and ResNet110 Convolutional Neural Network for further experimentation since it yields the best results. However, because a complete ablation study has not been done yet, the hyperparameters could still be tuned to increase the performance.

| Name | Rank-1 | Rank-5 | Rank-10 | mAP |
|---|---|---|---|---|
| ResNet56 | 7.6% | 27% | 40.4% | 10.46% |
| ResNet56 FC 1024 | 8.2% | 29.2% | 42.2% | 10.99% |
| ResNet56 SPP | 8.4% | 26.2% | 40.2% | 10.584% |
| ResNet110 | 8.6% | 29.6% | 44.2% | 10.95 |
| ResNet110 FC 1024 | 8.8% | 29.6% | 43.8% | 11.3 |
| ResNet110 SPP | 9.4% | 25.6% | 40.2% | 11.254 |

TABLE II: Summary of lightweight residual network classifier experiments on PKU Sketch Re-ID datasets (averaging from ten different runs).

To increase the performance of the classifier, we conducted an experiment making an ensemble from the tested ResNet56 and ResNet100 classifier. From the experiments conducted, the ensemble created could increase the performance of the model by 19%, as seen on table III.

| Name | Rank-1 | Rank-5 | Rank-10 | mAP |
|---|---|---|---|---|
| Ensemble SPP | 10 % | 26.2% | 41.2% | 10.46% |
| Ensemble FC 1024 | 10.2% | 26.4% | 41.4% | 10.46% |

TABLE III: Experiments using Ensemble of ResNet56 and ResNet100 classifier (averaging from ten different runs).

## IV. Comparison

| Name | Parameters | Rank-1 | Rank-5 | Rank-10 |
|---|---|---|---|---|
| Dense-HOG+LBP+rankSVM | 8.6M | 5.1% | 16.8% | 28.3% |
| Triplet SN | n/a | 9% | 26.8% | 43.2% |
| GN Siamese | 14M | 28.9% | 54% | 62.4% |
| Cross-Domain Adversarial | n/a | 34% | 56.3% | 72.5% |
| Ensemble SPP | 3M | 10% | 26.2% | 41.2% |
| Ensemble FC 1024 | 3M | 10.2% | 26.4% | 26.4 % |

TABLE IV: Comparison to other state of the art models

Table IV shows the performance comparison of our model to several state-of-the-art models. Although the ensemble models does not have the best performance in Rank-5 and Rank-10 precision, we decided to evaluate all models using it's Rank-1 precision.

## V. Conclusion

In this paper, we introduce the usage of lightweight Convolutional Neural Network to tackle the problem of Sketch re-identification. To address the difference of modality in sketch and real images, we use image-to-image translation or more specifically CycleGAN. In the training process, we managed to have better precision compared to DenseNet, a classical model with more than quadruple the number of parameters that our model has. Other than that, or model prevailed against Triplet SN, a model composed of three identical Sketch-a-Nets and is optimized by triplet ranking loss. Although our model has not achieved state-of-the-art performance, the information gained by the classifier is very high, which proves the classifier is more efficient than other methods.

## References


[1] "Kasus kriminal meningkat 7,04 persen dalam sepekan, salah satunya perampokan," 2020, https://nasional.kompas.com/read/2020/05/18/16253371/kasus-kriminal-meningkat-704-persen-dalam-sepekan-salah-satunya-perampokan.

[2] "Dua pekan terakhir, polri catat peningkatan kejahatan 11,80 persen," 2020, https://nasional.kompas.com/read/2020/04/20/20542321/dua-pekan-terakhir-polri-catat-peningkatan-kejahatan-1180-persen.

[3] "Ini alasan angka kriminalitas meningkat pekan lalu menurut polri," 2020, https://nasional.kompas.com/read/2020/05/18/16253371/kasus-kriminal-meningkat-704-persen-dalam-sepekan-salah-satunya-perampokan.

[4] "Dalam sepekan, polri catat peningkatan kejahatan jalanan di indonesia," 2020, https://nasional.kompas.com/read/2020/05/12/17363331/dalam-sepekan-polri-catat-peningkatan-kejahatan-jalanan-di-indonesia.

[5] Lu Pang, Yaowei Wang, Yi-Zhe Song, Tiejun Huang, Yonghong Tian, "Cross-domain adversarial feature learning for sketch re-identification," 2018, https://www.pkuml.org/resources/pkusketchreid-dataset.html, Last accessed on 2020-11-30.