

# Enhancing the Quality of Cellular Camera Video With Convolutional Neural Network

Hotman Parsaoran Tampubolon  
Universitas Prima Indonesia  
Medan, Indonesia  
[hotmant97@gmail.com](mailto:hotmant97@gmail.com)

Steven Eduard Gulo  
Universitas Prima Indonesia  
Medan, Indonesia  
[eduardstevengulo@gmail.com](mailto:eduardstevengulo@gmail.com)

Watas Sinurat  
Universitas Prima Indonesia  
Medan, Indonesia  
[watassinurat1005@gmail.com](mailto:watassinurat1005@gmail.com)

Befi Juniman Gulo  
Universitas Prima Indonesia  
Medan, Indonesia  
[befijunimangulo96@gmail.com](mailto:befijunimangulo96@gmail.com)

Amir Mahmud Husein  
Universitas Prima Indonesia  
Medan, Indonesia  
[amirmahmud@unprimdn.ac.id](mailto:amirmahmud@unprimdn.ac.id)

**Abstrak**— At present technological developments, especially in the field of computer vision, are showing significant performance such as the application of convolutional neural networks that have a very high degree of accuracy, for example improving video quality which recently has image restoration such as super resolution (VSR) thanks to deep learning with the aim of helping produce better visual videos. The use of video cameras for mobile devices is now increasingly highly developed. Nowadays mobile devices are experiencing a rapid increase in quality especially in cameras. However, physical limitations such as the small sensor size, compact lens and the lack of supporting hardware can prevent cellular devices from achieving good video camera quality results. For that many method approaches are applied, one of which is the CNN (Convolutional Neural Network) method. This method can improve the image of video recordings that have poor quality.

**Keywords**—*Convolutional neural network, computer vision, Improved video quality ;*

## I. INTRODUCTION

The development of research in the field of computer vision in recent years has shown sophisticated performance with a high degree of accuracy due to the application in various fields such as face recognition (Husein & Harahap, 2017) - (Wijaya, Husein, & Harahap, 2017), signature identification (Harahap, Husein, & Dharma, 2017) - (Husein & Harahap, 2017) and camera surveillance (Husein, Calvin, Raymond, & William, 2018), recently the application of convolutional neural networks, for example object detection (Husein & Harahap, 2017) - (Wijaya, Husein, & Harahap, 2017), (Harahap, Husein, & Dharma, 2017) - (Husein & Harahap, 2017), (Husein, Calvin, Raymond, & William, 2018) - (Jeong, Park, & kwak, 2017) - (Zhang, Wen, Bian,

Lei, & Z, 2018), and (Ren, He, Girshick, & Sun, 2015) - (Ahn, Kang, & Sohn, 2019), pedestrians (Haris, Shakhnarovich, & Ukita, 2019), (Caballero, et al., 2017). Recent image restorations such as super resolution (Kim, Lim, Na, & Kim, 2018) - (Shi, et al., 2016), (Ignatov, Kobyshev, Timofte, Vanhoey, & Gool, 2018) and super resolution video (Vorobjov, Zakharava, Bohush, & Ablameyko, 2018), experienced a significant increase thanks to deep learning with the aim of helping produce visual video applications. Super-Resolution (SR) is an imaging technique that converts low resolution images (LR) into higher resolution images where LR as input into the SR algorithm network to exploit internal information so as to produce

output images similar to high resolution (HR) (Kim, Lim, Na, & Kim, 2018)

Video SR, or multi-frame SR, assumes that the input is a series of successive frames at each time instance of a video sequence, the application of the SR image algorithm to individual video frames has been proven

high performance and efficiency (Shi, et al., 2016), in addition to improving the quality of cellular photos equivalent to the quality of DSLR cameras has been presented by (Ignatov, Kobyshev, Timofte, Vanhoey, & Gool, 2018) by proposing a CNN deep learning model. In the study (Vorobjov, Zakharava, Bohush, & Ablameyko, 2018) proposed the YOLO method for detecting small objects on video, improving video quality for motion detection is presented (Caballero, et al., 2017).

From the results of several studies indicate that the improvement of super resolution video quality has achieved high performance with the CNN deep learning approach, in this study we propose the CNN model approach to improve the quality of videos recorded using 3 (three) types of cellular on the highway. The main contribution in this research focuses on comparing video quality, then we apply the YOLOv3 framework for object detection based on video that has been upgraded with original video.

## II. LITERATURE REVIEW

### 2.1. Convolutional Neural Network (CNN)

Convolutional Neural Network is one of the machine learning methods of developing Multi Layer Perceptron (MLP) which is designed to process two-dimensional data. CNN is included in the Deep Neural Network type because of its deep network level and is widely implemented in image data. CNN has two methods; namely classification using feedforward and learning stages using backpropagation. Network architecture using CNN is shown in Figure 1. CNN structure consists of input, feature extraction process, classification process and output. Extraction process in CNN consists of several hidden layers, namely the convolution layer, activation function (ReLU), and pooling. CNN works hierarchically, so the output at the first convolution layer is used as input at the next convolution layer. In the classification process consists of fully-connected and activation functions (softmax) whose output is in the form of classification results

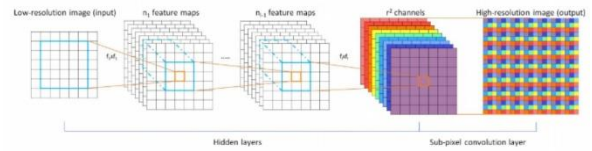


Figure 2.1 CNN Architecture

### 2.2 Convolution Layer

Filters are used to extract objects from the input image by the convolution layer. Filters contain weights whose function is to detect characters from objects such as color, the edge. Convolution will produce a linear transformation of the input image that matches the spatial information in the data. There are parameters that can be changed to modify the properties of each layer, namely filter size, stride and padding. Stride is a parameter that determines how many filters shift, if the value of stride is 1, then conv.filter will shift as many pixels horizontally then vertically. The smaller the stride, the more detailed information we get from an input, but requires more computing when compared to a large stride. Padding is While Padding or Zero Padding is a parameter that determines the number of pixels (containing the value 0) to be added on each side of the input. This is used for the purpose of manipulating the output dimensions of conv. layer (Feature Map). The results from the receptive field are single data. The output of this convolution process is used as input for the next convolution layer. An illustration of the stride and padding process is shown in Figure 2 and Figure 3.

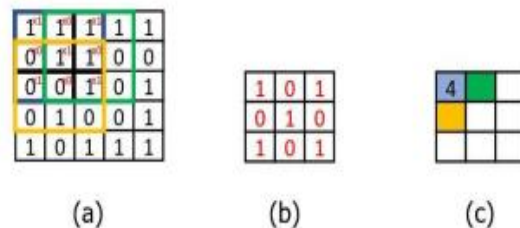


Figure 2.2. Convolution operation with stride 1 (a) Input data 5x5 (b) filter 3x3 (c) field receptive 3x3

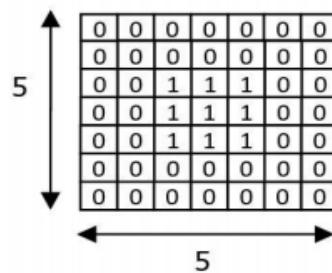


Figure 2.2. Zero padding 2 operation on 3x3 data.

### 2.3. ReLu Activation Function

The function used for activation in reLU, the reLU function is the output value of the neuron can be expressed as 0 if the input is negative. If the input value of the activation function is positive, the output of the neuron is the input value of the activation itself. ReLU activation function

is  $f(x) = \max(0, x)$  (Heaton, 2015). Because ReLU basically only creates a boundary at zero, meaning that if  $x \leq 0$  then  $x = 0$  and if  $x > 0$  then  $x = x$ .

### 2.4. Pooling

Pooling layer is usually after conv. layer. In principle, the pooling layer consists of a filter of a certain size and stride that shifts throughout the feature map area. Pooling commonly used is Max Pooling and Average Pooling. The purpose of using the pooling layer is to reduce the dimensions of the feature map (downsampling), thus speeding up the computation because the parameters that need to be updated are fewer and overcome overfitting.

### 2.5. Fully Connected Layer

The fully connected layer is a collection of convolution processes To determine which features are most correlated with a particular class, input from processes from the previous layer is used. The feature map that is produced from the feature extraction is still in the form of a multidimensional array, so it must do a "flatten" or reshape the feature map into a vector so that it can be used as input from the fully-connected layer.

The Fully-connected layer is the layer where all the neurons of activity from the previous layer are all connected with the neurons in the next layer as if a neural network can. Every activity from the previous layer needs to be converted into one-dimensional data before it can be connected to all neurons in the Fully-Connected layer. The Fully-Connected layer is usually used in the Multi-layer Perceptron method and aims to process the data so that it can be classified

## I. METHOD

In this study the data used are video recordings of 3 types of Oppo, Xiaomi, Samsung mobile devices on the highway. With the following specifications:

**Tabel 3.1.** Device Specifications

| HP Model | Resolution | Screen          | The Features             |
|----------|------------|-----------------|--------------------------|
| OPPO     | 13 MP      | (1520 x 720 px) | LED Flash, HDR, panorama |

|         |       |                 |                          |
|---------|-------|-----------------|--------------------------|
| XIOMI   | 13 MP | (1280 x 720 px) | LED Flash, HDR, panorama |
| SAMSUNG | 13 MP | (1280 x 720 px) | LED Flash, HDR, panorama |

Recorded data can be seen in Figure 4.5. The appearance of the image is the result of capture of the recorded video.

Recording process

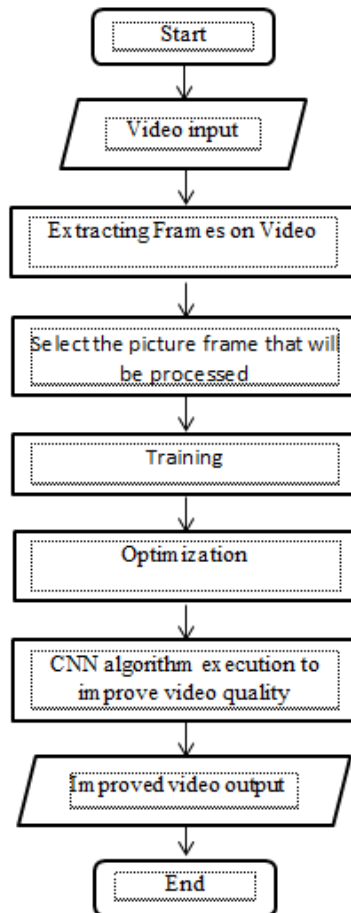
use a tripod to make the video recording results can be stable.

The making of the CNN model was carried out using the Tensorflow framework



**Figure 3.1.** Data Samples

### 3.1. Work Steps



**Figure 3.2.** Flowchart Video Quality Improvement Process

The flow of this research starts from inputting sample data in the form of video recordings on the highway. Recordings have a different quality level for each device used as input training data. Then design a network with CNN method to improve the quality of input video.

### 3.2. Sampling for input data

Data input used in this network is a sample of recorded video data from three mobile devices with their respective resolutions (1520 x 720 px) on the highway using a tripod to make the recording results stable. The total sample data used is 50 videos for each device, so the total data total is 150 data. The data is then used as training data. Training data is used for do the network learning process. The output from the

network is the video from the recording that has been upgraded through the CNN network.

## II. RESULTS AND DISCUSSION

### 4.1. Implementation CNN

There are three stages in implementing CNN, namely training, validation and testing. The training phase is the main stage for training the network to learn data input. Then the network is tested on validation data. If it gives good results, then the network can be used to classify data with test data.

### 4.2. Video Quality Improvement Process

Before carrying out the process of improving the quality of the objects first detected in the image with the aim that the resulting image has better quality. The test will be carried out using various types of videos from 3 mobile devices which are then extruded into images for each device. The following detailed results of the tests carried out. Following the extracted video data can be detailed in Figure 3.6. below this :



**Figure 3.3.** Three examples of original images (above) with upgraded images (below) taken from Oppo, Xiaomi and Samsung

In the testing process carried out using 3 types of mobile devices namely Oppo, Xiaomi, and Samsung each have good image quality after being improved by the Convolution Neural Network method. There are several factors that affect the recording quality of the three devices, such as lighting levels, color sharpness. The results of using the Convolution Neural Network method can improve the sharpness and color structure even though the results of video recordings of when this type of mobile device has a less good lighting level.

### III. CONCLUSION AND SUGGESTIONS

#### 5.1. Conclusion

After completing research on improving video quality on mobile devices, the author can draw the following conclusions :

1. CNN Method can improve the quality of an object contained in a digital image of a mobile device well.
2. The CNN method is able to distinguish objects that are focused in the database well in term of color, brightness, and contrast

#### 5.2. Suggestion

The few suggestions that I want to convey are as follows:

1. To increase the accuracy of recognizing an object in a dark place, it can add infrared / infrared light .
2. Optimizing execution time by improving the algorithm.

### IV. ACKNOWLEDGE

Journal of Research Improving the Quality of Cellular Camera Video With CNN, expresses sincere appreciation to all reviewers for selflessly contributing their expertise and time to the review process, which is very important to guarantee the quality and substantive impact of the journal. Journal editors and joint authors thank the reviewers for evaluating and evaluating the articles submitted for consideration in the publication process, regardless of the outcome (acceptance or rejection).

### V. REFERENCES

- Ahn, N., Kang, B., & Sohn, K.-A. (2019). *Efficient Deep Neural Network for Photo-realistic Image Super-Resolution*.
- Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., & Shi, W. (2017). Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation. *CPVR* (pp. 4778-4787). CVPR.
- Harahap, M., Husein, A. M., & Dharma, A. (2017). IDENTIFIKASI TANDA TANGAN DENGAN KOHONEN SOM BERBASIS PRINCIPAL COMPONENT ANALYSIS. *Semnastikom* (pp. 333-337). Jayapura: Aptikom.
- Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent Back-Projection Network for Video Super-Resolution. *CPVR* (pp. 3897-3906). CPVR.
- Husein, A. M., & Harahap, M. (2017). Penerapan Metode Distance Transform Pada Kernel Discriminant Analysis Untuk Pengenalan Pola Tulisan Tangan Angka Berbasis Principal Component Analysis. *Sinkron*, 31-36.
- Husein, A. M., & Harahap, M. (2017). PENGENALAN MULTI WAJAH BERDASARKAN KLASIFIKASI KOHONEN SOM DIOPTIMALKAN DENGAN ALGORITMA DISCRIMINANT ANALYSIS PCA. *QUERY: JURNAL SISTEM INFORMASI*, 33-39.
- Husein, A. M., Calvin, C. H., Raymond, L., & William, W. (2018). Motion detect application with frame difference method on a surveillance Camera. *International Conference on Mechanical, Electronics, Computer, and Industrial Technology (2018)* (pp. 1-10). Medan: Universitas Prima Indonesia.
- Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., & Gool, L. V. (2018). WESPE: Weakly Supervised Photo Enhancer for Digital Cameras. *CVPR*, (pp. 804-813).
- Jeong, J., Park, H., & kwak, N. (2017). Enhancement of SSD by concatenating feature maps for object detection. *British Machine Vision Conference (BMVC)* (pp. 76.1-76.12). London: BMVA Press.
- Kim, S. Y., Lim, J., Na, T., & Kim, M. (2018). *3DSRnet: Video Super-resolution using 3D Convolutional Neural Networks*. ArXiv.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 1-14). IEEE.
- Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., . . . Wang, Z. (2016). Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *ArXiv*, (pp. 1-10).
- Vorobjov, D., Zakharava, I., Bohush, R., & Ablameyko, S. (2018). An effective object detection algorithm for high resolution video by using convolutional neural network. *15th International Symposium on Neural Networks, ISNN 2018*, (pp. 503-510). Minsk, Belarus.
- Wijaya, B. A., Husein, A. M., & Harahap, M. K. (2017). Implementation Distance Transform Method in Kernel Discriminant Analysis for Face Recognition Using Kohonen SOM. *International Journal of Engineering Research & Technology (IJERT)*, 28-31.
- Zhang, S., Wen, L., Bian, X., Lei, Z., & Z, S. (2018). Single-Shot Refinement Neural Network for Object Detection. *CVPR* (pp. 4203-4212). Salt Lake: CVPR.