Automatic Cigarette Object Concealment in Video using R-CNN

Kadek Utari Widiarsini[1], Duman Care Khrisne[2], and I Made Arsa Suyadnya[3]

[1,2,3]Department of electrical Engineering, Faculty of Engineering,
Udayana University,
Bali, Indonesia
duman@unud.ac.id

Abstract - Cigarettes are packaged processed tobacco products, produced from the Nicotiana Tabacum, Nicotiana Rustica plants and other species or synthetics that contain nicotine with or without additives. Smoking is known to the public as one of the causes of death in the world that is quite large such as asthma, lung infections, oral cancer, throat cancer, lung cancer, heart attacks, strokes, dementia, erectile dysfunction (impotence), and so on. This research aims to build an application that can recognize cigarettes automatically and conceal pictures so that people especially minors are not affected by cigarettes. The application is built using the Region-based Convolutional Neural Network (R-CNN) method. The study uses images that have cigarette objects in them. The test is carried out to find out the application performance such as the level of application accuracy in recognizing cigarette objects. Based on the test results with a sample of 126 cigarette images, the application built is able to recognize cigarette objects by obtaining an accuracy value of 63%.

Index Terms—*Cigarettes, R-CNN, Mask R-CNN, Deep Learning.*

I. INTRODUCTION

Media has become a phenomenon in the communication process. Even human dependence on mass media is so great. The communication media most liked by the public is television.

The fact that adults spend about 20 hours per week shows that watching television is the number one activity. In Indonesia, according to the AGB Nielsen survey in 2006, 42.86% to 95.83% of Indonesians prefer to watch television. These results further show that nearly 8 out of 10 adults watch television every day. In the provisions of Law 33/2009, all films, whether feature films, documentaries, commercial films, or film advertisements (trailers, posters, and photos) must be censored before being shown to the public. This is very important because, in addition to positive benefits, the film also has a negative impact, so the role of censorship is needed. Film censorship can be considered the final filter for pre-show film work. Films shown to the public, especially those of a commercial nature, must pass censorship. The censorship process must be carried out during the process of editing, appraising themes, images, sounds which takes a lot of time and effort.

II. RELATED WORKS

This study refers to references related to existing research. Related research regarding object detection in images using several methods/approaches is as follows.

Several previous studies [1]–[5] have provided information that using CNN is a very effective method for carrying out the digital image recognition process. However, to do the recognition of specific objects in images we need more capabilities than CNN. therefore we try to use one of the object recognition techniques, namely Region-based Convolutional Neural Network.

The research conducted by [6] uses the Fast R-CNN method to classify and detect certain fashion items used by people in an image. Helder conducted 3677 train images per category and conducted 696 testing images per category. The results show that using the CNN method to detect fashion items used by someone produces an average precision of close of 78%, for pants by 65%, and an average of accessories such as glasses by 57%. The Fast R-CNN method is used to further shorten the time in object training.

III. PROPOSED APPROACH

In this study, we analyze the Detection is a process of checking or examining something using certain methods and techniques. Detection can be used for various problems, for

example in an image detection system, where the system identifies which is related to the image. The purpose of detection is to solve a problem in various ways depending on the method applied to produce a solution.

An object is a concept, an abstraction, or something that has clear boundaries and is intended for an application. Objects can be found in pictures or videos. According to [7] argues that "Images are everything that is manifested visually in the form of two dimensions as an outpouring of feelings or thoughts", whereas in the Big Indonesian Dictionary" Pictures are imitations of goods, animals, plants and so on.

An object detection system is a method that is used together by using certain methods in an image or video to achieve the desired results.

A. Moving object detection

Object detection is the process of detecting moving objects in a video or image sequence. Before doing the object detection process, The video must be converted into an image sequence. Figure 1 shows the video stages when converted into image sequences.

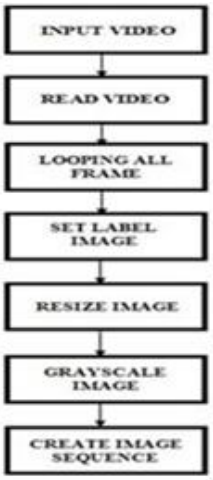

Fig. 1 Video Stages Becoming Image Sequence

Figure 1 explained to do the process of detecting objects, the video must be changed to the form of image sequences first. The first step carried out for the process of converting videos into image sequences begins with video input. The next stage is the read video process, in this process, the video will be read every frame. The process that is done after the video is read is looping all frames in the video. The next process is giving name tags and resizing, and grayscale for the video to be converted to an image sequence. The last step is to create a sequence of images from the video. The number of image sequences can be calculated from the length of video times the video frame rate.

B. Deep Learning

Deep Learning is one of the techniques in machine learning that utilizes many layers of nonlinear information processing to perform feature extraction, pattern processing, and classification [8] Deep Learning is one part of Machine Learning that utilizes Neural Network techniques or artificial neural networks in solving a problem. Deep learning techniques add more layers in the learning process to be able to produce models that can better represent labeled images, so that Deep Learning techniques provide a strong architecture in learning Supervised Triano Learning.

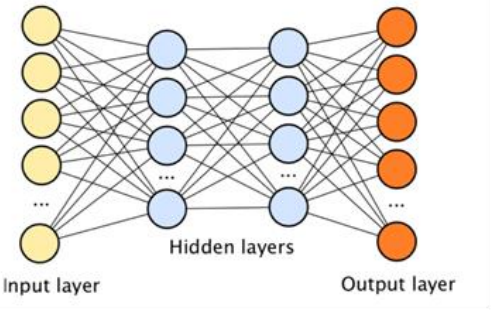

Fig. 2 Deep Learning

The idea of deep learning is to create a hierarchy of concepts with an architecture consisting of many layers,

arrange complex concepts from simpler concepts, so deep learning can represent a problem with a high degree of flexibility. The use of deep learning continues to grow until now, deep learning is often used by the research community and industry to solve various problems such as computer vision, speech recognition, and natural language processing. The use of deep-learning in a system, the CNN or Convolutional Neural Network method is very good at finding good features in the image to the next layer to form a nonlinear hypothesis that can increase the complexity of a model [9]

C. Convolutional Neural Network

The artificial neural network model that has several layers is called Multi-Layer Perceptron (MLP) where the neurons are fully connected so they have powerful classification capabilities. MLP still has shortcomings when receiving input in the form of images. The image must first go through several stages such as preprocessing, segmentation, and feature extraction to get optimal performance. This makes MLP has many free parameters that result from the formation of a full connection scheme between the input and feature maps of the related layers or in other words MLP has information overload in the architecture. The handling of these problems can be done by using another variation of MLP called Convolutional Neural Network (CNN) [10] CNN is inspired by a simple and complex cell visual mammalian cortex. This model can reduce the number of free parameters and can handle input image deformations such as translation, rotation, and scale.

Convolutional Neural Network has a way of working that is almost the same as MLP, but in CNN neurons are represented in 2D dots. MLP accepts one-dimensional input data and then propagates the data on the network to produce output. The quality of the mode is determined by each relationship between neurons in adjacent layers and has one-dimensional weight parameters. Each input on the layer is carried out linearly with the existing weight value, then the

computational results are transformed using non-linear operations called the activation function. Figure 3 is an MLP that has a layer with each layer having neurons.

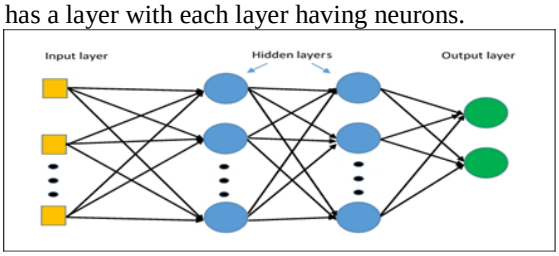

Fig. 3 Simple MLP Architecture

CNN is a method for transforming original images from layer by layer based on pixel values of images into class scoring for classification. Underlying a CNN is a convolution layer, which is called a convolution layer.

D. *Region-based Convolutional Neural Network (R-CNN)*

R-CNN is a method of detecting objects that enter the realm of computer vision-based on convolution networks or CNN. R-CNN itself was created in 2015 as an object detection method that combines the Region Proposal Network (RPN) and Convolutional Neural Network (CNN) algorithms [11] Over time this R-CNN method continues to be developed to improve performance both speed and accuracy in object detection. However, R-CNN and Fast R-CNN still have shortcomings, one of them is a bottleneck on RPN that cannot match the computing speed on CNN. The R-CNN Faster then came with optimization on the use of convolutional features to speed up the RPN process to reduce bottlenecks [12] An overview of the architecture of the R-CNN Faster can be seen in Figure 4. the following.

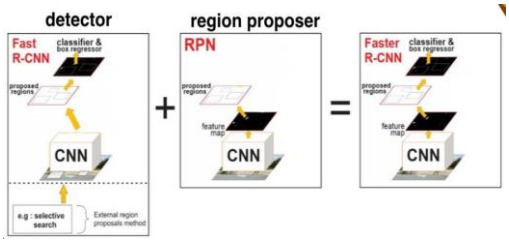

Fig. 4 R-CNN Faster Architecture

E. *R-CNN Mask*

The R-CNN mask has rapidly improved object detection and the results of semantic segmentation in a short amount of time. Much of this progress has been driven by strong baseline systems, such as Fast / Faster RCNN.

R-CNN mask, faster than R-CNN by adding branches to predict mask segmentation in each Region of Interest (RoI), in parallel with existing branches for classification and square regression. The R-CNN mask branch is a small FCN that is applied to each RoI, predicts mask segmentation in a pixel-to-pixel way. The R-CNN mask is easy to implement and trains to remember the faster R-CNN framework, which facilitates a variety of flexible architectural designs. In addition, the Mask R-CNN branch only adds a small computational overhead, enabling fast systems and fast experiments.

The R-CNN (regional convolutional neural network) mask is a framework that has two stages: the first stage is scanning images and generating proposals (areas that are likely to contain objects). And the second stage classifies proposals and produces bounding boxes and masks [13] An overview of the architecture of the R-CNN mask can be seen in Figure 5.

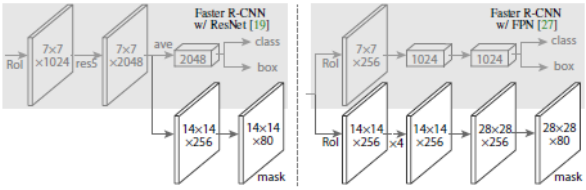

Fig. 5 Ilutration of mask R-CNN

F. *Python*

Python is an example of a high-level language. Other examples of high-level languages are Pascal, C ++, Pert, Java, and so on. Examples of low-level languages are machine language or assembly language. Simply stated, a computer can only execute programs written in the form of machine language. Therefore, if a program is written in the form of a high-level language, then the program must be processed first before it can run on a computer. This is one of the shortcomings of high-level language that requires time to process a program before the program is run. However, a high-level language has many advantages. High-level languages are easy to learn, easy to write, easy to read, and of course easy to find mistakes. High-level languages are also easily changed portable to match the machine running it. Unlike the machine language that can only be used for these machines. Many applications are written using high-level languages because of these advantages. The process of changing the form of a high-level language to low-level programming languages are of two types, namely interpreters and compilers. [14]

G. *TensorFlow*

TensorFlow is a library developed by the Google Brain Team in the Google Intelligent Machines research organization, to carry out machine learning and deep learning research [14]TensorFlow combines computational algebraic computational optimization techniques, to simplify the calculation of many mathematical expressions that require quite a long time to do the calculation. To facilitate the deeper development of Deep Learning, several CNN libraries can run on TensorFlow, such as Keras. Hard is a high level artificial neural network library written in the Python programming language. Not only on TensorFlow, but Keras can also run on CNTK or Theano.

H. *mAP*

mAP (mean average precision) is the average value of AP. AP (Average precision) is a popular metric in measuring the accuracy of the process of object detectors such as Faster R-CNN, SSD, and others. The AP calculates the average

accuracy value for recall values between 0 and 1. In several contexts. We calculate the AP of each class and averaged it. But in other contexts, AP and mAP can have values [15]

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP(q)}}{Q} \quad (1)$$

I. Box Masking

Box masking is a digital image encryption technique. Box masking is generally used in hiding information in images, and also to make the image invisible. Examples of box masking can be seen in Figure 6.

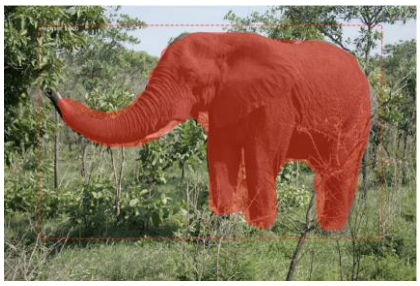

Fig. 6. Result of Box Masking Algorithm

IV. CONCLUSION

The results of this study are an application that can be used to recognize cigarette objects and provide an automatic masking effect. The application is built using the R-CNN method with the Regional Convolutional Neural Network architecture model. The cigarette object dataset used in this study consisted of 126 data, then the data is used as training and testing data. The results of the detection of cigarette images can automatically be seen in Figure 7, the red box in the picture is the result of the system's detection of objects that are considered cigarettes in the picture. Figure 8 shows the concealment of cigarette objects using box masking.

Fig. 7. Results of cigarette image detection applications

Figure 8 The results of the detection of cigarette images automatically after the masking box

ACKNOWLEDGMENT

The authors would like to express his deepest gratitude to the Department of electrical Engineering, Faculty of Engineering, Udayana University for support and constructive suggestions to help completion of this research.

REFERENCES

[1] D. C. Khrisne and I. M. A. Suyadnya, "Indonesian herbs and spices recognition using smaller VGGNet-like network," in *2018 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS)*, 2018, pp. 221–224.

[2] P. A. Wicaksana, I. M. Sudarma, and D. C. Khrisne, "PENGENALAN POLA MOTIF KAIN TENUN GRINGSING MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DENGAN MODEL ARSITEKTUR ALEXNET," *J. SPEKTRUM*, vol. 6, no. 3, pp. 159–168, 2019.

[3] I. M. Wismadi, D. C. Khrisne, and I. M. A. Suyadnya, "Detecting the ripeness of harvest-ready dragon fruit using smaller VGGNet-like network," *J. Electr. Electron. Informatics*, vol. 3, no. 2, pp. 35–38, 2020.

[4] D. C. Khrisne and T. Hendrawati, "Indonesian Alphabet Speech Recognition for Early Literacy using Convolutional Neural Network Approach," *J. Electr. Electron. Informatics*, vol. 4, no. 1, pp. 34–37.

[5] K. H. Indrani, D. C. Khrisne, and I. M. A. Suyadnya, "Android Based Application for Rhizome Medicinal Plant Recognition Using SqueezeNet," *J. Electr. Electron. Informatics*, vol. 4, no. 1, pp. 10–14.

[6] H. Filipe De Sousa Russa, "Computer Vision: Object recognition with deep learning applied to fashion items detection in images Master in Data Analysis," no. September, 2017.

[7] O. Hamalik, *Media Pendidikan*. 1986.

[8] L. Deng, D. Yu, and B. — Delft, "Deep Learning: Methods and Applications Foundations and Trends R in Signal Processing," *Signal Processing*, vol. 7, pp. 3–4, 2013, doi: 10.1561/2000000039.

[9] K. P. Danukusumo, "Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis GPU"," 2017.

[10] M. Zufar and B. Setiyono, "Convolutional Neural Networks Untuk Pengenalan Wajah Secara Real-time," *J. Sains dan Seni ITS*, vol. 5, no. 2, p. 128862, 2016.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[14] D. N. Rizky, "Deteksi Tanda Nomor Kendaraan Bermotor Pada Media Streaming Dengan Algoritma Convolutional Neural Network Menggunakan Tensorflow," no. March, 2018.

[15] Y. Yue, T. Finley, F. Radlinski, and T. Joachims, "A support vector method for optimizing average precision," *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR'07*, pp. 271–278, 2007, doi: 10.1145/1277741.1277790.