

Automatic Stranger Remover in Photo by Deep Learning Segmentation

Akeem Olowolayemo
Faculty of Cognitive Sciences
and Human Development,
Universiti Malaysia Sarawak,
Sarawak, Malaysia
oakeem@unimas.my

Saleh Alanazi
Faculty of Information &
Computer Science, University
Kebangsaan Malaysia
sala207@hotmail.com

Lim Yoong Kang, Doreen
Teoh Sim Ying
Faculty of Cognitive Sciences
and Human Development,
Universiti Malaysia Sarawak,
Sarawak, Malaysia
yoongkang0122@gmail.com

ABSTRACT

A new algorithm utilizing deep learning segmentation is proposed for removing stranger in digital photos. The proposed algorithm consists of four main steps which are SIFT descriptor alignment, deep learning segmentation, bounding boxes detection and Poisson blending. VGG 16-layer convolution network is adapted and modified by replacing the final output layer with the exact shape of the input layer for output image with stranger segmented. Automatic stranger remover has considerable potential in removing stranger in photo and returns a seamless photo for users in a faster and effective way.

CCS Concepts

• Computing methodologies → Image processing.

Keywords

Deep learning; Fully Convolutional Networks (FCN); Automatic Stranger Remover.

1. INTRODUCTION

Object removal from images or videos is one of the important tasks in photo editing. Manual processes are often employed in photo editing software. The success of the process hinges primarily on the expertise of the human experts performing the editing. From the foregoing, it is desirable that unwanted objects in images or videos can be automatically removed using a combination of processes. Many techniques and processes are involved for human objects removal in an image. Typical process involves multiple stages. Firstly, detection of the unwanted objects is required from the background of the images. Usually, background subtraction is used for object tracking. This is however challenging to apply when the image background is not static but dynamic.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICIGP 2018, February 24–26, 2018, Hong Kong, Hong Kong

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6367-9/18/02...\$15.00

<https://doi.org/10.1145/3191442.3191475>

Therefore, the system proposed in this work is to develop an automatic system that can remove stranger in images and then restore the removed part with multiple image inputs. The system proposed in this work is enhanced using deep learning for segmentation. A database is created and two types of input are inserted comprising the frame with the unwanted human object and another frame without the unwanted object. These two inputs are required so that the deep learning segmentation network can be performed.

The Automatic Stranger Remover proposed in this work typically involves four steps. The first step is the SWIFT descriptor Alignment, which is followed by the deep learning segmentation using the database created from the inputs. The subsequent step consists of the bounding boxes detection of the layer after segmentation. The last step is Poisson blending where the system blends the photo so it can replace the stranger masked with white colour. Finally, a seamless photo with the unwanted stranger removed will be output as the result. These steps are discussed in greater details after a review of related work in the subsequent sections.

2. RELATED WORK

2.1 Fully Convolutional Network for Semantic Segmentation

Long, Shelhamer and Darrell proposed fully convolutional networks (convnets) for semantic segmentation. The fully convolutional networks are very efficient in inference and learning for semantic segmentation [1]. A fully convolutional network is often trained from end to end or by considering pixels to pixels in the process of semantic segmentation. Besides, a fully convolutional network takes the input of arbitrary size and produces correspondingly similar sized output with efficient inference and learning.

Convolving the original filter with upsampled output does not produce the same result as shift and stitch but the original filter only realizes a reduced portion of the input. In order to reproduce the shift and stitch, rarefy the filter with the formula:

$$f'_{ij} = \begin{cases} f_{i/s, j/s} & \text{if } s \text{ divides both } i \text{ and } j; \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

Decreasing the subsample within a net is a tradeoff where the filter detected finer information but take longer time to compute. The shift and stitch technique is another kind of tradeoff where the output is denser without decreasing the receptive fields.