# Face and Information Content Detection for Automated Blurring

* CSE509 Digital Video Processing Final Project

Kuriakose Eldho
*SCAI*
*Arizona State University*
Tempe, Arizona
keldho@asu.edu

Raghav Kapoor
*SCAI*
*Arizona State University*
Tempe, Arizona
rkapoo22@asu.edu

Jisoo Lee
*SCAI*
*Arizona State University*
Tempe, Arizona
jlee815@asu.edu

***Abstract*—This project focuses on developing an automated system for face and information content detection in videos, with the ultimate goal of facilitating efficient automated blurring. The study aims to enhance accuracy in identifying and blurring faces and textual elements. The implementation of recent advancements in face and text detection from contemporary research further contributes to the project's cutting-edge approach. The proposed system addresses privacy concerns by automating the process of obscuring sensitive information, presenting a valuable solution for applications in video content moderation and privacy preservation.**

***Index Terms*—object tracking, face detection, text detection, Kalman filter**

## I. Introduction

Within the progression of artificial intelligence technology, effective data collection emerges as a pivotal component. Video recordings, ubiquitous in settings like CCTV and automobiles, form a substantial part of this data landscape. Nonetheless, instances arise where collected data harbors sensitive information, yet robust protective measures may be lacking. Notably, in specific scenarios like video footage from military bases, there arises a necessity to conceal pivotal segments when disseminating content to the public. This research project seeks to automatize the identification and obfuscation of critical information within video recordings, encompassing elements such as human faces and text, employing sophisticated deep-learning methodologies to enhance precision and efficiency in safeguarding confidential content.

## II. Methodology

The application of automatic detection and blurring encompasses four key stages. Our emphasis is on identifying and blurring sensitive information, specifically faces and text, which represent the first classes of interest. The detection process is primarily focused on localization rather than identification, as our objective is to identify all objects within the images. Subsequently, time information is integrated to enhance the overall detection performance. The final step involves applying blurring to the localized positions.

Figure 1 shows the model architecture involves feeding images extracted from the video into two types of detectors. The resulting bounding boxes' locations are refined using a Kalman filter. Ultimately, Gaussian blurring is applied to the updated bounding boxes to achieve the desired level of anonymization.

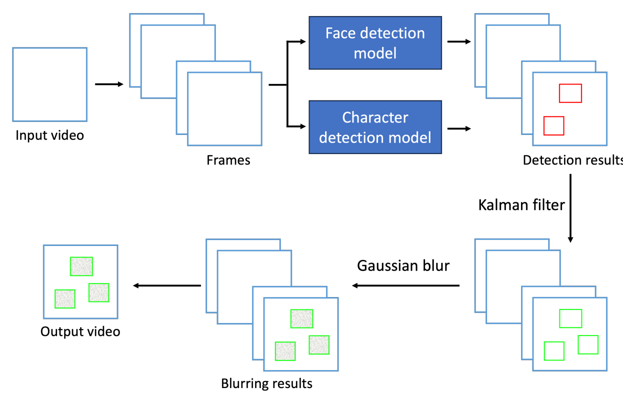


Fig. 1. Model Architecture

### A. Character Recognition

In the context of character detection, this study employs a pre-trained model integrated within the EasyOCR library. Specifically, the utilized model is based on the CRAFT [1] (Character Region Awareness for Text Detection) architecture, developed in 2019. CRAFT implements a weakly supervised learning approach within a VGG16 Deep Neural Network [2] to detect text within a given frame. Upon inputting a frame into the model, the system outputs the precise coordinates of bounding boxes encapsulating the identified text. This methodology not only leverages the established EasyOCR library but also integrates the robust capabilities of the CRAFT model to achieve accurate and efficient character detection within video frames.

Fig. 2 shows CRAFT input-output examples for a real world image with bouding box coordinates plotted on the input image

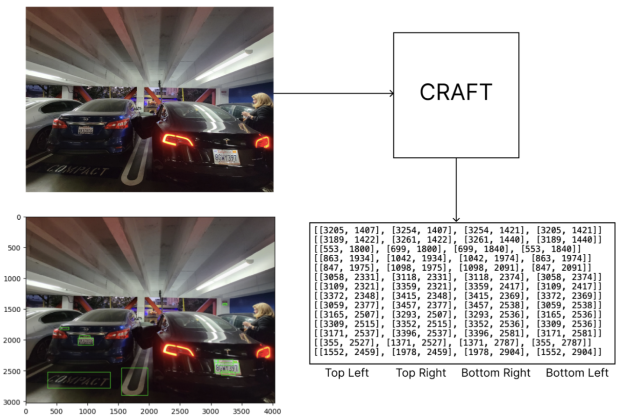


Fig. 2. Character Detection I/O with result plots for real world image

### B. Face Detection

For the face detection part of our project, we have used a pre-trained DNN module from OpenCV, which contains a 'Single Shot Multibox Detector' [3] object detection model with ResNet-10 [4] as its backbone architecture. This approach helps in detecting faces in real-time and with high accuracy.

Single Shot Multibox Detector (SSD) is an efficient and popular framework for object detection in images. SSD performs detection directly on feature maps of different scales. It predicts bounding boxes and class probabilities for objects at multiple scales within a single forward pass of the neural network. ResNet (Residual Network) is an architecture known for its deep convolutional neural networks that use residual learning blocks. The ResNet-10 backbone provides a balance between performance and computational efficiency, making it suitable for object detection tasks within the SSD framework.

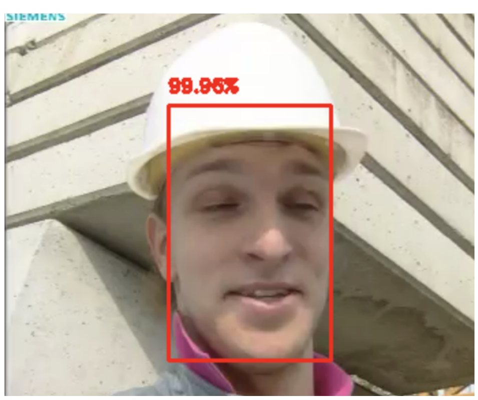


Fig. 3. Output of face detection using OpenCV-DNN

### C. Tracking

The problem of tracking moving objects in videos fundamentally originates from the challenge of detecting objects in images. Therefore, the effectiveness of object tracking heavily relies on the performance of object detection in images. With the advancements in deep learning technologies, numerous models for detecting objects in images have been developed. From notable examples like Faster R-CNN [5] to high-performing models like YOLO (You Only Look Once) [6], detectors have made significant strides in achieving rapid and accurate object detection.

In this project, as we are dealing with video data that includes temporal information rather than static images, we have chosen to treat the data as time series data. Leveraging this temporal aspect, we aim to enhance detection performance for each frame. A prominent method of incorporating temporal data into these detection results is the Kalman filter [7]. The Kalman filter is a mathematically effective algorithm utilized for estimating incomplete or noisy measurements, making it widely applied in various fields such as navigation, computer vision, and signal processing. Its effectiveness lies in its ability to enhance the estimation of measurements by mitigating the effects of incompleteness or noise.

In this application, the Kalman filter has been applied to the face detection part. In cases where the result of face detection is obtained or when faces are missed, the Kalman filter is utilized to correct or predict these occurrences.

The Kalman filter needs to be initialized for each object and applied to track each object in subsequent frames. To achieve continuous detection of multiple faces, we have devised an algorithm. If, within a given set of frames, the predominant number of faces is not detected, we apply Kalman filters corresponding to that specific number of faces. This approach is updated as new frames arrive, and variations, such as adjusting the window size to accommodate the predominant number of faces within it, are also feasible.

### D. Blurring

After applying the face detection model , character detection model and the Kalman Filter on a video frame, we get a list of bounding boxes. We use a Gaussian filter with a high sigma value to blur out the contents of the detected bounding boxes. This results in all detected sensitive information in the frame to be hidden.

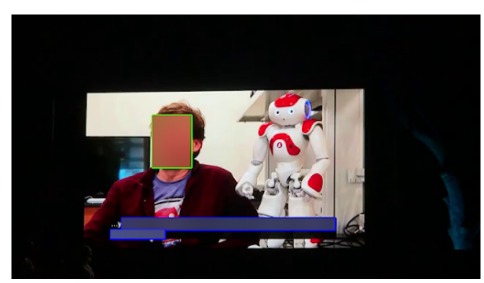

Fig. 4. Output after blurring out sensitive information

## III. Results

Before obtaining the final video result, we extracted images immediately after the application of the Kalman filter to assess its effectiveness. For face detection, we utilized OpenCV-DNN, and relying solely on the performance of this model, we observed that faces were not detected after the 8th frame.

However, following the application of the Kalman filter, we confirmed the ability to identify missing detections.

In Figure 5, the images in the first row depict the detection results without the application of the Kalman filter, indicated by green bounding boxes. The second row represents the images after the application of the Kalman filter. In this row, corrected bounding boxes and prediction-based bounding boxes for undetected cases are marked in green. The results solely relying on OpenCV-DNN are highlighted in red. The dataset was obtained from YouTube, featuring a scenario where only one face appears.

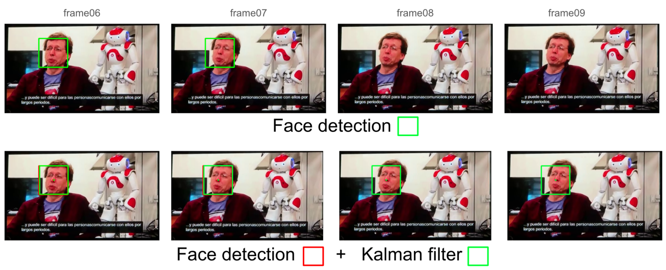


Fig. 5. Face detection and Kalman filter results

The final results are shown in Figure 6. The test data is a sports video from YouTube. After processing the video using our application, all informative parts have been blurred.


Fig. 6. Before and after hiding sensitive information in video frame

## IV. Conclusion

In conclusion, our research project, "Face and Information Content Detection for Automated Blurring," addresses the imperative need for robust data protection in artificial intelligence. The integration of sophisticated Video Processing methodologies, including the Kalman filter, has proven pivotal in enhancing detection precision. Notably, the CRAFT and OpenCV-DNN's limitations in character and face detection were mitigated by the Kalman filter, rectifying missed detections(for eg- beyond the 8th frame in figure 3)

Figure 3 vividly illustrates the transformative impact of the Kalman filter, marked by corrected bounding boxes and prediction-based boxes for undetected cases in green, contrasting with OpenCV-DNN only results in red. This dynamic adaptability fortifies the system's reliability in face and information content detection. In essence, our research contributes to advancing automated solutions for privacy preservation, emphasizing the importance of advanced techniques for enhanced data security in the era of artificial intelligence and widespread video recording.

## References


[1] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," 2019.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[4] J. Gong, W. Liu, M. Pei, C. Wu, and L. Guo, "Resnet10: A lightweight residual network for remote sensing image classification," in *2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 2022, pp. 975–978.

[5] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[6] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015. [Online]. Available: http://arxiv.org/abs/1506.02640

[7] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.