TRACKING AND RECOGNIZING REAL-TIME CREATIVITY

Problem Statement

- The emerging trend of social messaging applications along with time-constrained jobs, leads to requirement of improving the messaging environment such that it would become more expressive.
- Track the real-time drawn creativity of user over any social messaging platform to encourage less typing and more expressive theme.
- Recognize the objects/emojis saved as a result of tracking in such a way that user need not search for it in the keyboard.

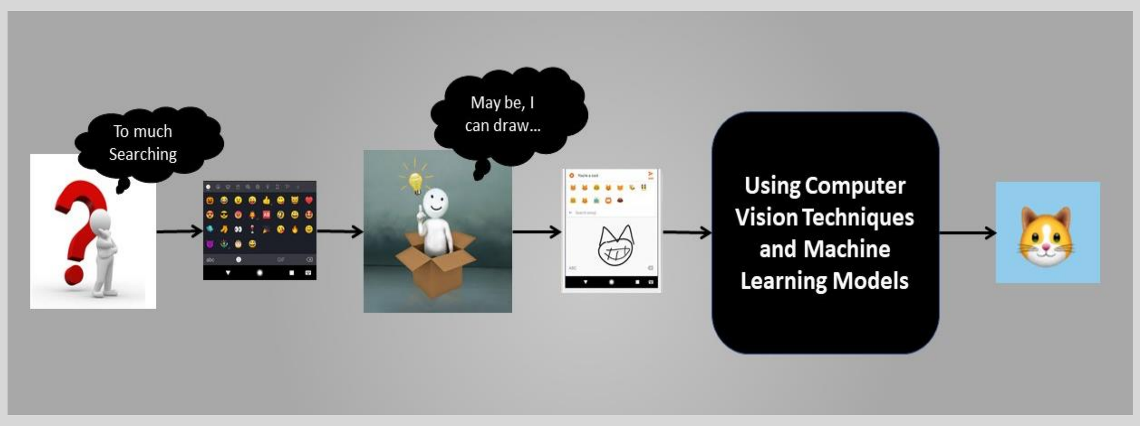

Figure 1 : Problem Statement

Literature Review

- Sketch recognition remained a topic of interest for research in the field of computer vision, but most of the sketch recognition algorithms depends on hand-crafted feature extraction, so deep neural network is proposed by Yongxin et. al.[2] for the sub-task, feature extraction.
- Air written data is way of writing characters and words in free space using finger or any object without aid of hand held device. Soham et al.[3] proposed an optimized algorithm for contour detection and tracking of real-time air written data using the Faster R-CNN.
- DNNs have not yet been effectively explored so a deep convolutional neural network is proposed by Omar et al.[4] which can be used for both classification and medium/high-level feature extraction of images.

Baseline

- The pin pointer pen is used for drawing of images in air which comprises of locating and tracking the motion of the pen using webcam. Apart from pen, a particular shaped object also helps in drawing the image in air. This technique is helpful when drawing in air is restricted for particular audience i.e. people provided with that object.
- To detect hand regions/object out of the frames, Background Subtraction technique such as Opencv's BackgroundSubtractorMOG2, BackgroundSubtractorkNN, and running median are helpful, the results shows that kNN Subtractor works better if only foreground detection is the required task but not good for tracking.
- Feature Tracking algorithm for the purpose of hand tracking to get the drawn image helps when the image has to be drawn using full hand and without considering any finger point.
- Naïve Bayes classifiers worked on the basis of Bayes' theorem, which describes the probability event. It is generative model which is based on prior knowledge of conditions. Naive Bayes Classifier are very simple and fast classification algorithm often suitable for high dimensional datasets.

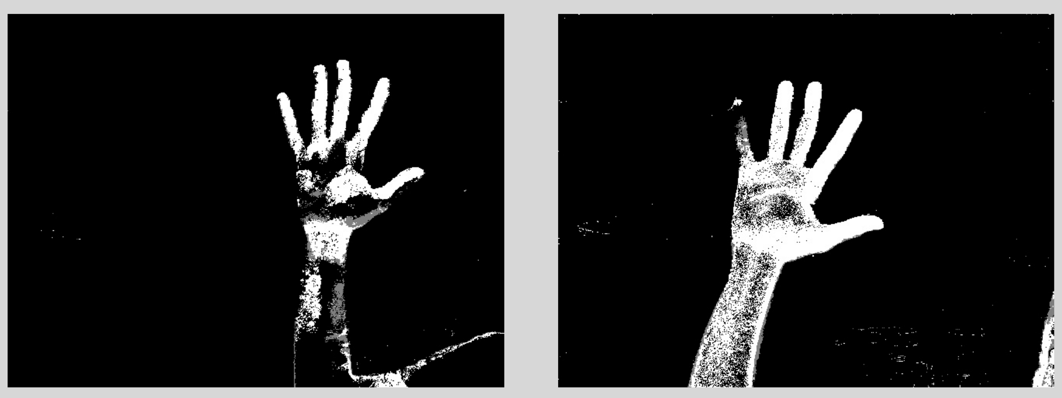

Figure 3: BackgroundSubtractorMOG2 & BackgroundSubtractorkNN

Data-Set Description

- World's largest doodling dataset, QuickDraw by Google[1] consists of 345 categories and over 50 million of drawings.
- The dataset consist of 28x28 gray-scale image form and stored in numpy(.npy) format.
- Currently, We sampled out 10 random categories from the whole dataset to reduce computational time ,but it can be extended later anytime for arbitrary any number of categories.

Dataset Visualization

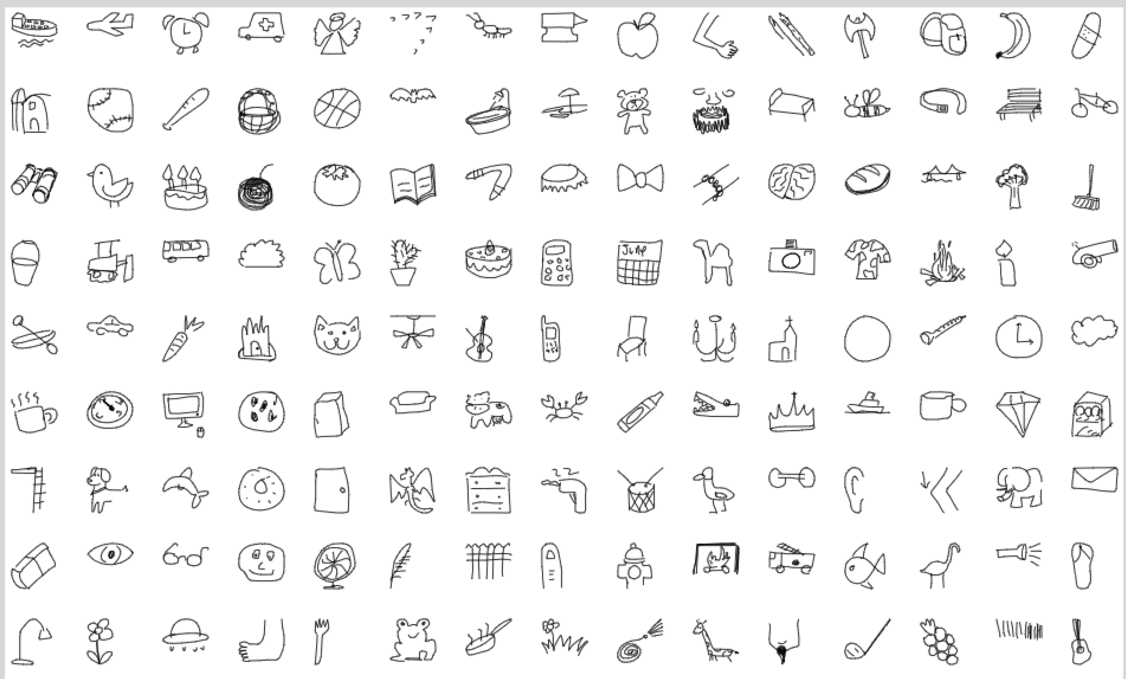

Figure.2: Quick-Draw Dataset

Proposed Algorithm

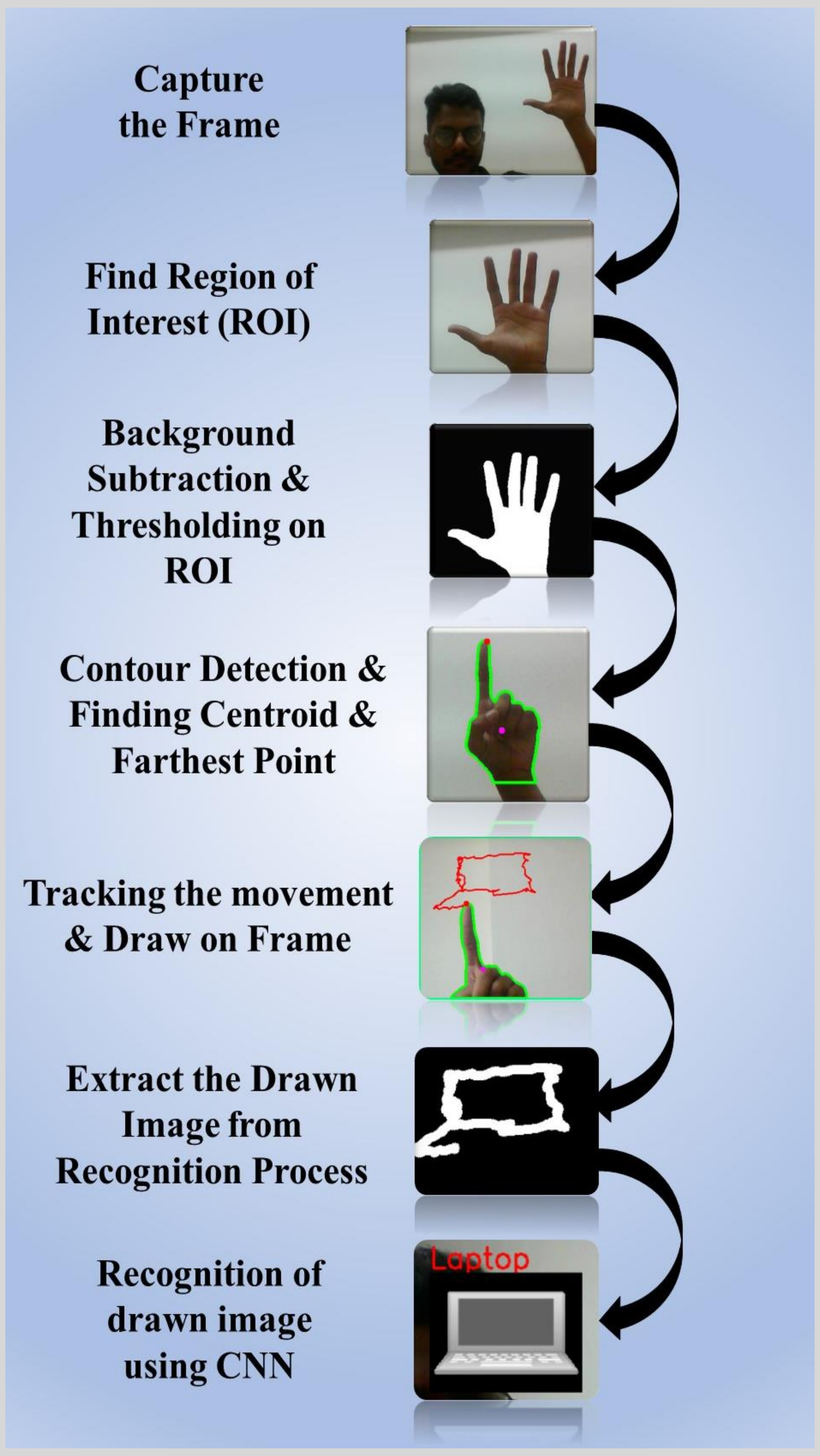

Figure 4 : Flow of Proposed Approach

- Step 1: Detect hand region out of the frame using running average differences among the frames.
- Step 2: Detect hand gestures to come up with framework of drawing using only single finger.
- Step 3: Tracking of drawing by detecting the finger in each frame and connecting the pin-point coordinates of the finger
- Step 4: Extract the drawn image from the window and pass it to the recognizer.

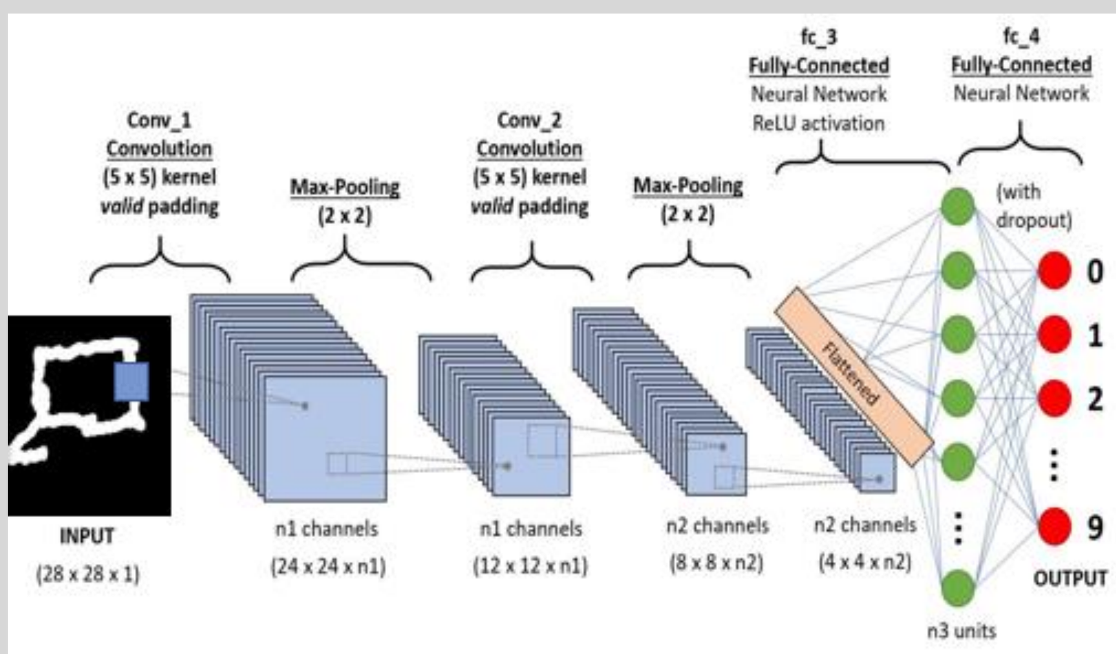

Figure 5 : Convolutional Neural Network Architecture

•Step 5: Recognize the drawn image using Convolutional Neural Network, for predicting the actual label of drawn image.
•Step 6: Return the actual image of the corresponding drawing.

Results & Analysis

Classifiers	Gaussian Naïve Bayes	Random Forest	XG_Boost	Convolutional Neural Network
Accuracy	66.50	90.188	91.67	94.65

Figure 6 : Performance of Different Classifiers

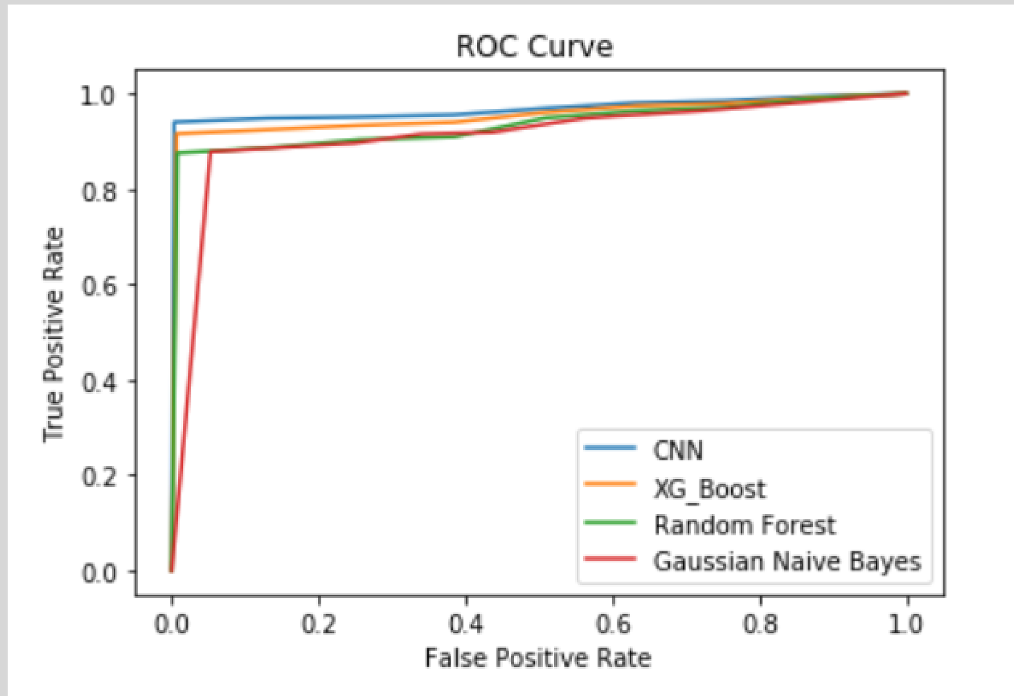

Figure 7 : ROC Curve for Different Models

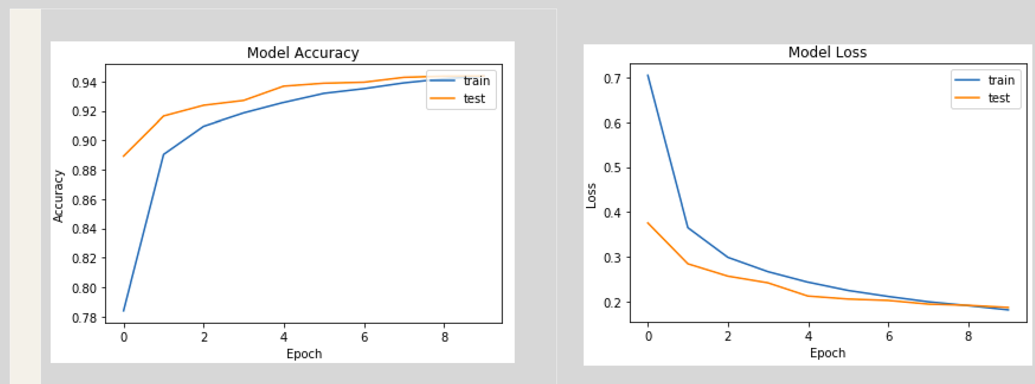

Figure 8 : Analysis of Accuracy and Loss per epoch in CNN

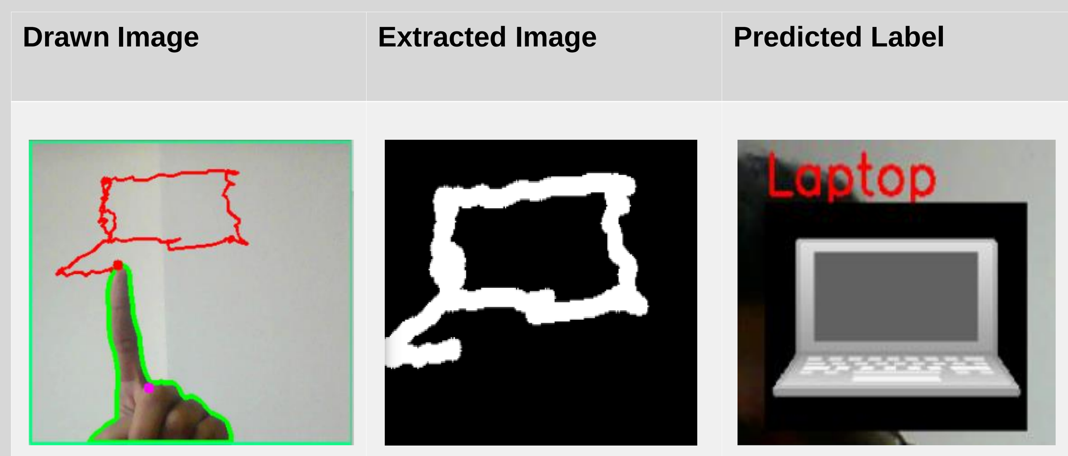

Figure 9: Predicted Object

Interpretation of Results

- Using finger to draw creativity instead of pin-point pen, increases the user convenience and also the target audience for the system.
- Background subtraction using running averages over the frames is able to distinguish finger from other background objects more reliably.
- Tracking of drawn image by detecting the hand in each frame makes the drawing accurate and doesn't lack the finger point in between which was occurring very frequently in the case of feature tracking algorithm.
- Using CNN as a recognizer is able to make classification among the categories more accurately.

Conclusion & Future Work

- The proposed methodology is giving outperforming results over the baselines very significantly. This encourages the use of finger over the pen/object, use of running averages concept for background subtraction, follow the finger detection in every frame for tracking, and use of CNN as a model for recognizing the drawn creativity accurately.
- The proposed approach can be extended to the recognition of social messaging apps emoji and help in social norms.

References

[1] https://quickdraw.withgoogle.com/data
[2] Yang, Yongxin, and Timothy M. Hospedales. "Deep neural networks for sketch recognition." arXiv preprint arXiv:1501.07873 (2015).
[3] Seddati, Omar, Stephane Dupont, and Sad Mahmoudi. "Deepsketch: deep convolutional neural networks for sketch recognition and similarity search." In Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on, pp. 1-6. IEEE, 2015.
[4] Mukherjee, Sohom, Arif Ahmed, Debi Prosad Dogra, Samarjit Kar, and Partha Pratim Roy. "Fingertip Detection and Tracking for Recognition of Air-Writing in Videos." arXiv preprint arXiv:1809.03016 (2018).