Person Identification Using Image and Voice

Sun Yitong

Background and Objective

Identification of people is woven into our daily life. Extracting facial features and identifying people remains a challenge when it is happening under some conditions such as when the environment is dark or when the face is covered by cloths.

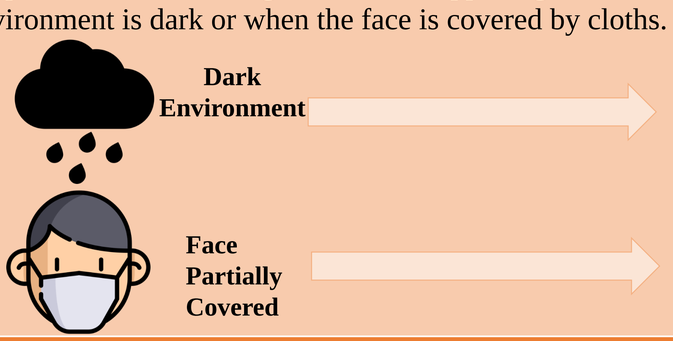
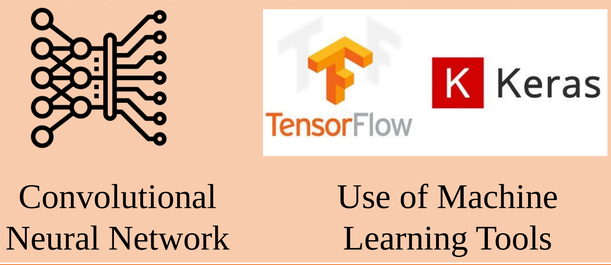


Figure 1

This project explores the possibility of applying deep learning methods combining both image and voice recognition systems to identify a person when one of them is not working well. The results have shown that when both facial features and vocal features are used, accuracy of identifying the person improves significantly.



Convolutional Neural Network

Use of Machine Learning Tools

Better Accuracy

Figure 2

Material and Methods

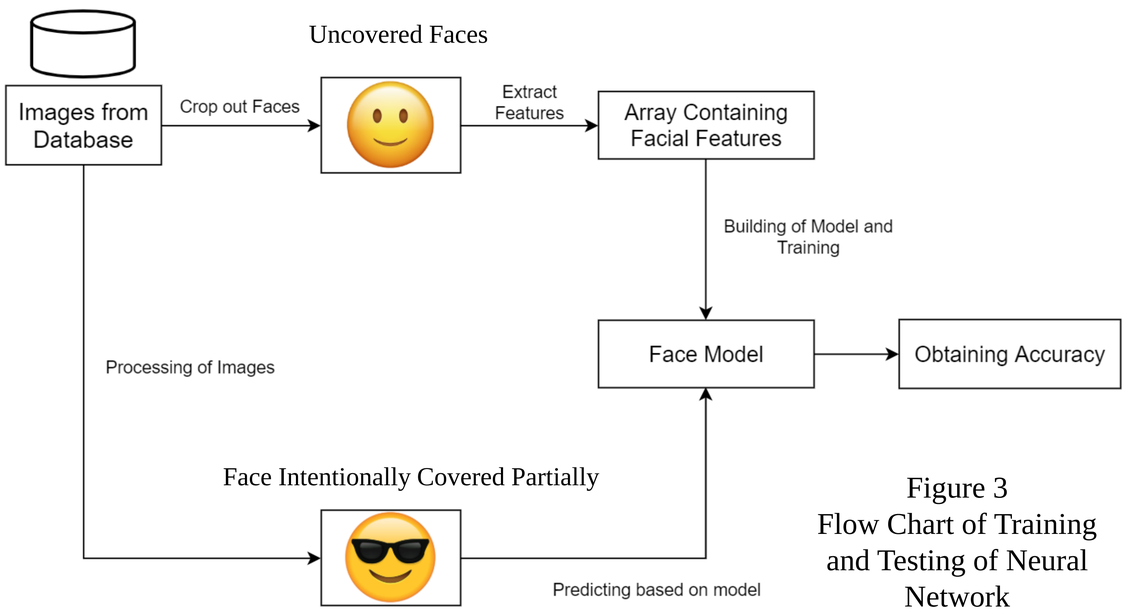


Figure 3
Flow Chart of Training and Testing of Neural Network

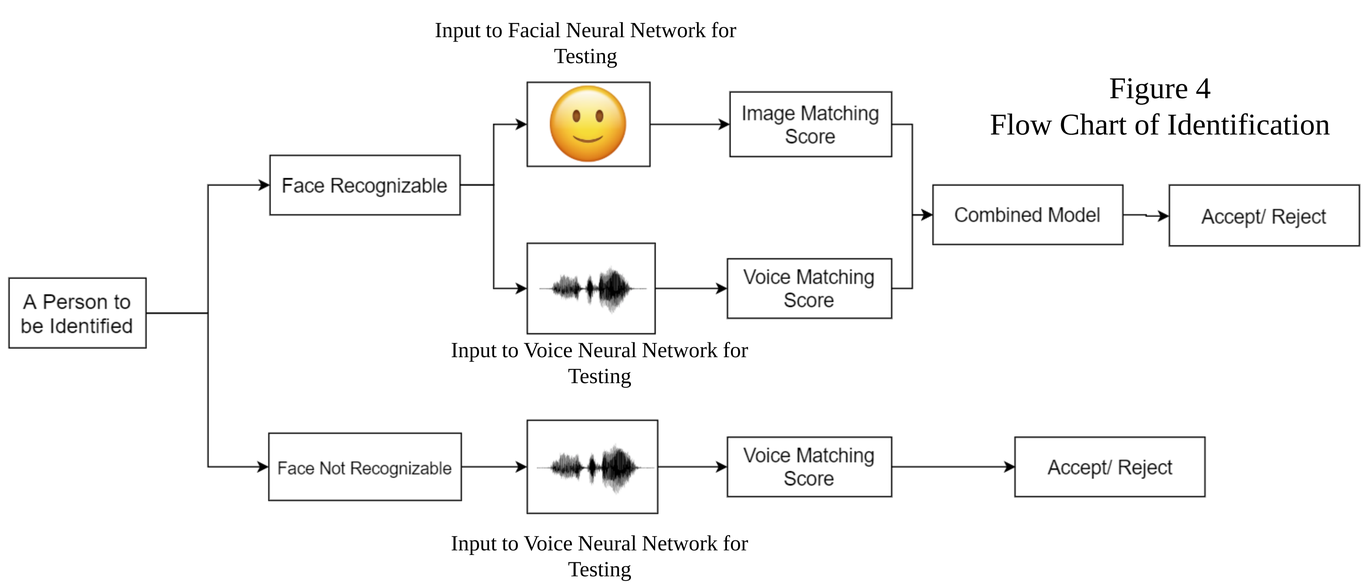


Figure 4
Flow Chart of Identification

Under the circumstances that the person's face can be seen, to combine both methods, the confidence value of the model can be calculated by multiplying the confidence value of both the face model and voice model. If the combined confidence level is below the threshold, the identification will be output as unknown.

$$Pcombined = \frac{Pface}{Pface\ max} + \frac{Pvoice}{Pvoice\ max}$$

$Pface$: the confidence value of facial recognition
$Pvoice$: the confidence value of voice recognition
$Pface\ max$: the maximum value of facial recognition for the image
$Pvoice\ max$: the maximum value of voice recognition for the same voice clip

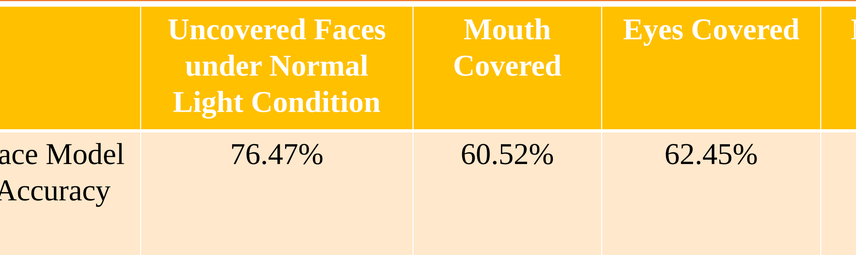
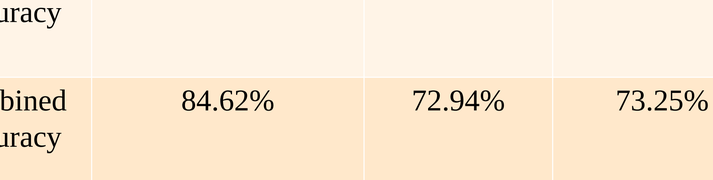


Figure 5
Processing of Images for Training



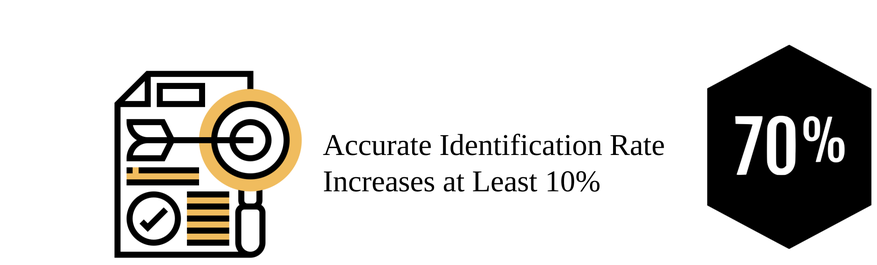


Figure 6
The processing of Voice Clips for Training

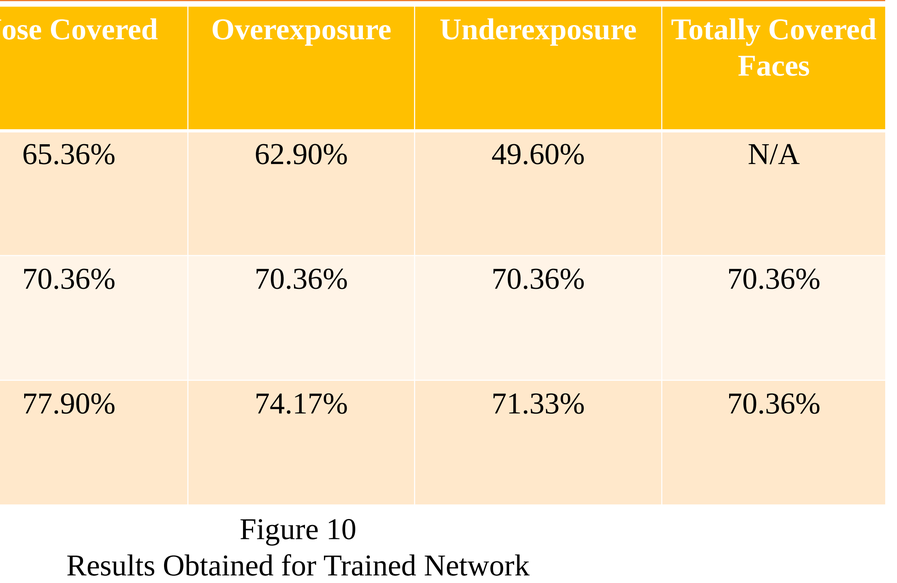
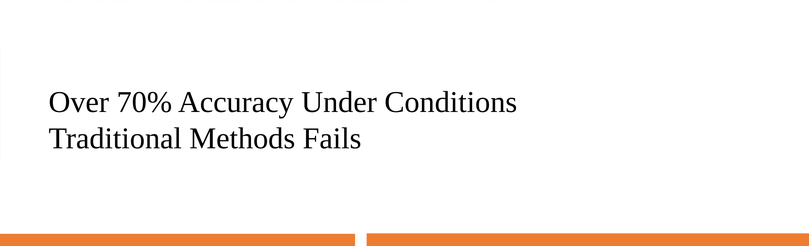


Figure 7
Processing of Images for Testing



Results and Discussion

	Uncovered Faces under Normal Light Condition	Mouth Covered	Eyes Covered	Nose Covered	Overexposure	Underexposure	Totally Covered Faces
Face Model Accuracy	76.47%	60.52%	62.45%	65.36%	62.90%	49.60%	N/A
Voice Model Accuracy	70.36%	70.36%	70.36%	70.36%	70.36%	70.36%	70.36%
Combined Accuracy	84.62%	72.94%	73.25%	77.90%	74.17%	71.33%	70.36%

Figure 10
Results Obtained for Trained Network

Figure 8
Face Model

Figure 9
Voice Model

Deep neural networks are used in both the training of face model and voice model. After generating the models for the facial features and voice features, test sets are put into the models to evaluate the accuracy of the models. The structures of the neural networks are shown in Figure 8 and Figure 9.

Accurate Identification Rate Increases at Least 10%

70%

Over 70% Accuracy Under Conditions Traditional Methods Fails

References

[1]CarstenKorfmacher,OxfordUniversity.PersonalIdentity
[2]DeLeeuw,Karl;Bergstra,Jan(2007).TheHistoryofInformationSecurity:AComprehensiveHandbook.
[3]VisualGeometryGroup,(2014).VGGFace2.Retrievedfrom http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/
[4]AISHELLTechInc,(2019).Retrievedfrom http://www.aishelltech.com/kysjcp.
[5]HaythamFayek,(2016).SpeechProcessingforMachineLearning:Filter banks,Mel-FrequencyCepstral Coefficients(MFCCs)andWhat'sIn-Between
[6]JamesLyons,(2013).Python_speech_features'sDocumentation

Future Work

Noise Cancelling for Better Clarity of Voice Clips to Obtain Better Accuracy

Living Body Detection e.g. Eye-blinking Detection to Ensure Real Person Presence