Real-Time Computer Vision ASL Translator

First-Year Engineering Design, Advisor: David Han

Ian Guiwo, Kevin Frank, Vincent Vassallo

Abstract

- As an increasing number of employers permanently transition to hybrid work environments [1], it is increasingly crucial to bridge the gap between the auditory disabled and their coworkers in a professional workplace. Audiologists predict that these demographics will only rise in time [2]; to combat these hardships, as well as the ever-increasing price of hearing aids [3], we have come up with a *Real-Time Computer Vision ASL Translator*.
- We have coded a **working prototype** that is compatible with all standard computers to promote ease-of-use and affordability.

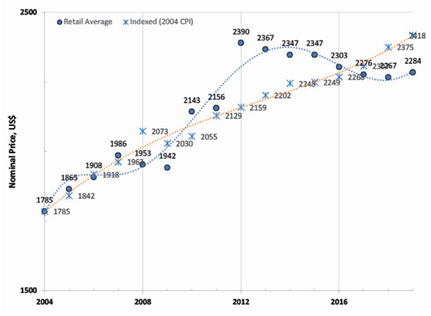


Fig. 1 *"...compares the unadjusted, average, single-unit retail hearing aid prices, in US dollars, and a 2004 average-priced, single unit retail hearing aid adjusted for inflation (blue asterisks) across all technology tiers"* [3]

Objectives

- Create an algorithm that recognizes American Sign Language
- Achieve an accuracy of at least 80%
- Translates the signs to both visual text and auditory speech
- Utilize Tensorflow and Keras libraries to identify signs
- Train using Google's Teachable Machine
- Use external webcam to imitate camera's functionality:

Webcam, used to capture photos and replicate hardware component

With this, we hope to assist the auditorily disabled and provide them a simple, affordable alternative to their current expensive counterparts.

Methodology

The program has been trained and repeatedly tested and retrained to improve upon its previous iterations.

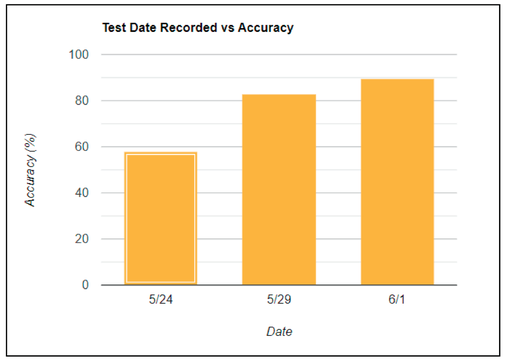


Fig.2 *Displays accuracy improvement over time. Accuracy improved to 58.06%, 82.75%, 89.65% respectively*

Retraining consisted of the implementation of more data. Photos were captured through class 'image_capture'. ~ 300 to 1000 photos utilized per sign, used to train Teachable Machine algorthim.

Societal Impact

Technologies that involve object identification and computer vision have been historically discriminant towards disabled people [4]. The goal of modern society is to bring equality and equal access to everyone - our product was designed with that same goal in mind. The target group we are aiming to help is the deaf community and anyone who relies on sign language to communicate. Our product would provide smooth translation and make holding a conversation much easier for those who use ASL. For example, our product can easily be attached and linked into online meetings when there is a deaf team member to ensure there is clear communication between everyone.

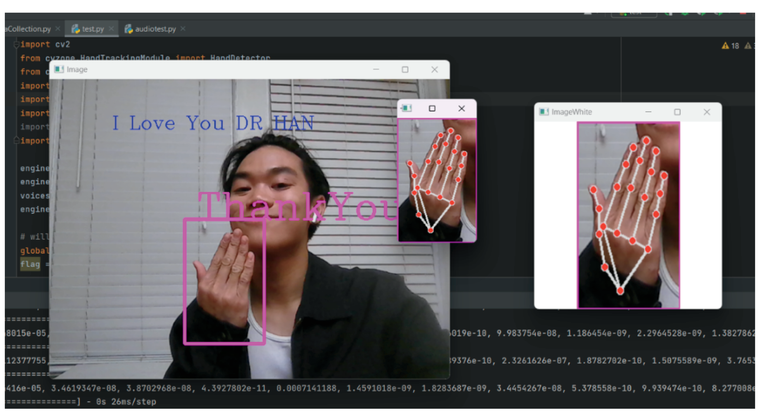
From an environmental and sustainability standpoint we want to create the most eco-friendly product possible. To manufacture, we can use recycled plastics like many other tech companies are doing to try and minimize waste. Like most cameras, the device can ultimately be repurposed rather than thrown away [5], promoting recyclability and a circular economy.

Results

```python
engine = pyttsx3.init()
engine.setProperty('rate', 125)
voices = engine.getProperty('voices')
engine.setProperty('voice', voices[1].id)

# will be above  #camera
global flag
flag = False

1 usage
def speak_to_me(words):
    engine.say(' '.join(words))
    engine.runAndWait()

1 usage
def display_text(words, flag):
    nw = labels[index]

    # if a word is detected not splicing is required
    if len(nw) > 1:
        if keyboard.is_pressed('a'):
            words.append(nw)

    # flag is default false, when flag is true letters will build a word a ---> an in the list
    # when flag is false a new word will begin a ----> a n
    elif flag:
        if keyboard.is_pressed('a'):
            words[-1] += nw
    else:
        if keyboard.is_pressed('a'):
            flag = True
            words.append(nw)

    return words, flag
```

speak_to_me and display_text functions

- Translates ASL with at least an **89.65**% accuracy
- Resizes given image to fit 300x300 px box size (main function sourced from [6])
- Displays algorithm's prediction with **cvzone** ClassificationModule
- Original function prints on screen prediction to text
- Speaks text out loud using **pyttsx3**
- Simple controls to capture images, record letters and clear console



Sample Test, "I Love You Dr Han"

```python
imgWhite = np.ones((imgSize, imgSize, 3), np.uint8)*255
imgCrop = img[y - offset_: y + h + offset, x - offset_: x + w + offset]

aspectRatio = h/w
if aspectRatio > 1:
    k = imgSize/h
    wCal = math.ceil(k*w)
    imgResize = cv2.resize(imgCrop, (wCal, imgSize))
    wGap = math.ceil((imgSize - wCal)/2)
    imgWhite[:, wGap:wCal+wGap] = imgResize
    prediction, index = classifier.getPrediction(imgWhite, draw_=False)
    print(prediction, index)

else:
    k = imgSize / w
    hCal = math.ceil(k * h)
    imgResize = cv2.resize(imgCrop, (imgSize, hCal))
    hGap = math.ceil((imgSize - hCal) / 2)
    imgWhite[hGap:hCal + hGap, :] = imgResize
    prediction, index = classifier.getPrediction(imgWhite, draw_=False)

cv2.putText(imgOutput, labels[index], (x,y-20), cv2.FONT_HERSHEY_COMPLEX, 2, (255,0,255), 2)
cv2.rectangle(imgOutput, (x-offset,y-offset), (x+w+offset, y+h+offset), (255,0,255), 4)

words, flag= display_text(words,flag)
cv2.putText(imgOutput, ' '.join(words), (100, 80), cv2.FONT_HERSHEY_COMPLEX, 1, (255, 0, 0), 1)
if (len(words) > 3 and not flag):
    speak_to_me(words)
    words = []
elif keyboard.is_pressed('z'):
    words[-1] = ''
cv2.imshow("ImageCrop", imgCrop)
cv2.imshow("ImageWhite", imgWhite)
```

Main Function, resizing and displaying image

Conclusion

We were able to integrate our code into Teachable Machine's algorithm and ultimately produce a functional ASL translator with an accuracy of at least 89.65%. We accomplished our goals of producing a translator from ASL to text and voice, as well as learned about computer vision and the training involved in machine learning.

References

[1] M. Amanda, "How Sign Language Evolves as Our World Does," *The New York Times*, July 27, 2022. [Online]. Available: https://www.nytimes.com/interactive/2022/07/26/us/american-sign-language-changes.html

[2] World Health Organization, "WHO: 1 in 4 people projected to have hearing problems by 2050," World Health Organization, March 2021. [Online]. Available: https://www.who.int/news/item/02-03-2021-who-1-in-4-people-projected-to-have-hearing-problems-by-2050

[3] A. Amyn, "An Analysis of US Hearing Aid Pricing: Insights and Commentary for the Practicing Manager," Audiology Practices, 2023, [Online]. Available: https://www.audiologypractices.org/an-analysis-of-us-hearing-aid-pricing

[4] B. Lydia et al. "Ableism And Disability Discrimination In New Surveillance Technologies," Center for Democracy and Technology, [Online]. Available: https://cdt.org/insights/ableism-and-disability-discrimination-in-new-surveillance-technologies.

[5] "What type of camera can be recycled?" Cohen USA, 2023, [Online]. Available: https://www.cohenusa.com/recyclable-items/cameras/ (Accessed May 12, 2023).

[6] M. Hassan, "Easy Hand Sign Detection | American Sign Language ASL | Computer Vision," YouTube, July 4, 2022. Available: https://www.youtube.com/watch?v=wa2ARoUUdU8&list=LL&index=11&t=2874s