

[Deep Learning] Hand gesture recognition



Yacine BENAFFANE

Follow



Jan 14, 2020 · 5 min read



Figure 1: Hand gesture (<https://www.npr.org/2019/09/26/764728163/the-ok-hand-gesture-is-now-listed-as-a-symbol-of-hate>)

Gesture recognition is a hot topic in computer vision and pattern recognition. Although great progress has been made recently, fast and robust hand gesture recognition remains an open problem, since the existing methods have not presented a practical compromise between the performance and the efficiency (source: <https://arxiv.org/abs/1901.04622>).

The project started in a compagny that wanted to create an AI capable to predict the sign language. So i want to share my experience, because it's an interesting aspect of our society, the signs will be able to help, depending on the case, to establish and improve existing communication. So with an AI capable to predict these signs, which will ease communication and machine manipulation for speech-impaired persons

We can use similar gestures as in sign language for my own product and create a customizable range of gestures for the model to analyze!

A very similar approach to create the project as I am planning to do it!

The idea was to construct a neural network model to classify the signs, from the images captured by a camera of phone in real time. The user can

communicate with another person using sign to text. We can add after text to speech.

Dataset

The biggest problem in deep learning when you want to train a good model for prediction task, is to find a large dataset.

After searching in the internet, we did not find many datasets. The available one for free is for the alphabet A-Z and numbers 0–9. But the sign language is very rich, it contains a lot of representation.

Example from the free datasets:

- <https://www.kaggle.com/datamunge/sign-language-mnist> Data sets for future reference!
- <https://www.kaggle.com/ardamavi/sign-language-digits-dataset>



Figure 2: Dataset hand gesture (<https://www.kaggle.com/datamunge/sign-language-mnist>)

The dataset must contain European signs, because the project targets European users.

For an example, we decided to launch a demo to see what we could do afterwards. We created a simple model with ConvNets (convolutional neural networks) architecture to classify the static images. We obtained a good accuracy, but after when we wanted to add more cases to our model, we faced several constraints for the realization of this project.

The demo created is similar to the following video.

Although hidden here, this video shows an actual project that implemented the recognition of hand gestures by using convolutional neural networks and other parameters such as angular distance between fingers, analysis of gaps, and other potential complications in hand gestures. This project has helped me visualize some of the concepts when creating a hand gestures project as well as understand what the thought process behind the project would really look like.

Video 1: Simple hand recognition

The EgoGesture dataset

After a deeper research, we found the EgoGesture dataset, it's the most complete, it contains 2,081 RGB-D videos, 24,161 gesture samples and 2,953,224 frames from 50 distinct subjects. It design 83 classes of static or dynamic gestures focused on interaction with wearable devices. It's available only for research projects.

Could be very beneficial to my own research project



Figure 3: Some examples of the 6 scenes from EgoGesture
(<http://www.nlpr.ia.ac.cn/iva/yfzhang/datasets/egogesture.html>)

Link: <http://www.nlpr.ia.ac.cn/iva/yfzhang/datasets/egogesture.html>

NVIDIA Dynamic Hand Gesture dataset

There is an other dataset available also exclusively for research projects, created by NVidia:

Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural... Automatic detection and classification of dynamic hand gestures in real-world systems intended for human computer... research.nvidia.com	
--	--

Their gesture recognition system achieves an accuracy of 83.8%, it competing state-of-the-art algorithms, and approaching human accuracy of 88.4%.

Technical challenges

Here is the list of constraints for the realization of the project:

- Real-time recognition of dynamic hand gestures from video streams is a challenging task, because there is no indication when a gesture starts and ends in the video. For a demonstration of continuous data. See this video example about dynamic gesture



Figure 4: Example dynamic hand gesture (<https://www.youtube.com/watch?v=bO3TgU1s7hM&feature=youtu.be>)

- We need a dataset that contains dynamic gestures focused on interaction.
- There are signs where we have to use two hands and face expression. We need to synchronize them.

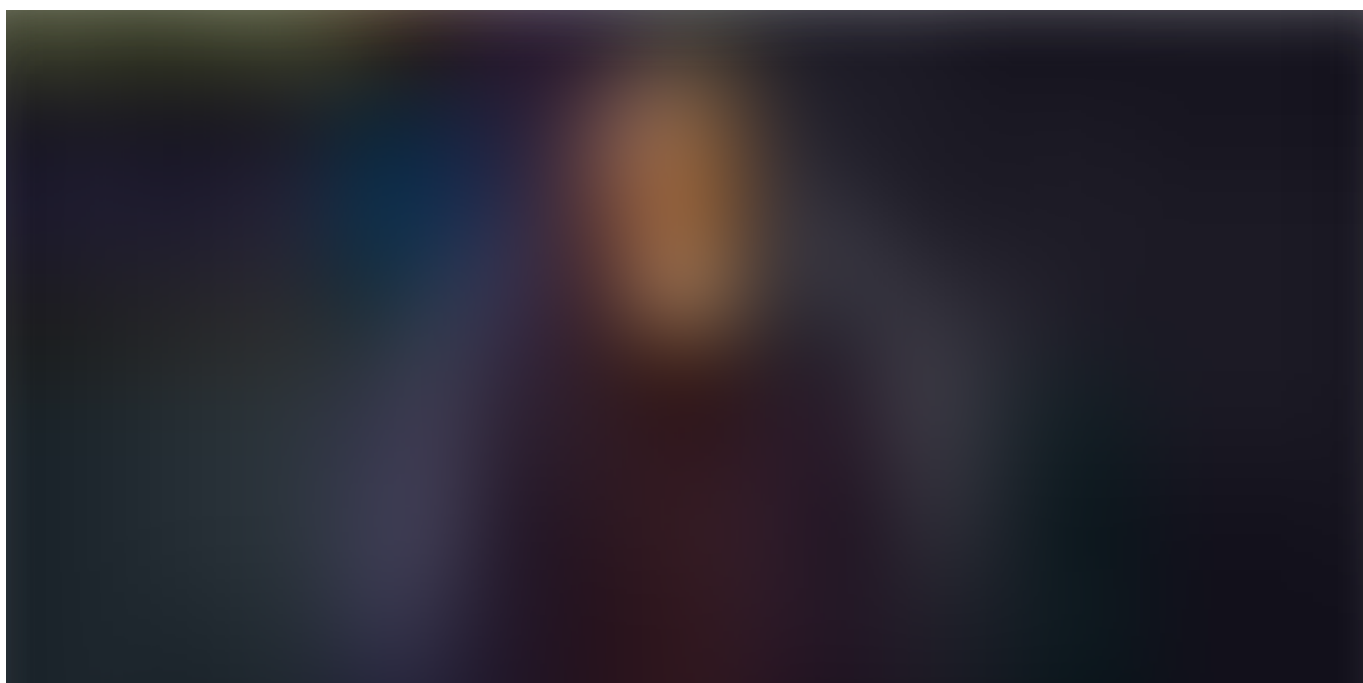




Figure 5: Using Hand and face expression for sign language(<https://www.youtube.com/watch?v=4cc3Stf3inQ>)

- We have a lot of signs that have the same gesture but different significations, for **example number 2, and “V”**.

Think about ways we can distinguish between very similar gestures



Figure 6: Example hand gesture with number 2 and v letter (<https://www.istockphoto.com/photo/american-sign-language-alphabet-gm136617721-18797954>)

- There are others signs that **depend on preceding signs**. So, for a complete recognition we have to make sense of the **position of signs in the phrase**.
- We need a 3D recognition of gestures.

- Developing an **efficient algorithm** for this task involves performance and optimization challenges.

The steps of realization

- At first, we have to detect the hand, it is a difficult task, but fortunately Google has published a **good 2D/3D hand detection model** in the mediapipe framework, this saves a lot of time:

<https://github.com/google/mediapipe>

We could use Google's hand detection model as inspiration for our own algorithm



Figure 7: Hand detection (<https://github.com/google/mediapipe>)

- Second, a classifier which is a deep CNN (resnet, inception, ...) to classify the detected gestures (segmented and continuous data).
- Facultative: maybe we need to add another architecture LSTM for remember also of previous geste.

Similar project

Kintrans (Professional)

It was launched in 2014, and it's always under development . It took several years to arrive to a mature version. It has already acquired funding from investors, and has Intel as sponsor. Microsoft has been an integral part of KinTrans' development, embedded at various steps along the way. From using the Microsoft Kinect 3D depth camera to providing hosting credits.

Link: <https://www.kintrans.com>

Real-time-GesRec (Research, open source)

This open source project for Real-time Hand Gesture Recognition use PyTorch on EgoGesture, NvGesture and Jester.

- Link github: <https://github.com/ahmetgunduz/Real-time-GesRec>
- Link research paper: <https://arxiv.org/abs/1901.10323>

Conclusion

This a very good and interesting project, with several technical challenges and open-research questions.

Reviewer

Okba Bekhelifi

Don't forget to give us your 🙌!



Thanks to Okba Bekhelifi.

Sign up for Latest from Becoming Human AI

By Becoming Human: Artificial Intelligence Magazine

Watch AI & Bot Conference for Free [Take a look.](#)

Get this newsletter

Artificial Intelligence AI Deep Learning Machine Learning Machine Intelligence

Learn more.

Medium is an open platform where 170 million readers come to find insightful and dynamic thinking. Here, expert and undiscovered voices alike dive into the heart of any topic and bring new ideas to the surface. [Learn more](#)

Make Medium yours.

Follow the writers, publications, and topics that matter to you, and you'll see them on your homepage and in your inbox. [Explore](#)

Write a story on Medium.

If you have a story to tell, knowledge to share, or a perspective to offer — welcome home. It's easy and free to post your thinking on any topic. [Start a blog](#)