GEORGETOWN UNIVERSITY

Real-Time Sign Language Detection

Group 1

Brendan Baker, Justin Ceresa,
Klaas van Kempen, Matt Moriarty

Introduction

Sign Languages

- Visual communication using fingers, hands, arm, head, body and facial expressions
- Often used by speech and hearing-impaired individuals
- More than **137** documented sign languages
- Automated sign language translation is a complex problem with many applications

Deep Learning and Sign Language

Sign Language Recognition: A Deep Survey
Razieh Rastgoo [a], Kourosh Kiani [a,*], Sergio Escalera [b]
[a] Electrical and Computer Engineering Department, Semnan University, Semnan, 3513119111, Iran
[b] Department of Mathematics and Informatics, Universitat de Barcelona and Computer Vision Center, 585, 08007 Barcelona, Spain

Expert Systems With Applications 164 (2021) 113794

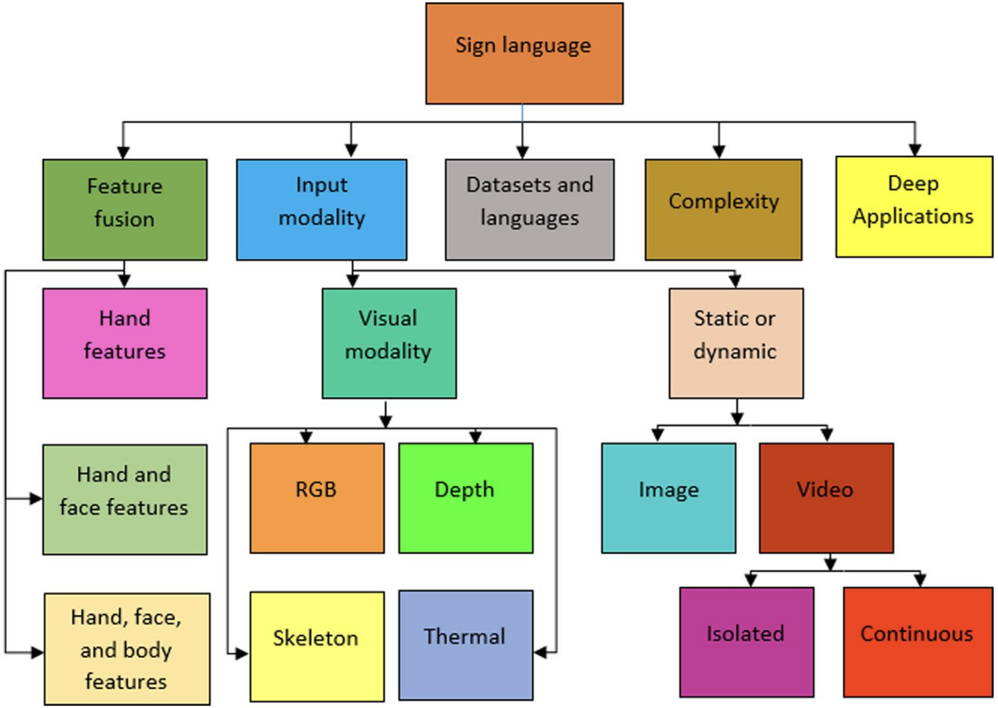


Fig. 1. Taxonomy of deep models for sign language recognition.

Applications:
- Translation
- Interpreting services
- HCI
- Virtual Reality

Deep Learning and Sign Language

Sign Language Recognition: A Deep Survey
Razieh Rastgoo [a], Kourosh Kiani [a,*], Sergio Escalera [b]
[a] Electrical and Computer Engineering Department, Semnan University, Semnan, 3513119111, Iran
[b] Department of Mathematics and Informatics, Universitat de Barcelona and Computer Vision Center, 585, 08007 Barcelona, Spain

 Expert Systems With Applications 164 (2021) 113794

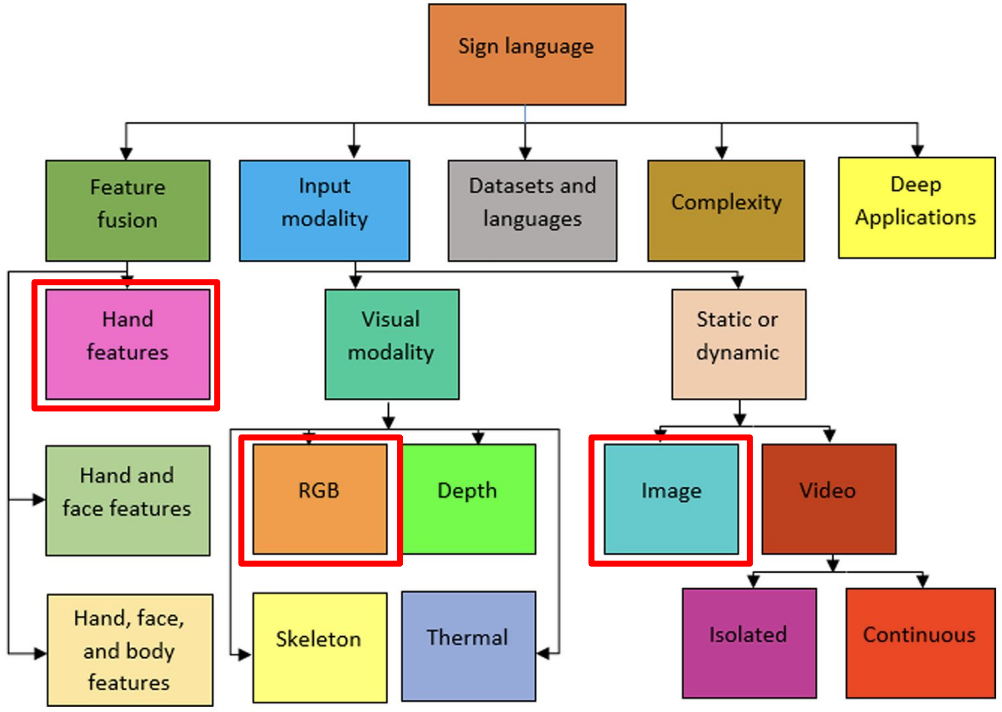


Fig. 1. Taxonomy of deep models for sign language recognition.

Applications:
- Translation
- Interpreting services
- HCI
- Virtual Reality

Datasets, Models, Goals

Table 4 (*continued*).

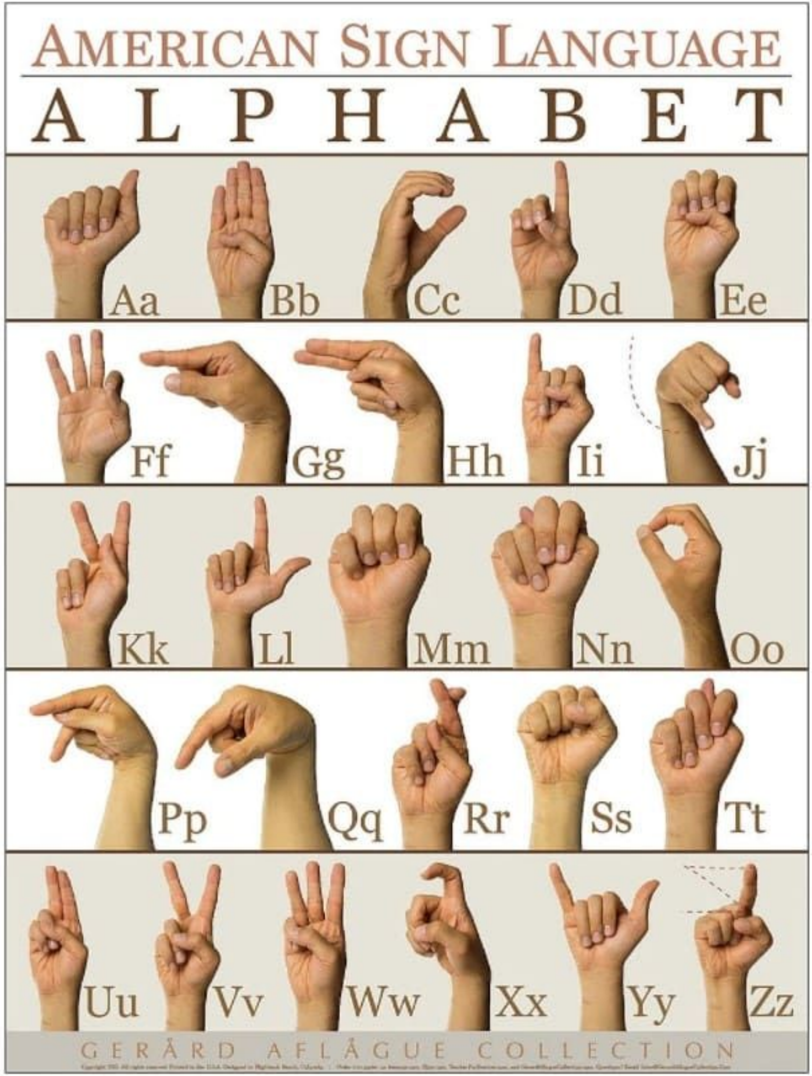
Dataset	Year	Ref.	Goal	Model	Modality	Results
Real video samples	2015	(Tagliasacchi et al., 2015)	HT	CNN	2D, Depth	8 mm (MSE)
	2019	(Ferreira et al., 2019)	HS	CNN	static, RGB, Depth	3.17, 92.61
RWTH-PHOENIX-Weather	2015	(Koller, Ney et al., 2015)	HSR	CNN	2D, RGB	55.70 (Precision)
BigHand2.2M	2017	(Yuan et al., 2017)	HP	CNN	2D, Depth	17.1 (error)
	2018	(Baek et al., 2018)	HP	GAN	static, Depth	13.7 mm
Human3.6M	2018	(Wang et al., 2018)	HP	CNN	2D, Depth	62.8 mm
UBC3V	2018	(Marı n Jimeneza et al., 2018)	HP	CNN	3D, Depth	88.2 (AUC)
Massey 2012	2018	(Rastgoo et al., 2018)	HR	RBM	2D, RGB, Depth	99.31
SL Surrey	2018	(Rastgoo et al., 2018)	HR	RBM	2D, RGB, Depth	97.56
ASL Fingerspelling A	2018	(Rastgoo et al., 2018)	HR	RBM	2D, RGB, Depth	98.13
OUHANDS,	2018	(Dadashzadeh, Tavakoli Targhi, & Tahmasbi, 2018)	HG	CNN	2D, Depth	86.46
Egohands	2017	(Dibia, 2017)	HT	CNN	static, RGB	96.86 (mAP)
Dexter	2018	(Mueller et al., 2018)	HT	CNN	static, RGB	64.0 (AUC)
EgoDexter	2018	(Mueller et al., 2018)	HT	CNN	static, RGB	54.0 (AUC)
RHD	2018	(Spurr et al., 2018)	HP	VAE	static, RGB, Depth	84.9(AUC)
B2RGB-SH	2019	(Li et al., 2019)	HP	CNN	static, RGB	7.18 (err)
DHG-14/28 Dataset	2019	(Chen, Zhao, Peng, Yuan, & Metaxas, 2019)	HG	CNN	dynamic, RGB	91.9
SHREC'17 Track Dataset	2019	(Chen et al., 2019)	HG	CNN	dynamic, RGB	94.4
RWTH-BOSTON-50	2019	(Lim et al., 2019)	HS	CNN	dynamic, RGB	89.33
ASLLVD	2019	(Lim et al., 2019)	HS	CNN	dynamic, RGB	31.50
EgoGesture	2019	(Kopuklu, Gunduz, Kose, & Rigoll, 2019)	HG	CNN	dynamic, RGB	94.03
NVIDIA benchmarks	2019	(Kopuklu et al., 2019)	HG	CNN	dynamic, RGB	83.83
isoGD	2020	(Elboushaki, Hannane, Afdel, & Koutti, 2020)	HG	CNN	dynamic, RGB, Depth	72.53
SKIG	2020	(Elboushaki et al., 2020)	HG	CNN	dynamic, RGB, Depth	99.72
NATOPS	2020	(Elboushaki et al., 2020)	HG	CNN	dynamic, RGB, Depth	95.87
SBU	2020	(Elboushaki et al., 2020)	HG	CNN	dynamic, RGB, Depth	97.51
RKS-PERSIANSIGN	2020	(Rastgoo et al., 2020a)	HSR	SSD, 2DCNN, 3DCNN, LSTM	dynamic, RGB	99.80

HT: Hand Tracking, HD: Hand Detection, HSR: Hand Sign Recognition, HG: Hand Gesture,
RHR: Real-time Hand Recognition.

Project

Current Project

- Real-time Detection/Translation of characters from American Sign Language (ASL)
- Known as “fingerspelling” or dactylology
- Goals:
  - Isolate hand
  - Classify characters
  - Deploy application



Data

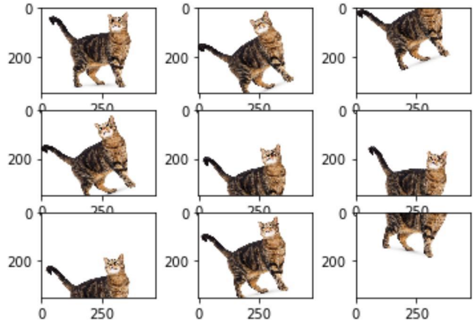
Kaggle Data

- 34,627 images
- 28x28 Grayscale Images
- No J and Z signs (motion required)
- ~1400 images of each sign

Methods

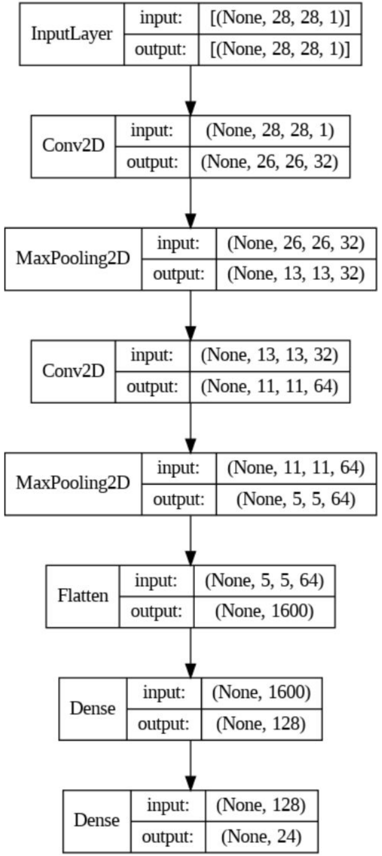
Data Preparation

- Image Augmentation
  - Zooming
  - Rotation
  - Shifting
- Image Normalization
- Reshaping



Model Architecture and Training

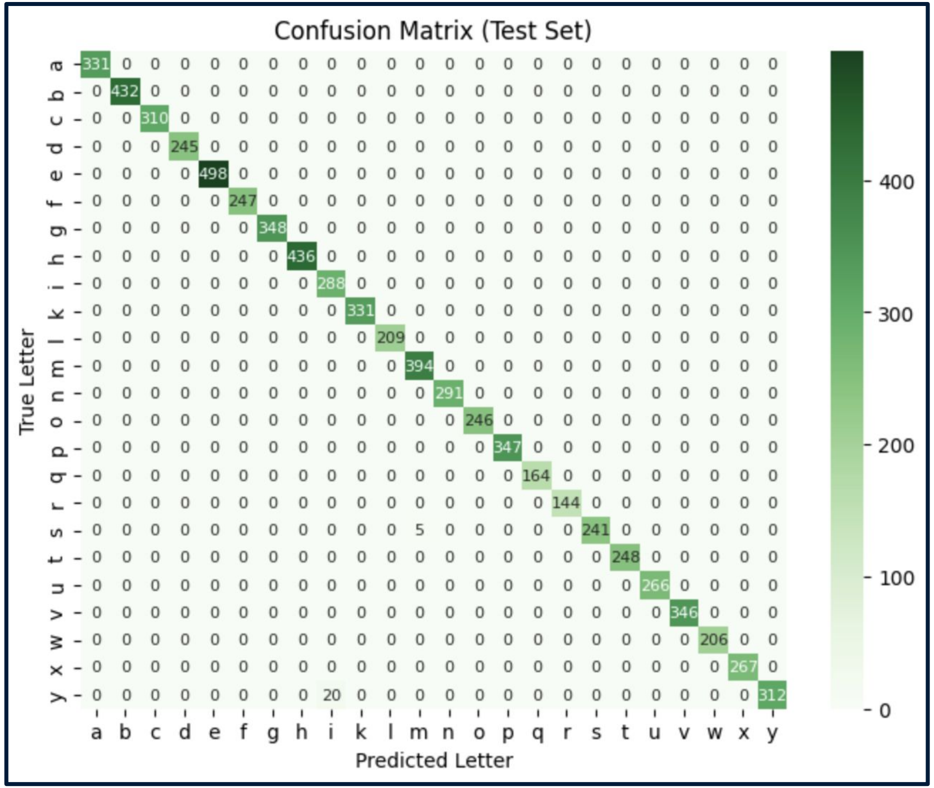
- CNN architecture
- Optimizer: **Adam**
- Loss Function: **Categorical Cross Entropy**
- Trained for 20 epochs (w/ early stopping)



Results

Results

	Accuracy
Training	99.9%
Testing	99.7%



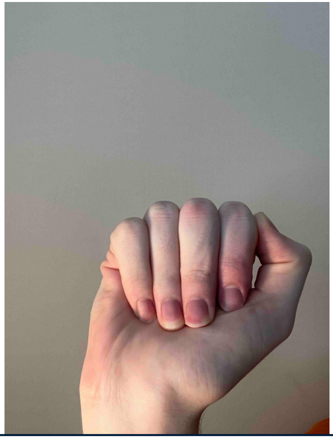
Live Image Processing

Data Preparation

- 📸 Take an image using a camera (not pre-prepared image)
- 🚀 Use pre-trained YOLOv5 model for object detection
- 🌾 Crop image to detected bounding box
- ☑️ Make image greyscale and downsample for expected input size
- 💻 Feed into sign language CNN just like pre-developed images

Data Preparation

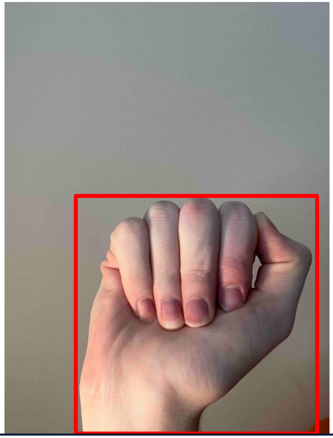
Original Image



(3000, 2000, 3)

YOLO →

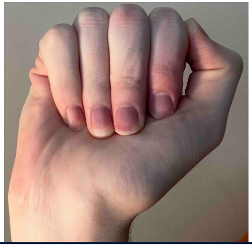
Object Detection



(3000, 2000, 3)

Crop →

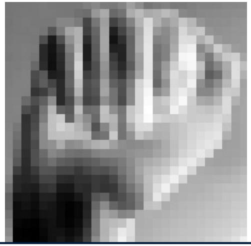
Cropped Image



(1600, 1500, 3)

Reduce →

Reduced Image



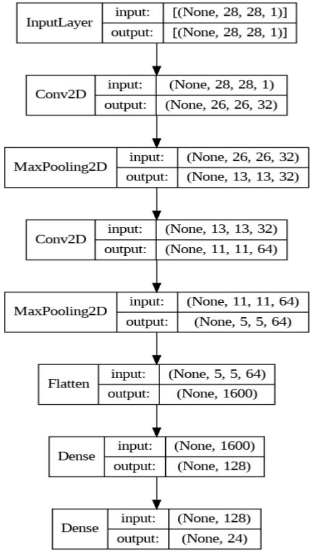
(28, 28, 1)

Model Evaluation

Test Image

Model Input →

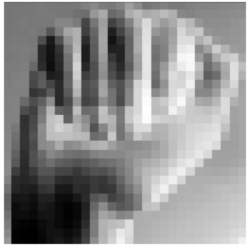
Sign Language CNN



Prediction →

‘D’

Reduced Image

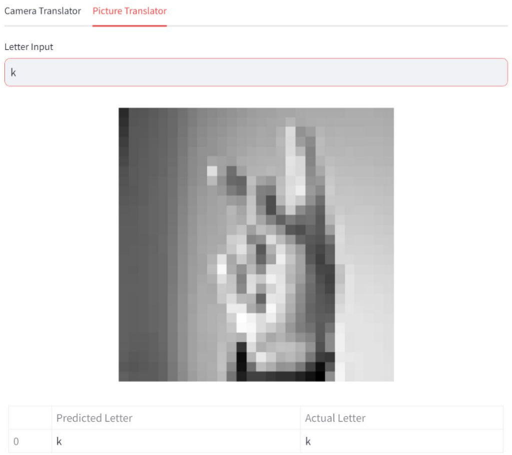


Model Input →

Prediction →

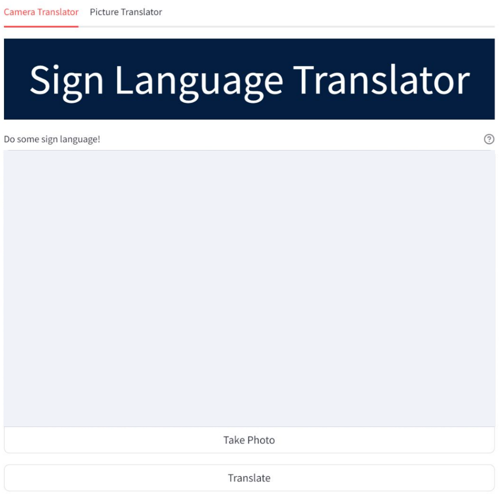
‘A’

Streamlit Application Deployment



Streamlit Application

- User-friendly way to interact with our sign language detection pipeline



Demo!

Future Work

- Incorporate motion
- Use a model that masks hands
- Allow for continuous translation

Concluding Remarks

- Processing live image data presents its own challenges
- Building a pipeline requires both architectural and accessibility considerations
- Accessibility is a priority!

Thank You!

References

- Bilgin, M., & Mutludoğan, K. (2019, October). American sign language character recognition with capsule networks. In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-6). IEEE.
- Cihan Camgoz, N., Hadfield, S., Koller, O., & Bowden, R. (2017). Subunets: End-to-end hand shape and continuous sign language recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 3056-3065).
- Elboushaki, A., Hannane, R., Afdel, K., & Koutti, L. (2020). MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. *Expert Systems with Applications*, *139*, 112829.
- Fenlon, J., & Wilkinson, E. (2015). Sign languages in the world. *Sociolinguistics and Deaf communities*, *1*, 5.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
- Morford, J. P., & MacFarlane, J. (2003). Frequency characteristics of American sign language. *Sign Language Studies*, 213-225.
- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, *164*, 113794.