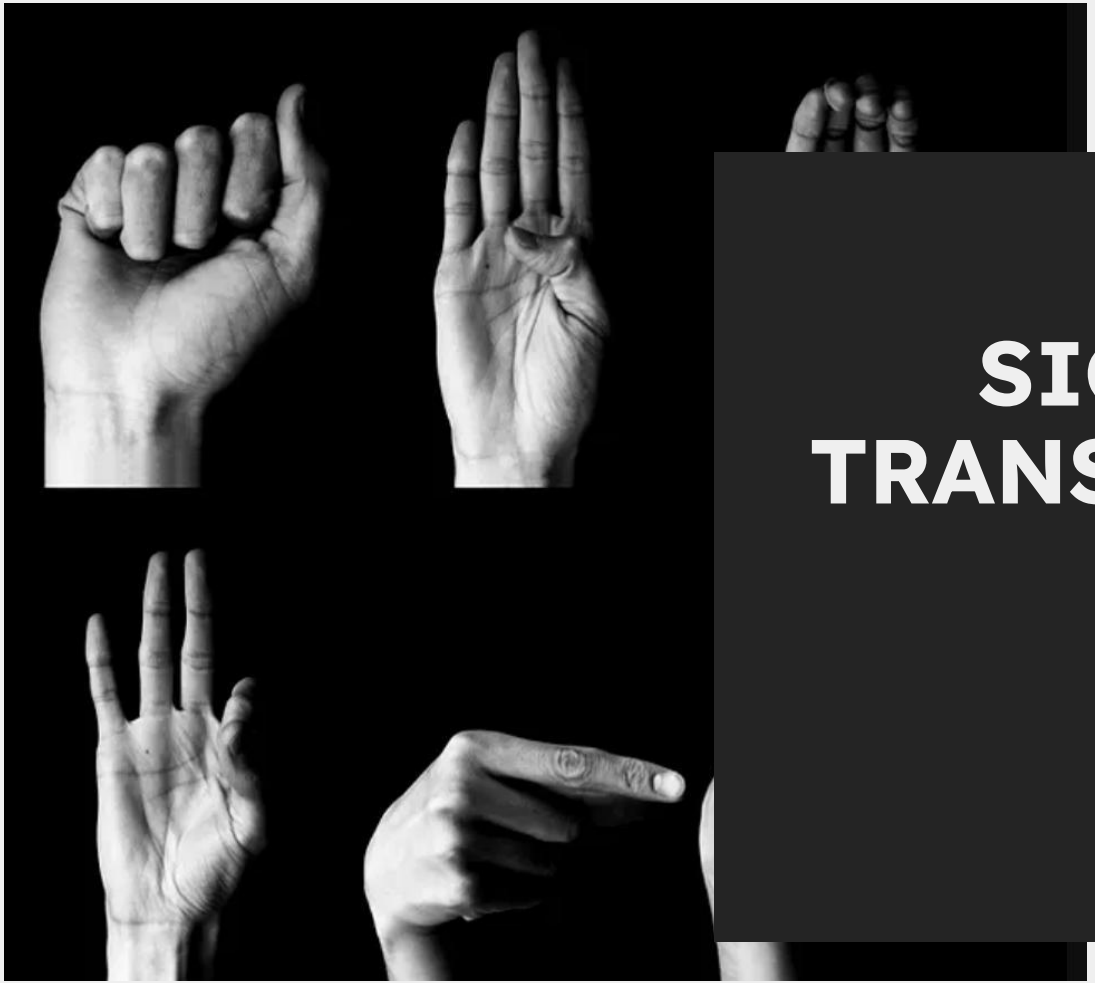


AUTOMATE SIGN LANGUAGE TRANSLATION INTO TEXT

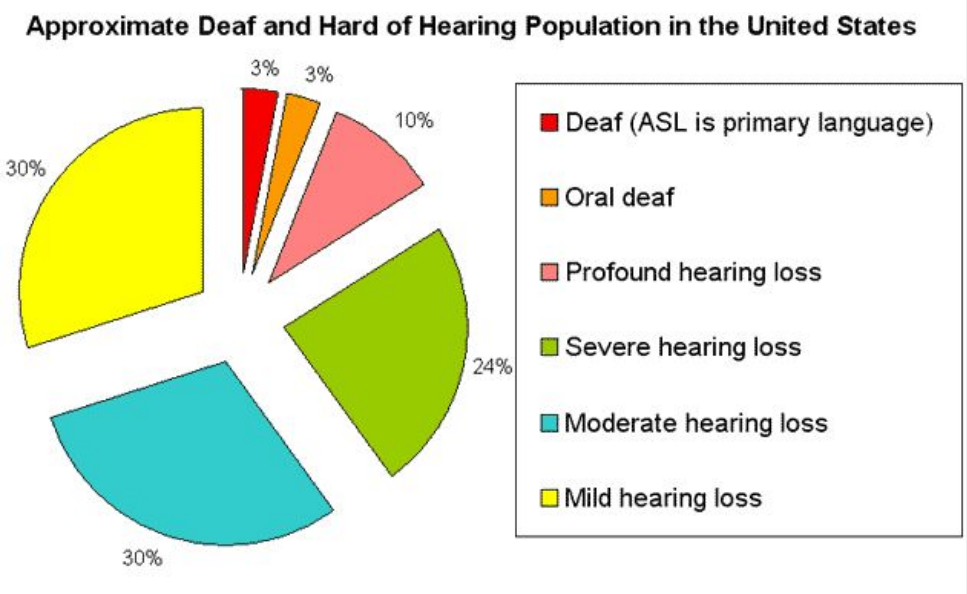
October, 2022
FourthBrain, Inc.

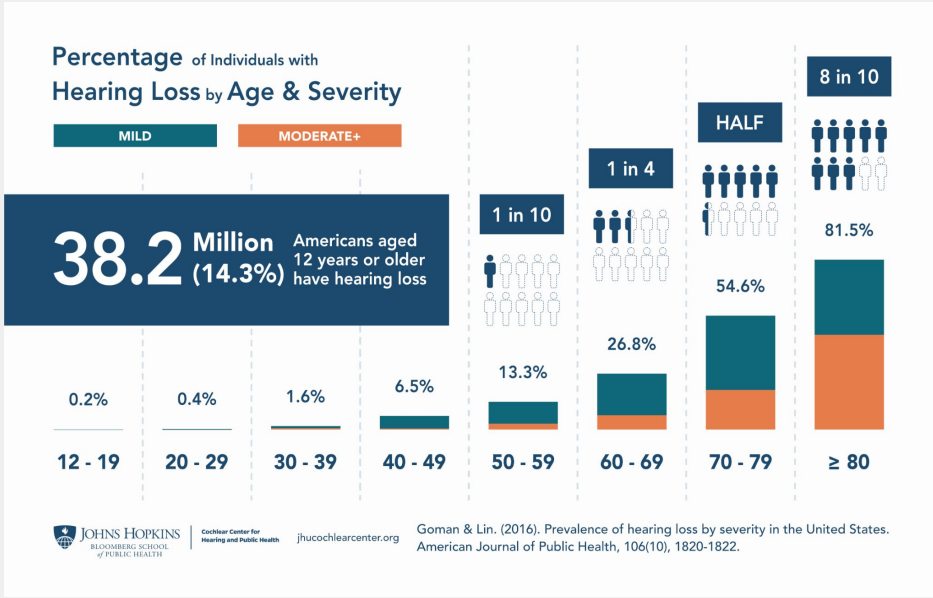
Sheilah K.

AGENDA

1. PROBLEM
2. SOLUTION
3. DATA + MODEL
4. DEMO
5. SYSTEM DESIGN
6. ETHICAL CONSIDERATIONS
7. FUTURE WORK
8. Q&A

OVERVIEW





Sources:

- World Health Organization
- US Census Bureau
- National Institute on Deafness and Other Communication Disorders

01

PROBLEM

BUSINESS VALUE: OPPORTUNITY

- **Large addressable market:**
  - 14 % of the US population aged 12 and above, have reported some form of hearing loss in both ears (NIDCD surveys)
  - Roughly 1 million are functionally deaf — or unable to hear normal conversation even when using a hearing aid

- **Inefficient existing solutions:**
  - Hindrance to visual communication– background noise, lighting, pace of conversation, number of speakers, accents, facial hair
  - Cost– hiring translators, auditory-enhancing equipment

- **Lack of autonomy due to socio-economic challenges:**
  - Unemployment or underemployment
  - Education gaps
  - Delayed healthcare access
  - Social isolation– communication barriers

02

SOLUTION

- Deep-learning based tool that can automatically translate American sign language gestures into English text
- “real -time” usage for a more accessible and inclusive world through AI –comes in handy during emergencies, doctor appointments, job interviews etc.
- **Solution 1: Real-time prediction using webcam**
- **Solution 2: Upload photo via FastAPI web application for prediction**
- Metrics : accuracy and quick inference

Data processing :

- 2, 520 RGB samples - hand gestures
- 70 samples for each of the 36 classes (A-Z, 0-9)
- open- sourced from kaggle
- in–place augmentation and scaling in batches

03

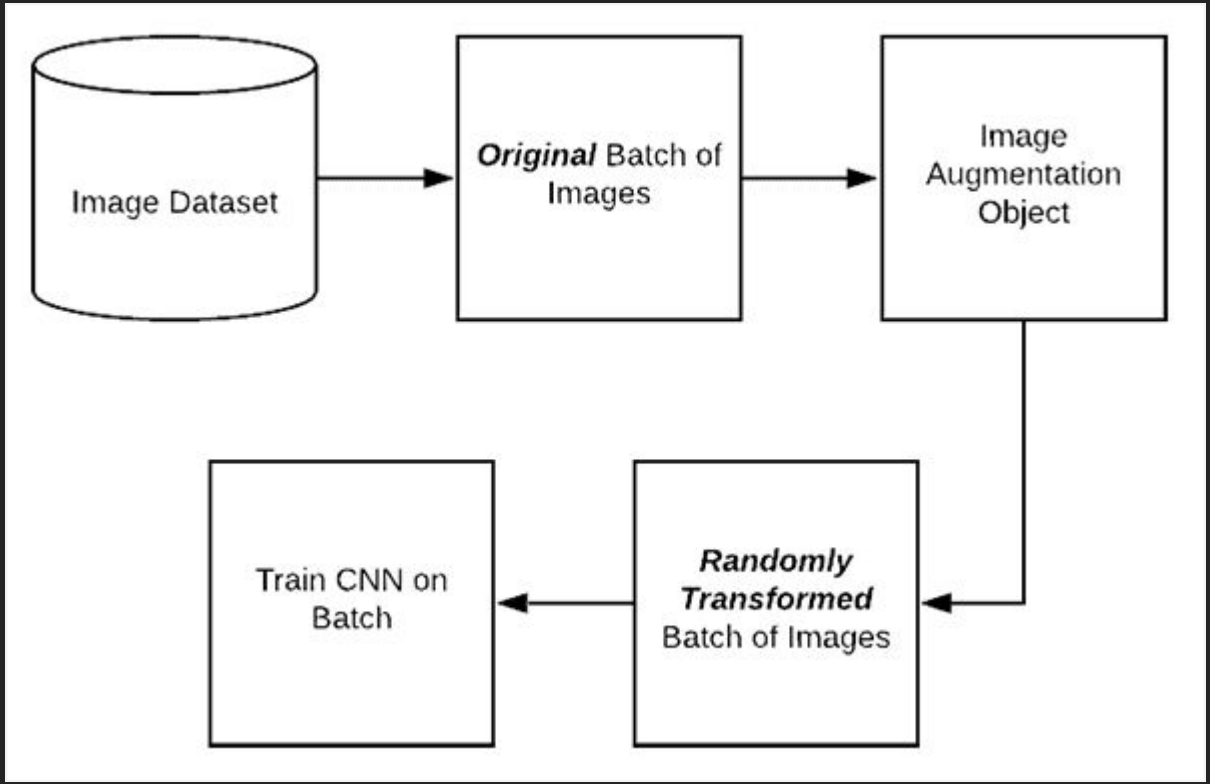
Model-training :

- Transfer-learning with pre-trained CNN models:
  - vgg16
  - Mobilenet

- Very little fine-tuning:
  - Frozen weights of most layers

DATA + MODEL

Data flow



Keras

ImageDataGenerator

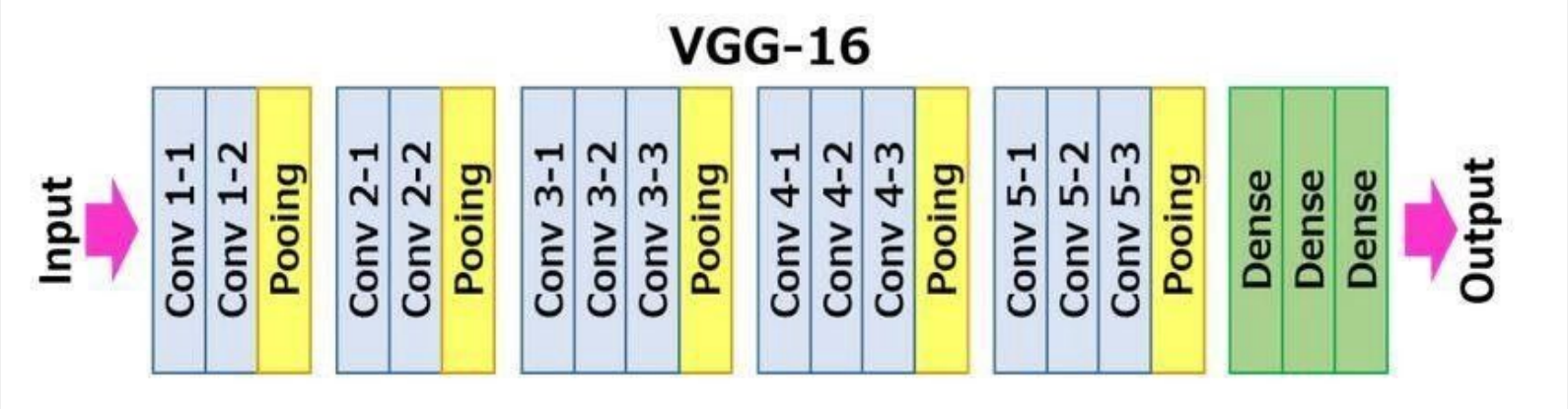
Preprocess_input function

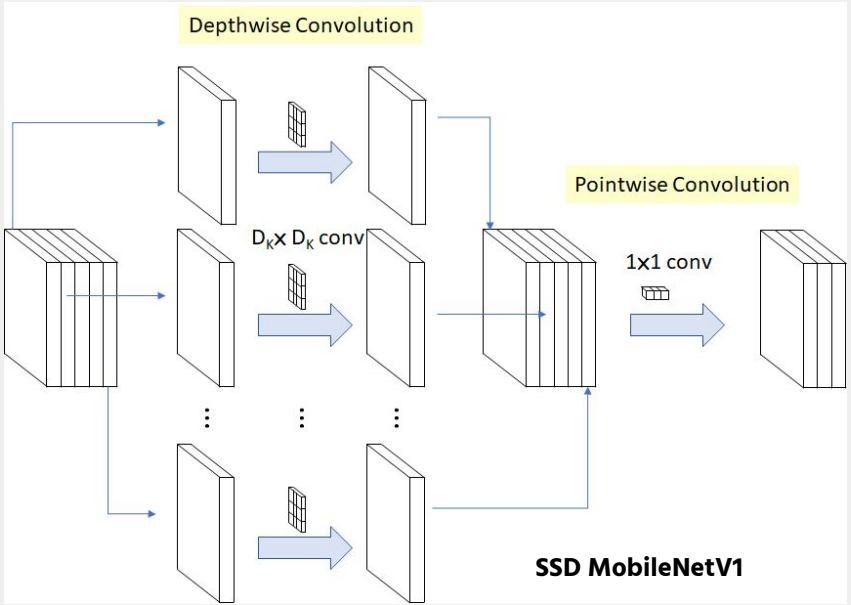
Normalization

In-place augmentation

Resize Input shape

Batch size



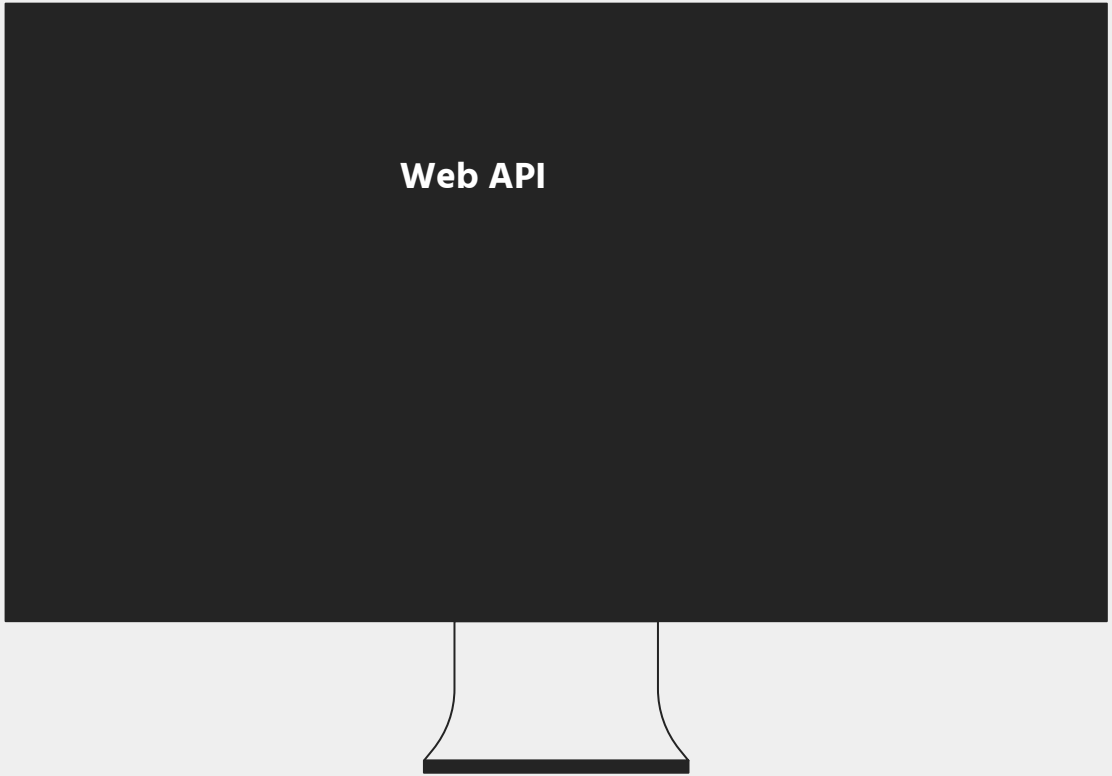


Model	Accuracy
MobileNetV1	0.88
VGG16	0.20

Conclusion: MobileNet is 32X smaller than VGG16 yet is 4X faster, which means it's efficient

LIVE DEMO

04

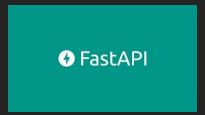


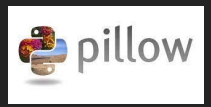
05

SYSTEM DESIGN

ML Stack

Data → Pre-processing → Machine Learning → Deployment





06

ETHICAL CONSIDERATIONS

- Racial bias: based on skin color and hand features, include images from diverse racial backgrounds
- Confirmation bias : data is well-structured and curated for image classification
- Inaccurate translations: compromised communication and understanding for deaf users
- Unknown origins: data might have violated ethical consent when collecting the images

07 FUTURE WORK

- Train on more models/ Extensive hyperparameter tuning to improve accuracy (VGG16)
- Can this model be used to translate gestures into complete sentences?
- Incorporate explainability component
- Deploy real-time system in AWS EC2 instance

QUESTIONS