

# Sign2Text



*Belal Chaudhary*

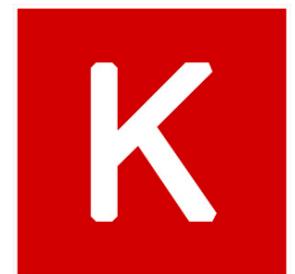
18/10/2017

# Real-time translation of Sign Language into text

## **Objectives:**

- Translate ASL from images into text ('Sign2Text')
- Focus on the ASL alphabet (A-Z)

# Pipeline



Pillow





**Iteration 1**

**ASL BU Corpus**  
~1500 images

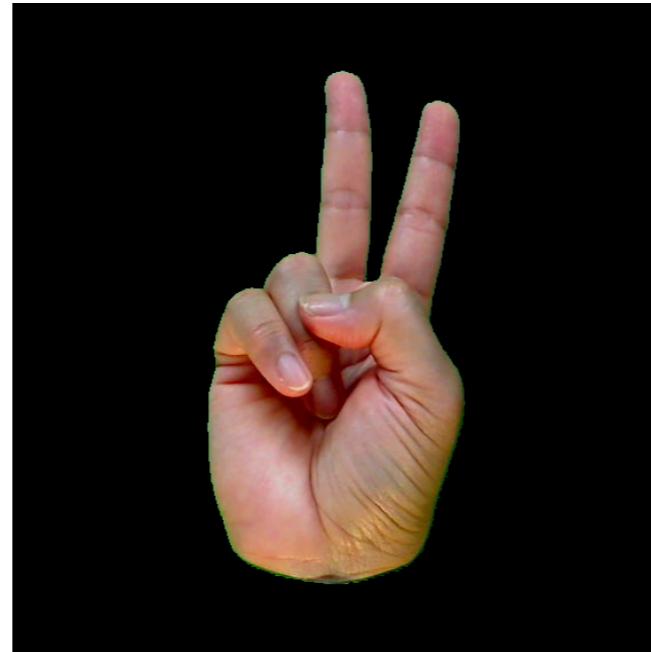




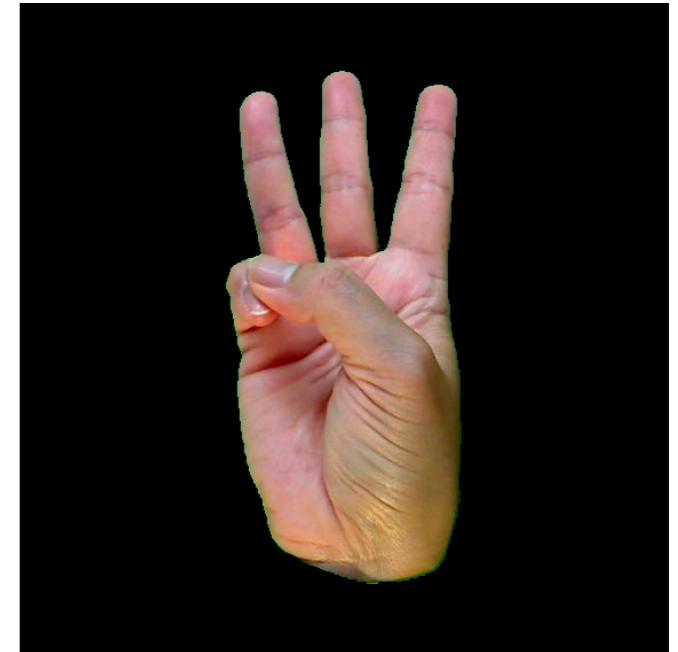
**U**



**V**



**W**





**A**



**M**



**N**



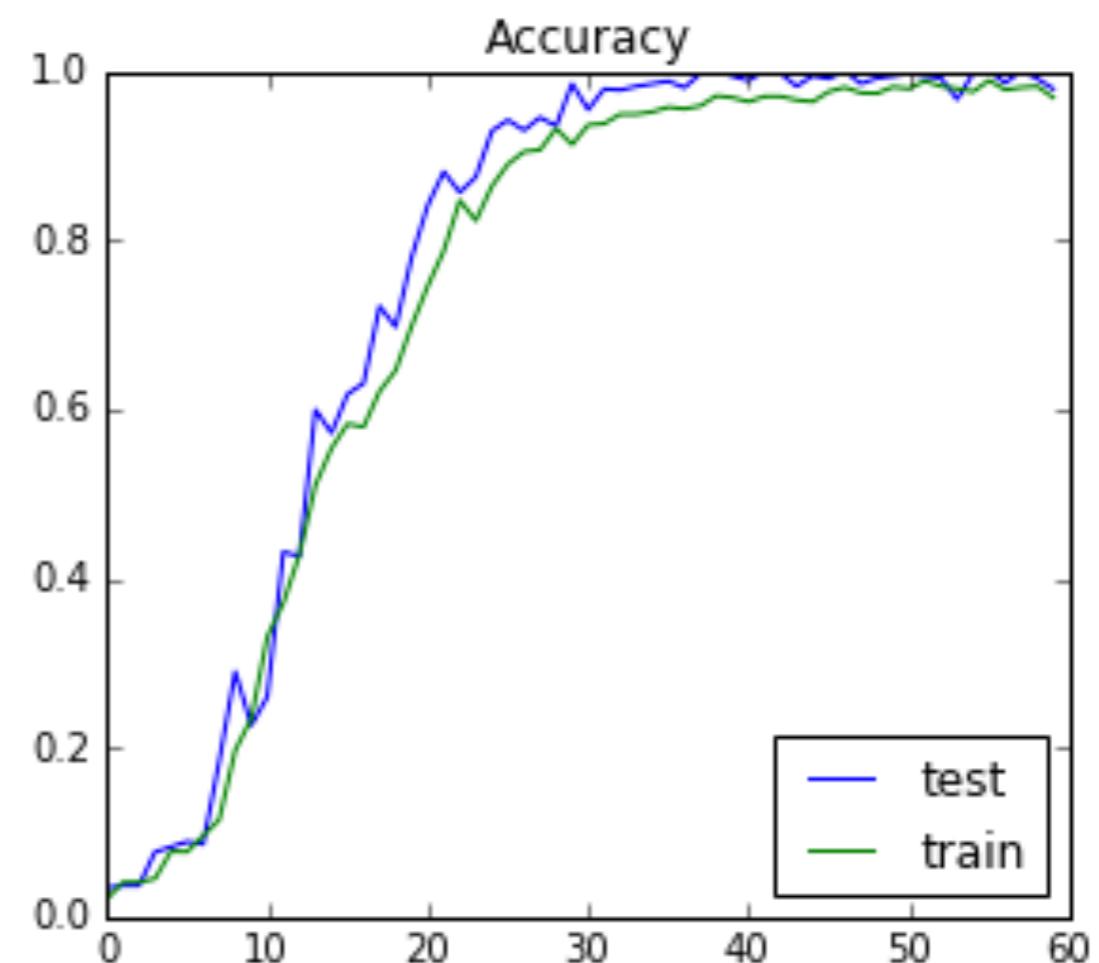
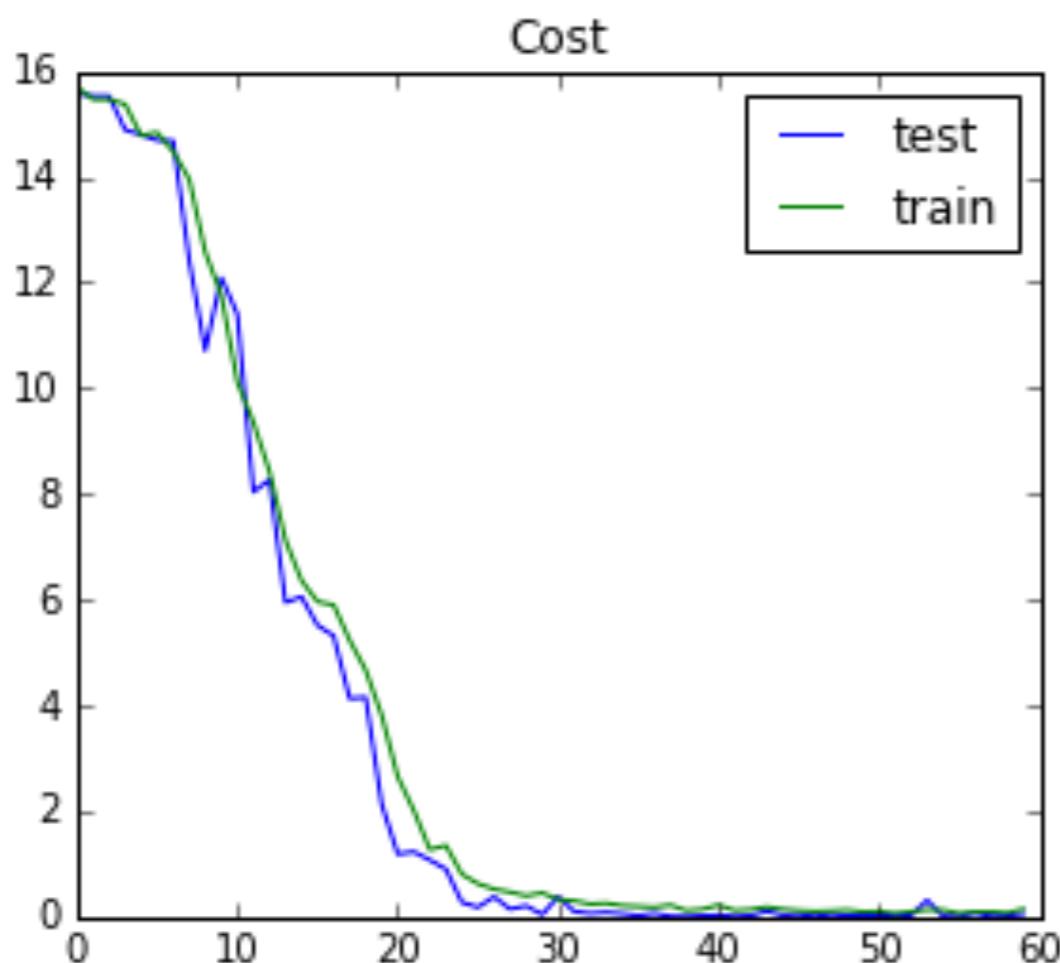


**Raw image**

**Padding**

**Re-size for NN**

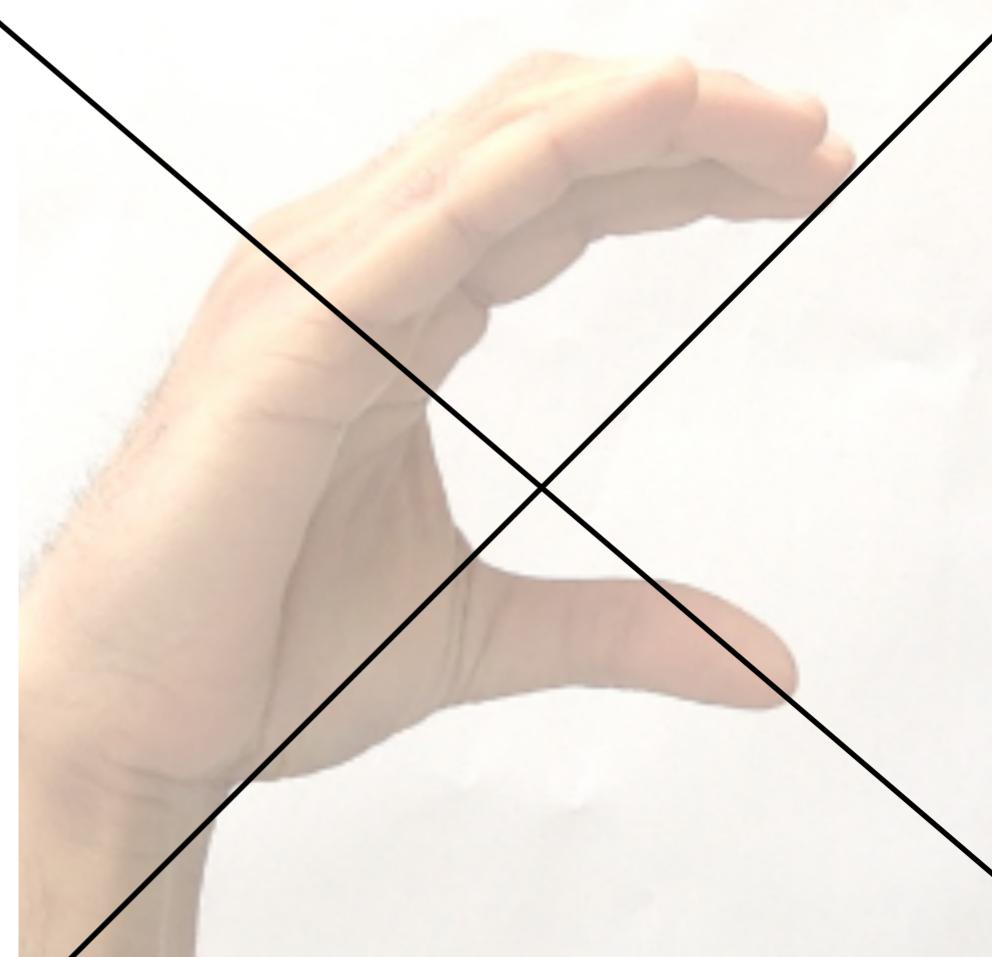
- ~1500 images
- ~55 / class





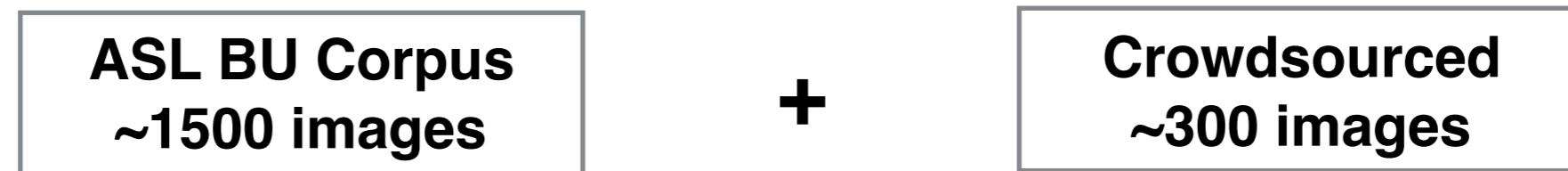
```
video_capture = cv2.VideoCapture(0)
```

```
while True:  
    # Capture frame-by-frame  
    ret, frame = video_capture.read()  
    # Do pre-processing  
    # ...  
    # Do classification  
    # ...  
    my_model.predict(frame, batch_size=1)  
    # Visualize the resulting frame  
    cv2.imshow('Video', frame)  
    # Press 'q' to exit live loop  
    if cv2.waitKey(10) & 0xFF == ord('q'):  
        break
```



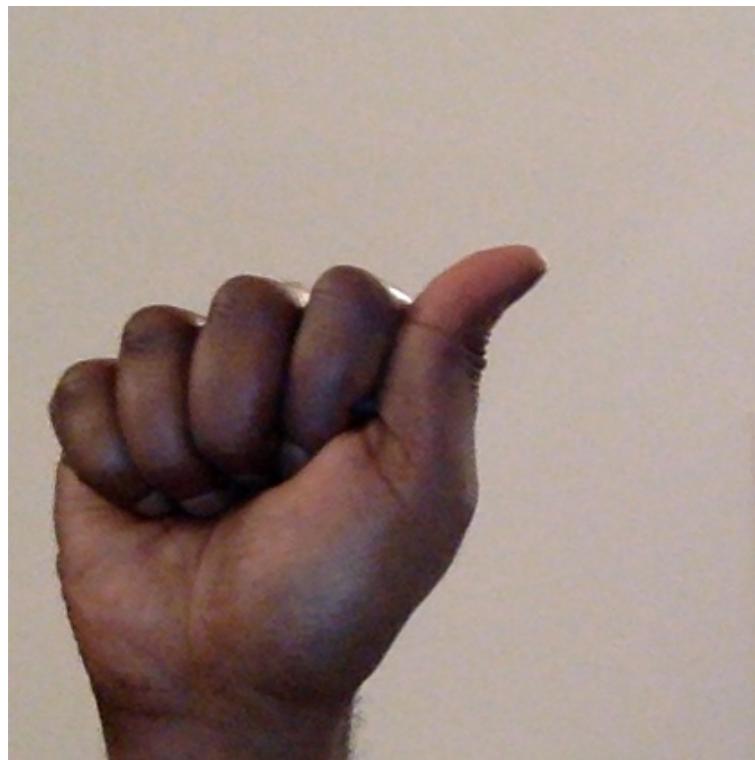


**Iteration 2**





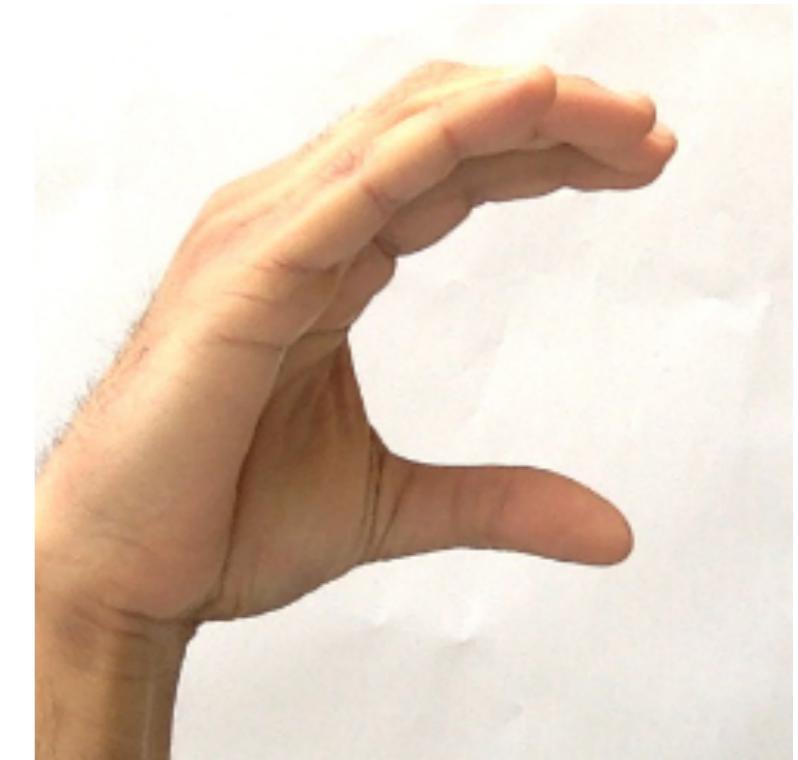
**A**



**B**



**C**





- ~2000 images
- ~75 / class



```
from keras.applications import VGG16, ResNet50, InceptionV3, Xception, MobileNet
```

| Model             | Size   | Top-1 Accuracy | Top-5 Accuracy | Parameters  | Depth |
|-------------------|--------|----------------|----------------|-------------|-------|
| Xception          | 88 MB  | 0.790          | 0.945          | 22,910,480  | 126   |
| VGG16             | 528 MB | 0.715          | 0.901          | 138,357,544 | 23    |
| VGG19             | 549 MB | 0.727          | 0.910          | 143,667,240 | 26    |
| ResNet50          | 99 MB  | 0.759          | 0.929          | 25,636,712  | 168   |
| InceptionV3       | 92 MB  | 0.788          | 0.944          | 23,851,784  | 159   |
| InceptionResNetV2 | 215 MB | 0.804          | 0.953          | 55,873,736  | 572   |
| MobileNet         | 17 MB  | 0.665          | 0.871          | 4,253,864   | 88    |

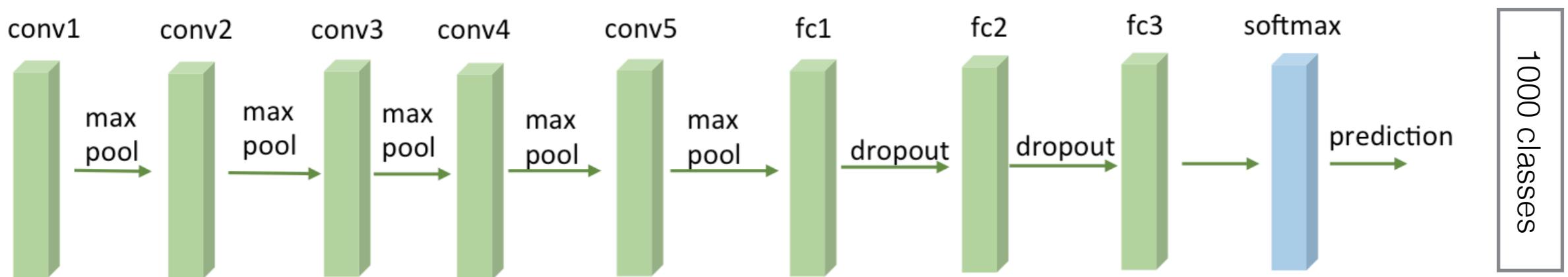
The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset



## Transfer learning

## VGG16 Architecture

each conv includes 3 convolutional layers

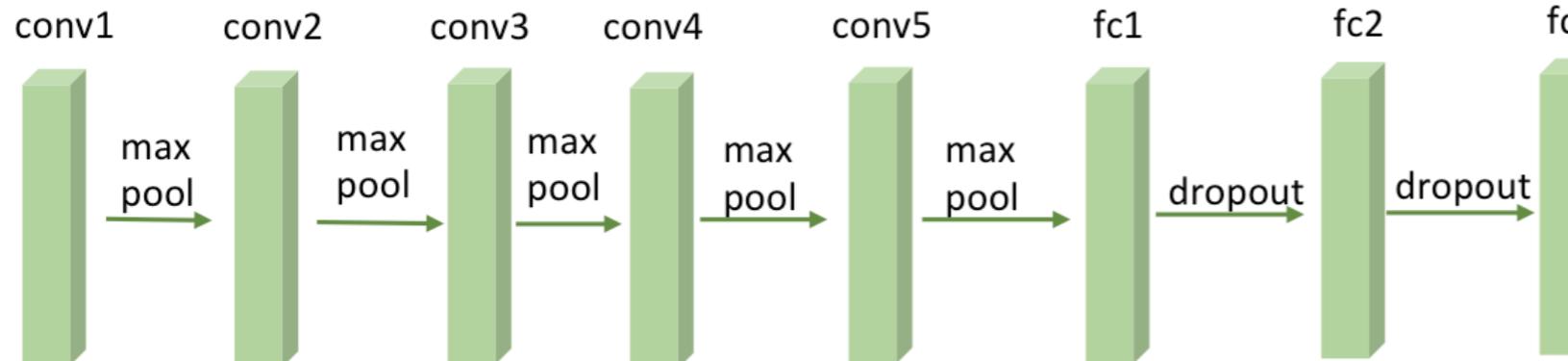




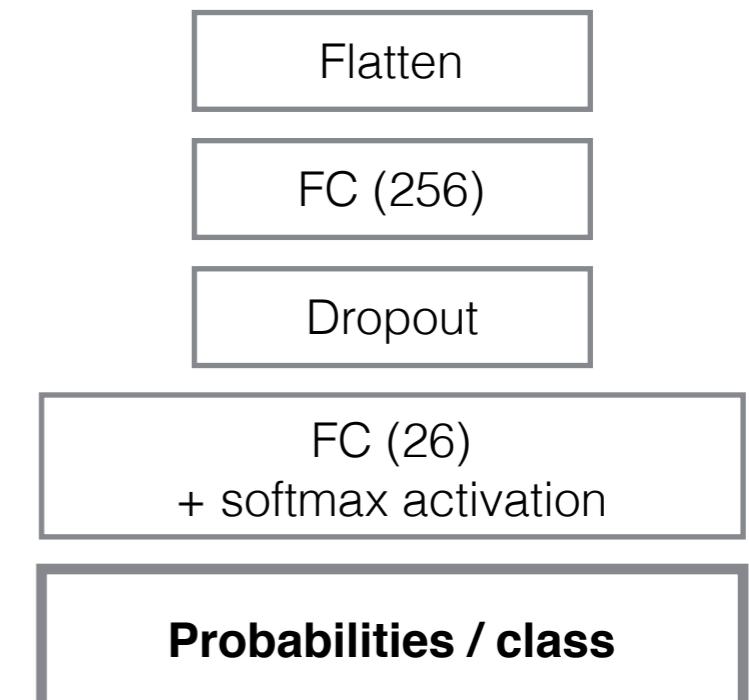
## Transfer learning

### VGG16 architecture

each conv includes 3 convolutional layers



***Classification block***

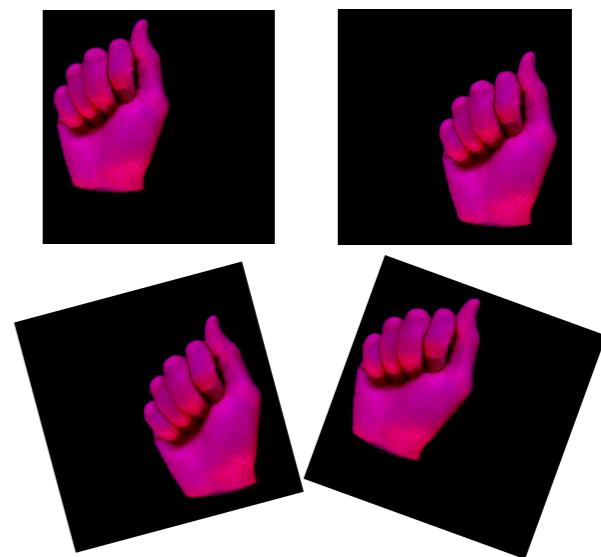




# DATA AUGMENTATION

---

→ applied every epoch





| Model       | Average<br>FPS | Classification Block |          | Random Forest |
|-------------|----------------|----------------------|----------|---------------|
|             |                | Accuracy             | Accuracy | Accuracy      |
| VGG16       | 1.69           | 0.997                |          | 0.853         |
| ResNet50    | 2.87           | 0.999                |          | 0.799         |
| InceptionV3 | -              | -                    |          | 0.818         |
| Xception    | -              | -                    |          | 0.767         |
| MobileNet   | 5.66           | 0.997                |          | 0.687         |

Testing done using: MacBook Pro  
 16 GB 1600 MHz DDR3  
 2.2 GHz Intel Core i7



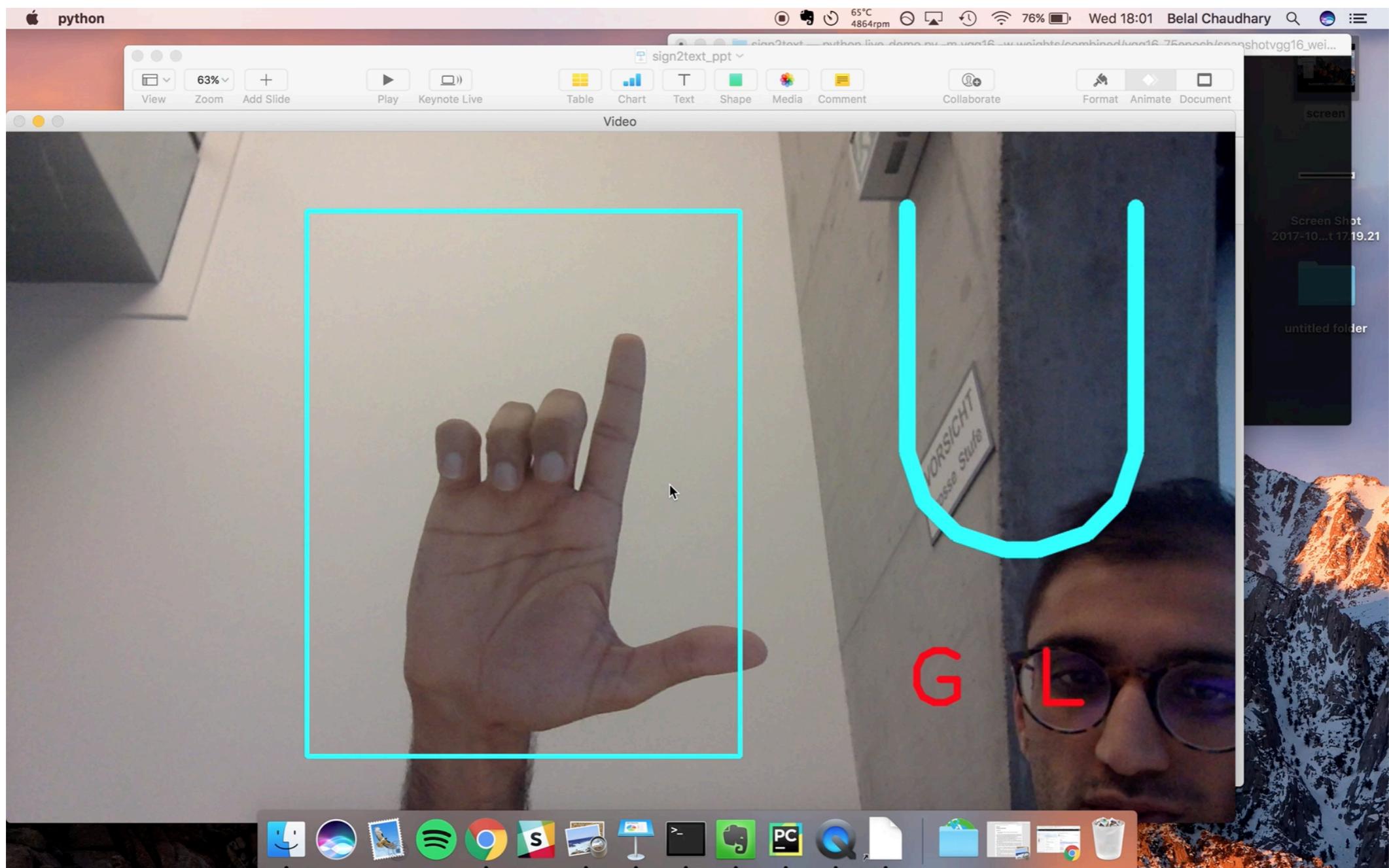
```
video_capture = cv2.VideoCapture(0)
```

```
while True:  
    # Capture frame-by-frame  
    ret, frame = video_capture.read()  
    # Do pre-processing  
    # ...  
    # Do classification  
    # ...  
    my_model.predict(frame, batch_size=1)  
    # Visualize the resulting frame  
    cv2.imshow('Video', frame)  
    # Press 'q' to exit live loop  
    if cv2.waitKey(10) & 0xFF == ord('q'):  
        break
```

**Public repo:** <https://github.com/BelalC/sign2text>



## VGG16 Demo - ~2 frames/ second



# Future work

- More data! (& augmentation)
- Word & sentence-level translation
- Other sign languages
- Hand tracking
- Port to mobile/embedded systems
- Speech2Text

# Thanks

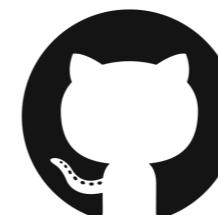
- Volunteer signers, interpreters, signing community
- Boston University

# Let's talk...

- Data science
- Deep learning
- Text, images, time-series
- Biotechnology



DATA SCIENCE RETREAT®  
SINCE 2014



/BelaIC



/belalchaudhary

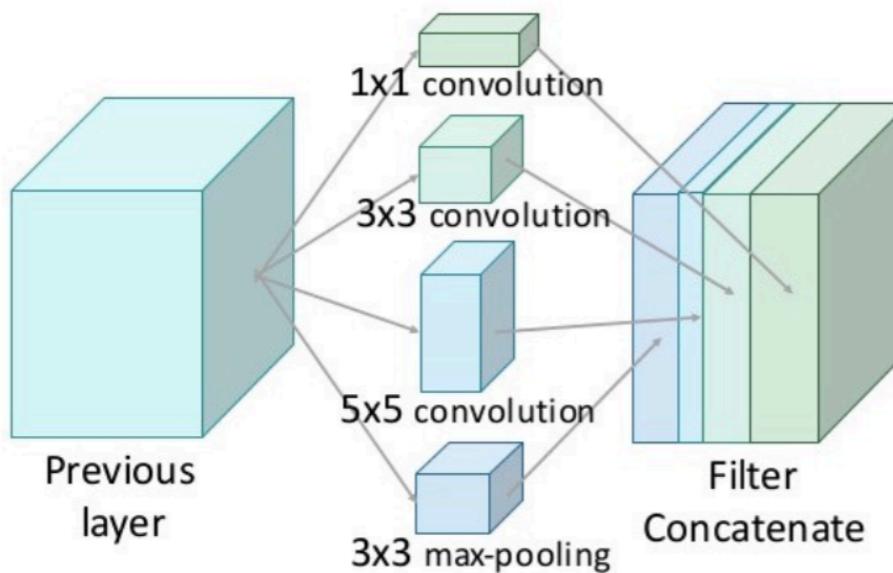
IIMHPLab

ENABLING YOU TO SHAPE THE FUTURE

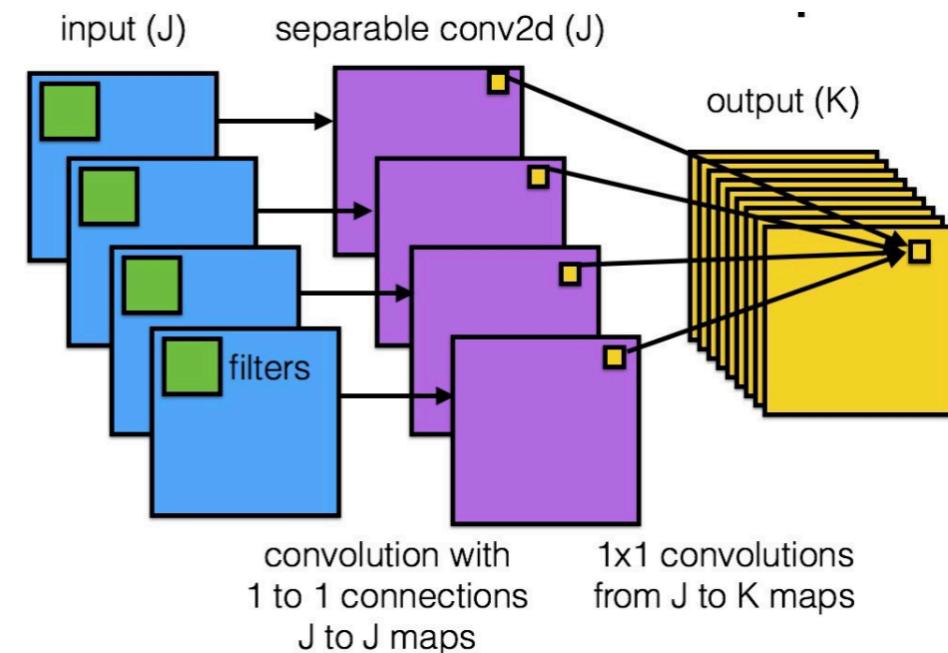
# Appendix

- **Transfer learning with Keras:**
  - <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>
- **Deep learning architectures:**
  - <https://medium.com/towards-data-science/an-intuitive-guide-to-deep-network-architectures-65fdc477db41>

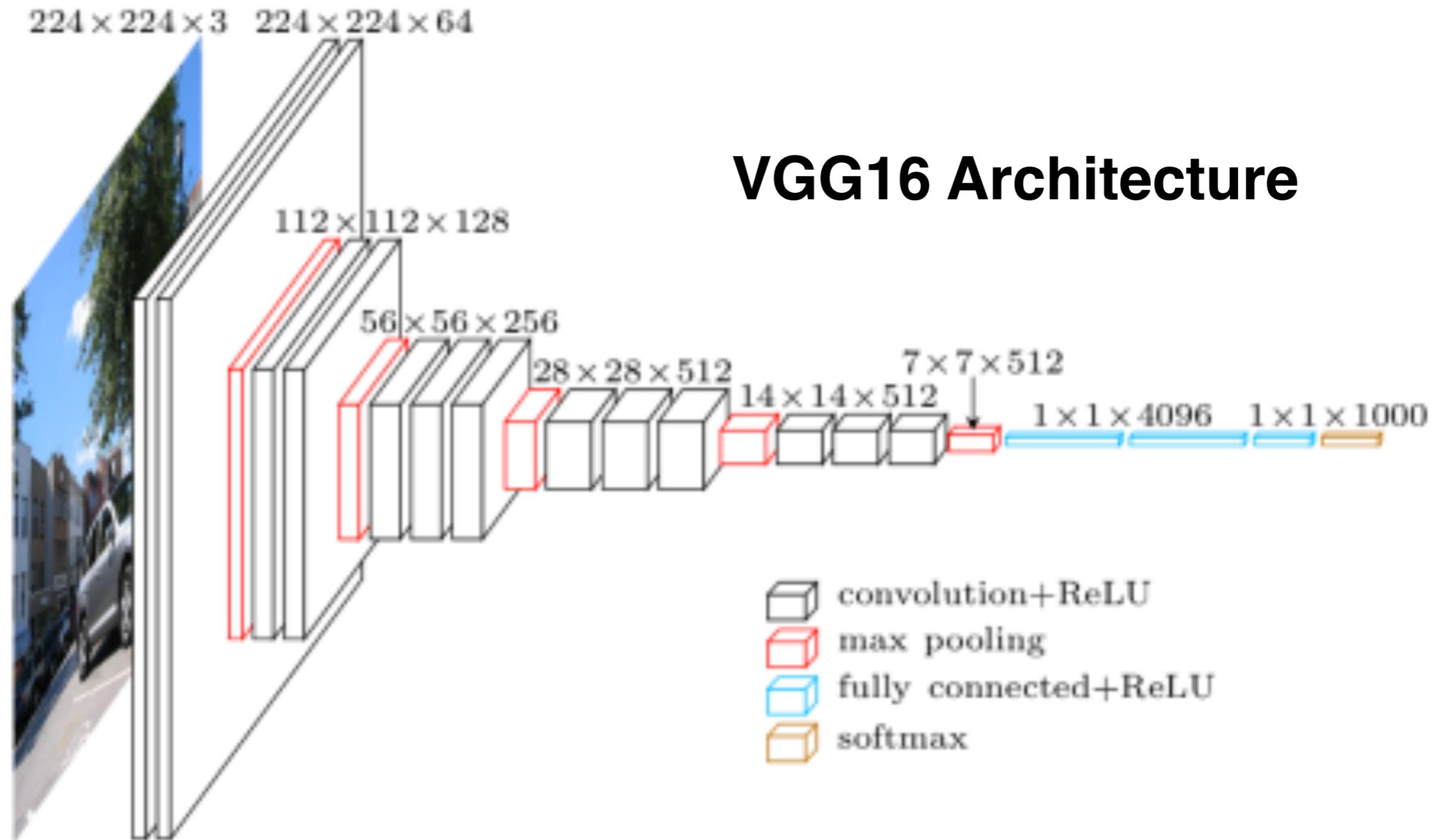
**'Naive' Inception Module**



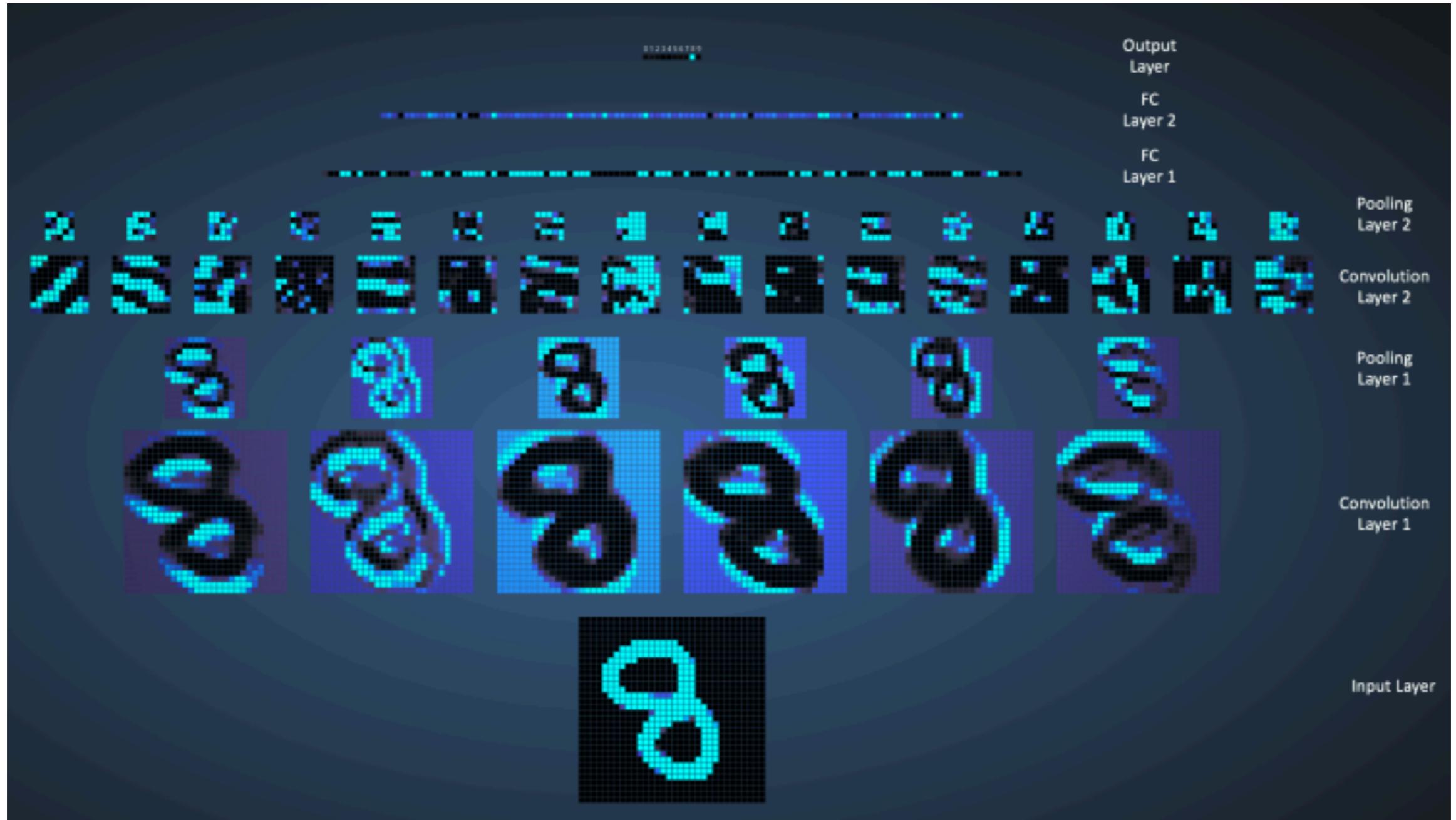
**Xception network**



# Appendix



# Appendix



## Convolutional neural networks, transfer learning & more

- <http://setosa.io/ev/image-kernels/>
- <http://cs231n.github.io/>
- <http://cs231n.github.io/transfer-learning/>