A. Architecture of the ConvNet Model Used

We use the architecture shown in Figure A.1 to train *Bob's* image classification model. The number of parameters to be learned for each layer is shown in Figure A.2. While training the model, each fully connected (FC) layer is separated by a dropout layer. We use Softmax as the activation for the last FC layer for training and use approximated Softmax for inference in-the-clear as well as private inference.

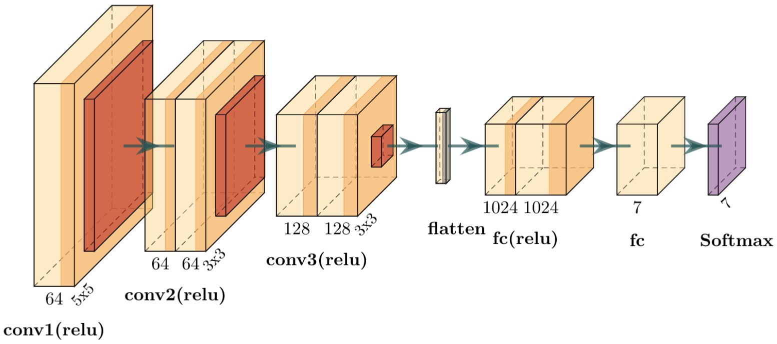
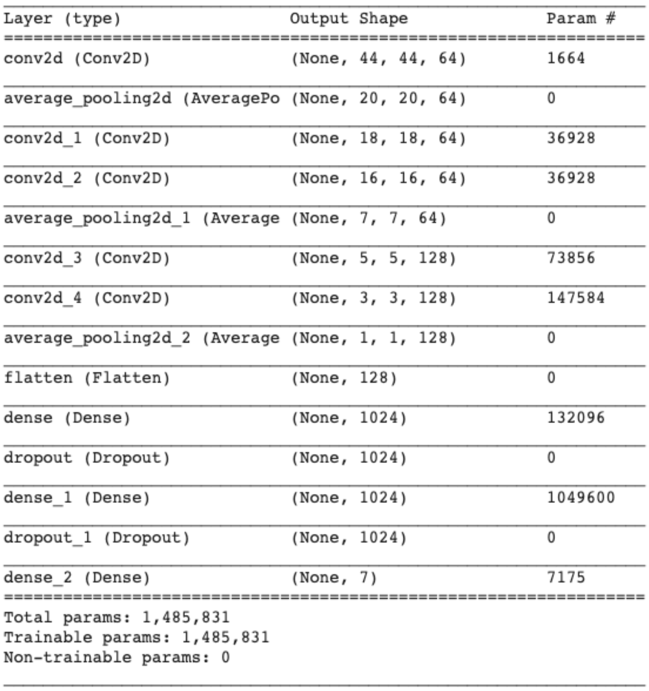


Figure A.1. Diagram of ConvNet architecture for frame classification



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 44, 44, 64)        1664
_________________________________________________________________
average_pooling2d (AveragePo (None, 20, 20, 64)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 18, 18, 64)        36928
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 16, 16, 64)        36928
_________________________________________________________________
average_pooling2d_1 (Average (None, 7, 7, 64)          0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 5, 5, 128)         73856
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 3, 3, 128)         147584
_________________________________________________________________
average_pooling2d_2 (Average (None, 1, 1, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 1024)              132096
_________________________________________________________________
dropout (Dropout)            (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              1049600
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 7)                 7175
=================================================================
Total params: 1,485,831
Trainable params: 1,485,831
Non-trainable params: 0
_________________________________________________________________
```

Figure A.2. Details of ConvNet architecture for frame classification

B. Source Code

We provide our source code which includes:

- a readme file
- code for training the model in-the-clear
- code for classifying a RAVDESS video in-the-clear

- code for extracting inputs for use with the open source MPC framework MP-SPDZ[1]
- code for the MPC protocols implemented in MP-SPDZ

---

[1] `https://github.com/data61/MP-SPDZ`