# Desafío 14

## Etiquetado y reconocimiento de acciones en vídeos

Iyán Álvarez

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

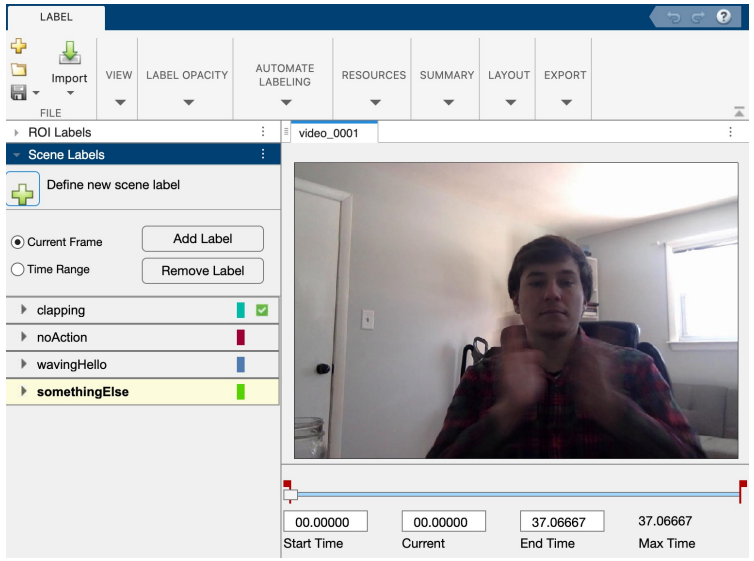
# Trabajo realizado

**Un script, funciones y etiquetado mediante Video Labeler.**

**El script “AccionesVideo_main” entrena un modelo SlowFast basado en la red resnet50-3d para la detección de cuatro acciones diferentes en videos y evalúa su rendimiento.**

**El etiquetado de escenas se ha realizado utilizando la herramienta Video Labeler de Matlab. Se han definido 4 etiquetas de escena y se ha definido los timeframes del video en el que se realiza cada una de las acciones.**

**Las funciones de apoyo han sido extraídas del ejemplo propuesto y se han realizado pequeñas modificaciones en alguna de ellas.**


# Datos

**Los datos utilizados han sido los videos de ejemplo propuestos en el ejemplo propuesto de Matlab: “Gesture Recognition using Videos and Deep Learning”.**

# Definición de SlowFast

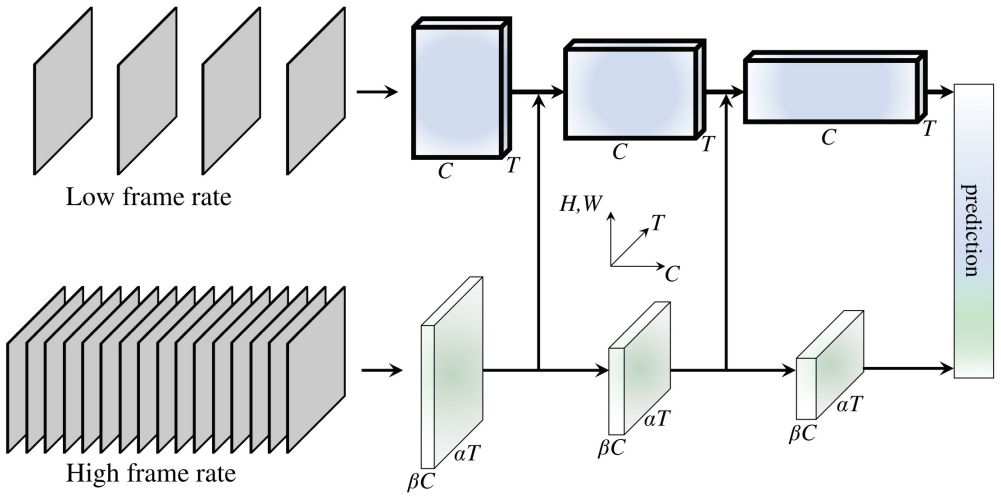
SlowFast Video Classifier es un modelo de red neuronal desarrollado para el reconocimiento de acciones en videos.

Utiliza una combinación de dos tipos de redes convolucionales:

- Slow: una "lenta" que captura detalles temporales finos a baja velocidad
- Fast: otra "rápida" que captura cambios rápidos de acciones.

Esta combinación permite al modelo procesar videos de manera más eficiente y precisa, capturando tanto detalles temporales finos como cambios rápidos de acción.

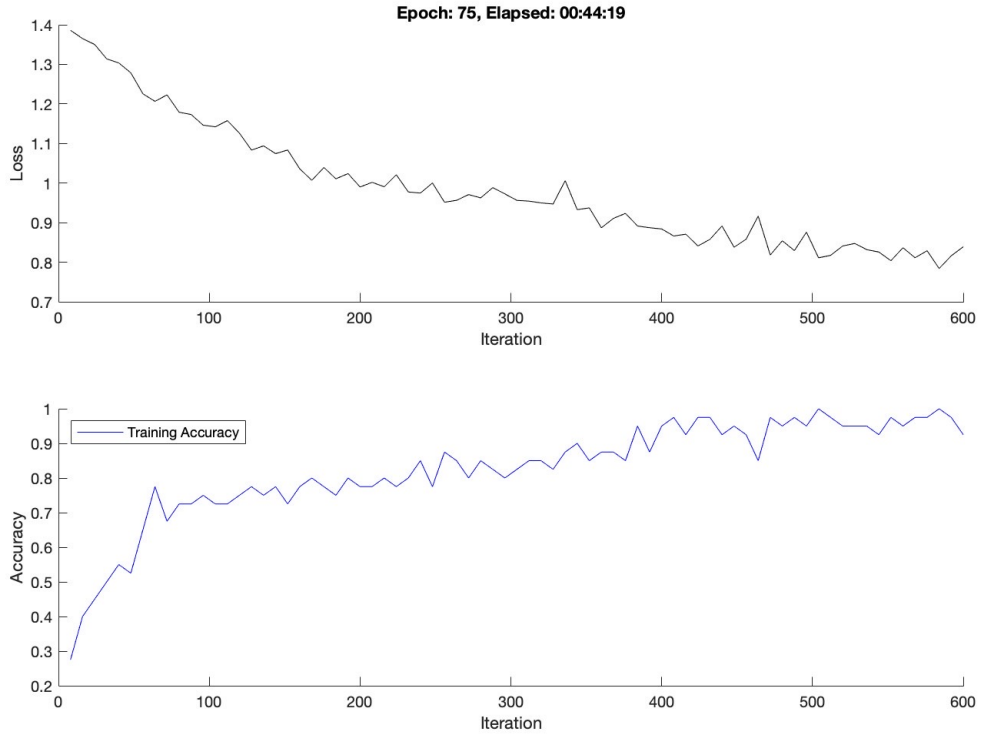
El SlowFast Video Classifier se basa en la arquitectura de redes neuronales ResNet50-3D. Esta arquitectura es una extensión tridimensional de la popular ResNet50, que se utiliza para el reconocimiento de imágenes estáticas. Esta versión está diseñada específicamente para el procesamiento de videos, capturando tanto la información espacial como temporal de los fotogramas.



# Entrenamiento de SlowFast

## Valores importantes sobre los datos:

- numFrames = 16;
- frameSize = [112,112];
- numChannels = 3;




## Valores importantes sobre el entrenamiento:

- Classes = categorical({'noAction', 'wavingHello’, 'clapping', 'somethingElse'})
- MiniBatchSize = 5
- NumIterations = 600
- Momentum = 0.9
- MinLearningRate = 1e-4
- MaxLearningRate = 1e-3

```
Training on CPU.
Training on parallel cluster 'Processes'.
NumIterations:600
MiniBatchSize:5
Classes:noAction,wavingHello,clapping,somethingElse
|=========================================================================================|
| Epoch | Iteration | Time Elapsed | Mini-Batch | Mini-Batch |   Base Learning   | Train Time |
|       |           |  (hh:mm:ss)  |  Accuracy  |    Loss    |       Rate        | (hh:mm:ss) |
|=========================================================================================|
|   1   |     8     |   00:01:00   |   27.50%   |   1.4012   | 0.0009858624225 |  00:01:00  |
|   2   |    16     |   00:01:34   |   22.50%   |   1.3872   | 0.0009443380060 |  00:00:34  |
|   3   |    24     |   00:02:08   |   32.50%   |   1.3618   | 0.0008780358823 |  00:00:33  |
|   4   |    32     |   00:02:41   |   45.00%   |   1.3270   | 0.0007911220577 |  00:00:32  |
|   5   |    40     |   00:03:13   |   57.50%   |   1.3010   | 0.0006890576475 |  00:00:32  |
|   6   |    48     |   00:03:46   |   62.50%   |   1.2895   | 0.0005782557338 |  00:00:33  |
|   7   |    56     |   00:04:19   |   67.50%   |   1.2524   | 0.0004656784084 |  00:00:32  |
|   8   |    64     |   00:04:51   |   67.50%   |   1.2377   | 0.0003583993188 |  00:00:31  |
|   9   |    72     |   00:05:24   |   70.00%   |   1.2088   | 0.0002631592046 |  00:00:32  |
|  10   |    80     |   00:05:56   |   82.50%   |   1.2092   | 0.0001859423525 |  00:00:32  |
|  11   |    88     |   00:06:30   |   72.50%   |   1.1993   | 0.0001316005814 |  00:00:33  |
|  12   |    96     |   00:07:03   |   72.50%   |   1.2022   | 0.0001035483844 |  00:00:32  |
|  13   |    104    |   00:07:35   |   65.00%   |   1.1958   | 0.0009991120278 |  00:00:32  |
|  14   |    112    |   00:08:07   |   75.00%   |   1.1470   | 0.0009920292628 |  00:00:32  |
|  15   |    120    |   00:08:40   |   80.00%   |   1.1188   | 0.0009779754323 |  00:00:32  |
|  16   |    128    |   00:09:13   |   72.50%   |   1.1742   | 0.0009571721736 |  00:00:32  |
|  17   |    136    |   00:09:47   |   75.00%   |   1.1041   | 0.0009299475665 |  00:00:33  |
|  18   |    144    |   00:10:20   |   87.50%   |   1.0454   | 0.0008967309592 |  00:00:33  |
|  19   |    152    |   00:10:54   |   87.50%   |   1.0037   | 0.0008580461977 |  00:00:33  |
|  20   |    160    |   00:11:27   |   82.50%   |   1.0315   | 0.0008145033635 |  00:00:32  |
|  21   |    168    |   00:12:00   |   85.00%   |   1.0211   | 0.0007667891533 |  00:00:33  |
|  22   |    176    |   00:12:33   |   82.50%   |   1.0162   | 0.0007156560487 |  00:00:32  |
|  23   |    184    |   00:13:06   |   92.50%   |   0.9805   | 0.0006619104492 |  00:00:33  |
|  24   |    192    |   00:13:39   |   90.00%   |   0.9507   | 0.0006063999551 |  00:00:33  |
|  25   |    200    |   00:14:12   |   92.50%   |   0.9442   | 0.0005500000000 |  00:00:33  |
|  26   |    208    |   00:14:45   |   95.00%   |   0.8911   | 0.0004936000449 |  00:00:33  |
|  27   |    216    |   00:15:18   |   97.50%   |   0.9034   | 0.0004380895508 |  00:00:32  |
|  28   |    224    |   00:15:51   |   97.50%   |   0.8629   | 0.0003843439513 |  00:00:32  |
|  29   |    232    |   00:16:24   |  100.00%   |   0.8740   | 0.0003332108467 |  00:00:32  |
|  30   |    240    |   00:16:58   |   80.00%   |   0.9780   | 0.0002854966365 |  00:00:33  |
```

# Evaluación de SlowFast

El modelo obtiene una accuracy media de 0.98 en el conjunto de test.

Además analizando la matriz de confusión podemos observar que las predicciones obtenidas son casi siempre acertadas en las acciones “clapping” y “wavingHello”.

Mientras que en los casos de ”noAction” y “somethingElse” puede existir alguna que otra equivocación más debido a que la definición de estas acciones es más abstracta y extensa.

Por lo general, los resultados son excepcionales.