

# PRÁCTICA PROFESIONAL II

EL5901-2 PRÁCTICA  
PROFESIONAL II

Diego Campanini García

# ACTIVIDAD DURANTE LA PRÁCTICA

Realizar un sistema de detección de acciones de personas



**RoboCup**  
LEIPZIG  
GERMANY **2016**  
30 JUNE to 4 JULY



# METODOLOGÍA

1. Investigar sobre el estado del arte en detección de acciones de personas
2. Buscar implementaciones disponibles en la red
3. Instalar el framework Caffe
4. Probar modelos encontrados
5. Caracterizar estadísticamente el sistema implementado
6. Implementar el sistema tal que funcione con una cámara

Caffe



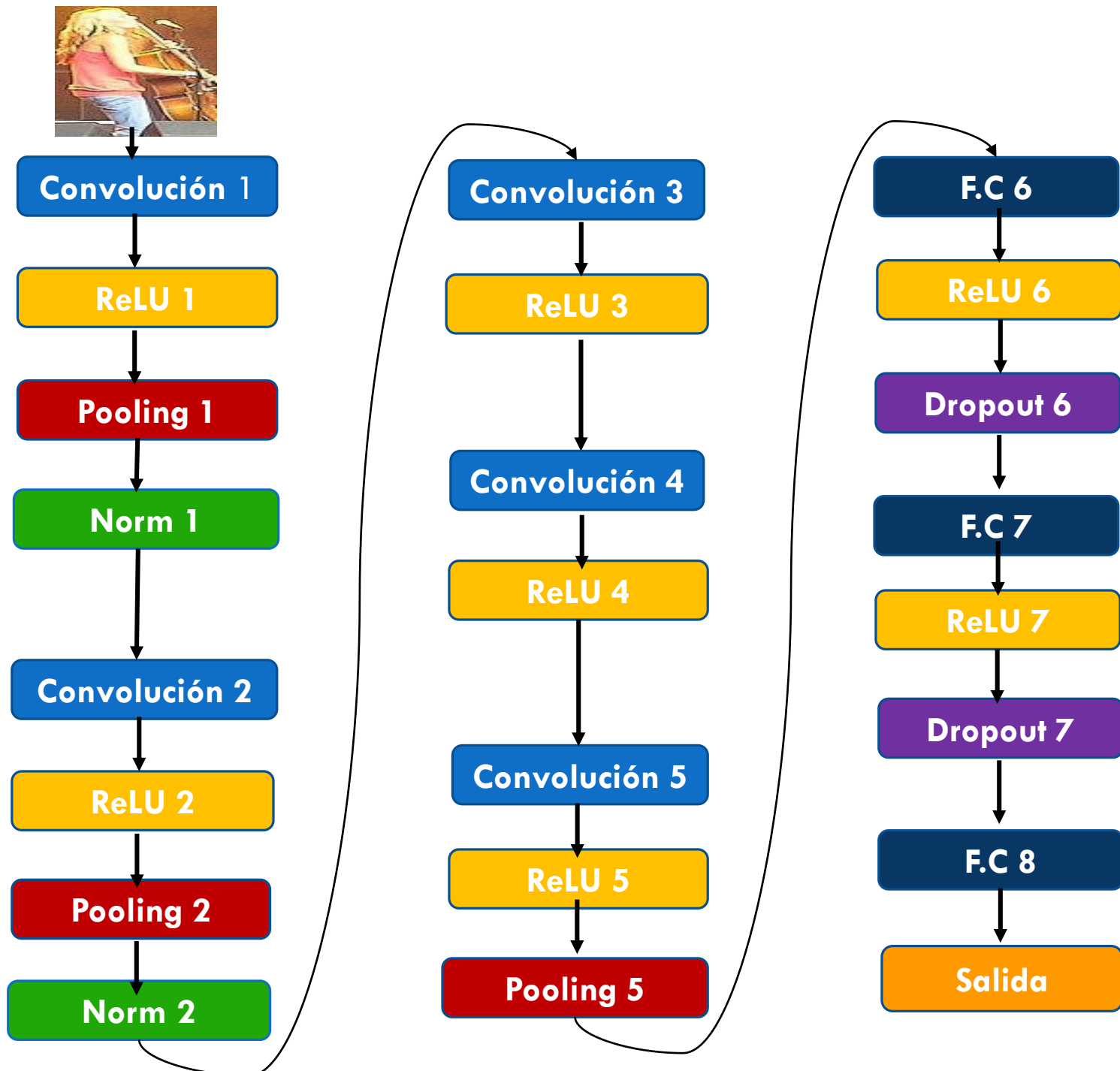






# MODELO IMPLEMENTADO

- Utiliza CNN mediante el framework caffe y corresponde al trabajo de Georgia Gkioxari et al [1], de la Universidad de Berkeley, USA.
- La CNN consiste en redes neuronales que realizan una convolución sobre la imagen para extraer características y poder clasificar las clases para las cuales se entrena la red
- La red cuenta con 16 capas, divididas en convolucionales, pooling, normalización, full connection y softmax.



# RESULTADOS

1. Primero se cortaron las imágenes tal que solo se consideren personas
2. Se etiquetó la base de datos prueba de 5360 imágenes
3. Obtención del average precisión y de la matriz de confusión del sistema









# BASE DE DATOS DE PRUEBA

Acción	N° Imágenes	%
Hablar por teléfono	455	8.48
Tocar un Instrumento	636	11.86
Leer	552	10.29
Andar en Bicicleta o Motocicleta	594	11.08
Montar a caballo	540	10.07
Correr	583	10.87
Tomar una fotografía	453	8.45
Usar computador	449	8.37
Caminar	601	11.21
Saltar	497	9.27

**Tabla 1:** Distribución de las clases en el conjunto de prueba

# RESULTADOS

	Hablar por teléfono	Tocar Instrumento	Leer	Andar en Bicicleta	Montar a caballo	Correr	Tomar Fotografía	Usar Computador	Caminar	Saltar	mAP
AP	43.0	70.5	37.7	86.1	82.5	79.9	44.1	52.0	56.6	69.9	62.23

**Tabla 2:** Average Precision del sistema medido en el conjunto de prueba del desafío PASCAL VOC 2012

	Hablar por teléfono	Tocar Instrumento	Leer	Andar en Bicicleta	Montar a caballo	Correr	Tomar Fotografía	Usar Computador	Caminar	Saltar	mAP
Standford	75.7	44.8	66.6	44.4	93.2	94.2	87.9	38.4	70.6	75.6	69.1
Oxford	77.0	50.4	65.3	39.5	94.1	95.9	87.7	42.7	68.8	74.5	69.6
Oquad et al [4]	74.8	46.0	75.6	45.3	93.5	95.0	86.5	49.3	66.7	69.5	70.2
Action R-CNN	76.2	47.4	77.5	42.2	94.9	94.3	87.0	52.9	66.5	66.5	70.5

**Tabla 3:** Average Precision de otros sistemas para PASCAL VOC 2012



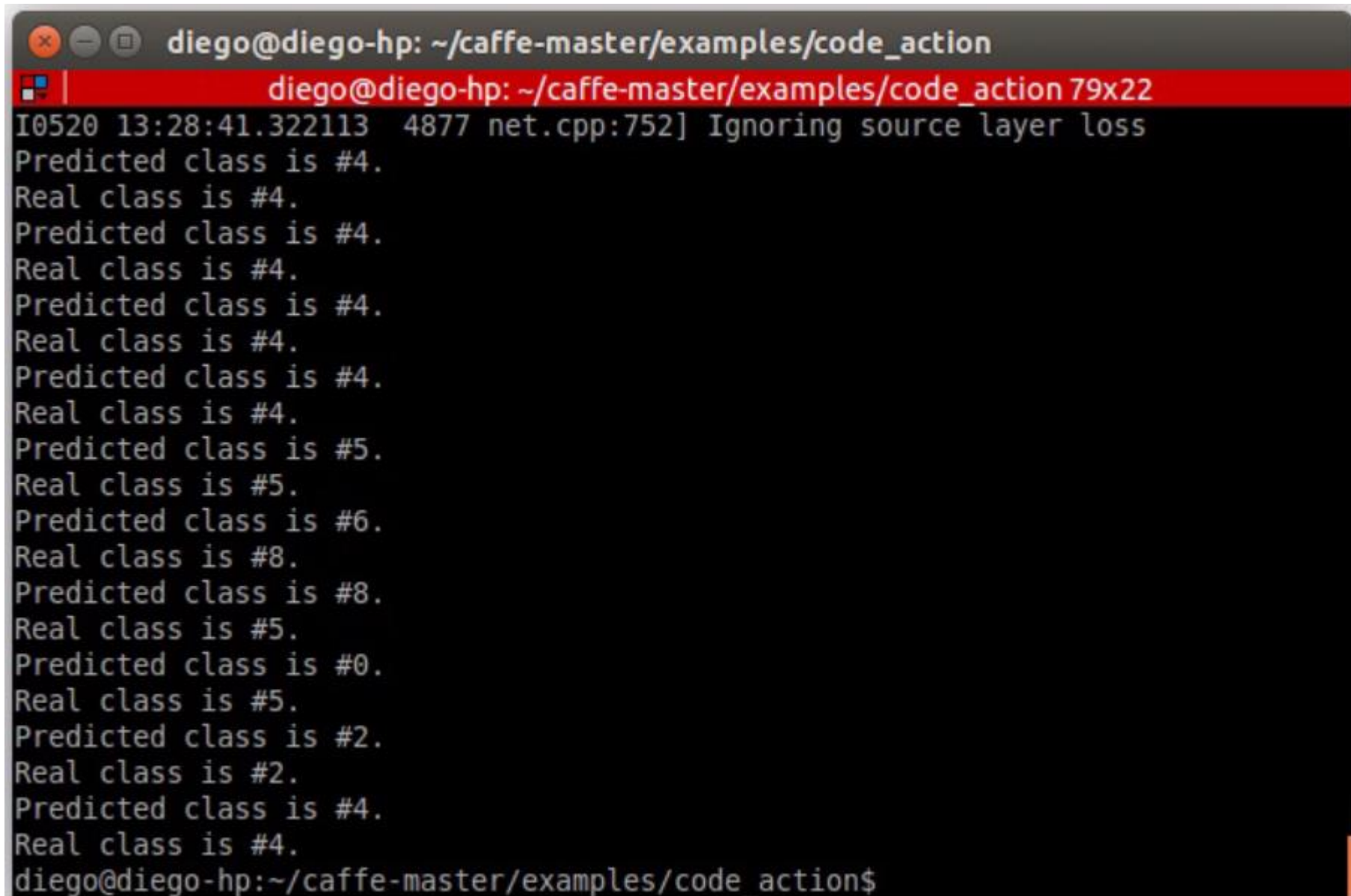
# RESULTADOS

		Predicho									
		1	2	3	4	5	6	7	8	9	10
Real	1	237 (4.42 %)	41 (0.76 %)	44 (0.82 %)	5 (0.09 %)	13 (0.24 %)	5 (0.09 %)	48 (0.89 %)	27 (0.50 %)	25 (0.46 %)	10 (0.18 %)
	2	27 (0.50 %)	460 (8.58 %)	44 (0.82 %)	11 (0.20 %)	18 (0.33 %)	3 (0.05 %)	10 (0.18 %)	30 (0.55 %)	19 (0.35 %)	14 (0.26 %)
	3	54 (0.98 %)	70 (1.30 %)	213 (3.97 %)	6 (0.11 %)	31 (0.57 %)	8 (0.14 %)	18 (0.33 %)	122 (2.27 %)	16 (0.29 %)	15 (0.27 %)
	4	1 (0.01 %)	10 (0.18 %)	8 (0.14 %)	505 (9.42 %)	34 (0.63 %)	5 (0.09 %)	3 (0.05 %)	3 (0.05 %)	17 (0.31 %)	8 (0.14 %)
	5	5 (0.09 %)	10 (0.18 %)	9 (0.16 %)	45 (0.83 %)	441 (8.22 %)	1 (0.01 %)	10 (0.18 %)	7 (0.13 %)	7 (0.13 %)	5 (0.09 %)
	6	6 (0.11 %)	3 (0.05 %)	20 (0.37 %)	12 (0.22 %)	17 (0.31 %)	419 (7.81 %)	15 (0.27 %)	3 (0.05 %)	70 (1.30 %)	18 (0.33 %)
	7	45 (0.83 %)	55 (1.02 %)	22 (0.41 %)	20 (0.37 %)	35 (0.65 %)	6 (0.11 %)	192 (3.58 %)	20 (0.37 %)	42 (0.78 %)	16 (0.29 %)
	8	35 (0.65 %)	43 (0.80 %)	73 (1.36 %)	2 (0.03 %)	8 (0.14 %)	5 (0.09 %)	15 (0.27 %)	257 (4.79 %)	5 (0.09 %)	6 (0.11 %)
	9	19 (0.35 %)	21 (0.39 %)	18 (0.33 %)	41 (0.76 %)	32 (0.59 %)	52 (0.97 %)	27 (0.50 %)	14 (0.26 %)	349 (6.51 %)	28 (0.52 %)
	10	9 (0.16 %)	17 (0.31 %)	15 (0.27 %)	25 (0.46 %)	15 (0.27 %)	23 (0.42 %)	15 (0.27 %)	9 (0.16 %)	27 (0.50 %)	342 (6.38 %)

Tabla 4: Matriz de confusión del sistema implementado

1: Hablar por teléfono, 2: Tocar un Instrumento, 3: Leer, 4: Andar en Bicicleta,  
5: Montar a caballo, 6: Correr, 7: Tomar fotografía, 8: Usar Computador, 9: Caminar, 10: Saltar

# SISTEMA IMPLEMENTADO



```
diego@diego-hp: ~/caffe-master/examples/code_action
diego@diego-hp: ~/caffe-master/examples/code_action 79x22
I0520 13:28:41.322113 4877 net.cpp:752] Ignoring source layer loss
Predicted class is #4.
Real class is #4.
Predicted class is #4.
Real class is #4.
Predicted class is #4.
Real class is #4.
Predicted class is #4.
Real class is #4.
Predicted class is #5.
Real class is #5.
Predicted class is #6.
Real class is #8.
Predicted class is #8.
Real class is #5.
Predicted class is #0.
Real class is #5.
Predicted class is #2.
Real class is #2.
Predicted class is #4.
Real class is #4.
diego@diego-hp:~/caffe-master/examples/code_action$
```

The image shows a terminal window with a dark background and light-colored text. The window title bar at the top indicates the user is 'diego' on a machine named 'diego-hp', and the current directory is '~/caffe-master/examples/code\_action'. The terminal output shows the execution of a Caffe model. It starts with a message 'I0520 13:28:41.322113 4877 net.cpp:752] Ignoring source layer loss'. This is followed by a series of 16 pairs of lines, each pair representing a prediction and a real class. The first 8 pairs show 'Predicted class is #4.' and 'Real class is #4.'. The next 2 pairs show 'Predicted class is #5.' and 'Real class is #5.'. The next 2 pairs show 'Predicted class is #6.' and 'Real class is #8.'. The next 2 pairs show 'Predicted class is #8.' and 'Real class is #5.'. The final 2 pairs show 'Predicted class is #0.' and 'Real class is #5.', and 'Predicted class is #2.' and 'Real class is #2.'. The last line shows 'Predicted class is #4.' and 'Real class is #4.'. The prompt 'diego@diego-hp:~/caffe-master/examples/code\_action\$' is visible at the bottom.

# SISTEMA IMPLEMENTADO



# CONCLUSIONES

- Las CNN son capaces de mejorar los resultados previos en la clasificación de acciones de personas en el desafío PASCAL VOC 2012
- Se comprobó que se obtienen resultados inferiores al utilizar solamente CNN para la clasificación. Se logra mejores resultados al utilizar CNN como extractor de características y SVM para clasificar.
- Trabajo futuro implementar un detector de personas, más el sistema de detección de acciones.



# BIBLIOGRAFÍA

1. Georgia Gkioxari, Bharath Hariharan, Ross Girshick and Jitendra Malik.  
R-CNNs for Pose Estimation and Action Detection
2. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner  
Gradient-Based Learning Applied to Document Recognition.