

Reconociendo la actividad humana en videos

Trabajo Fin de Máster Universitario en Inteligencia Artificial

Presentado por: Madariaga Lasala, Pilar

Director: Gallego Gómez, Jenaro



Teoría del aprendizaje por observación

¿Cómo aprende el ser humano?



Inteligencia Artificial



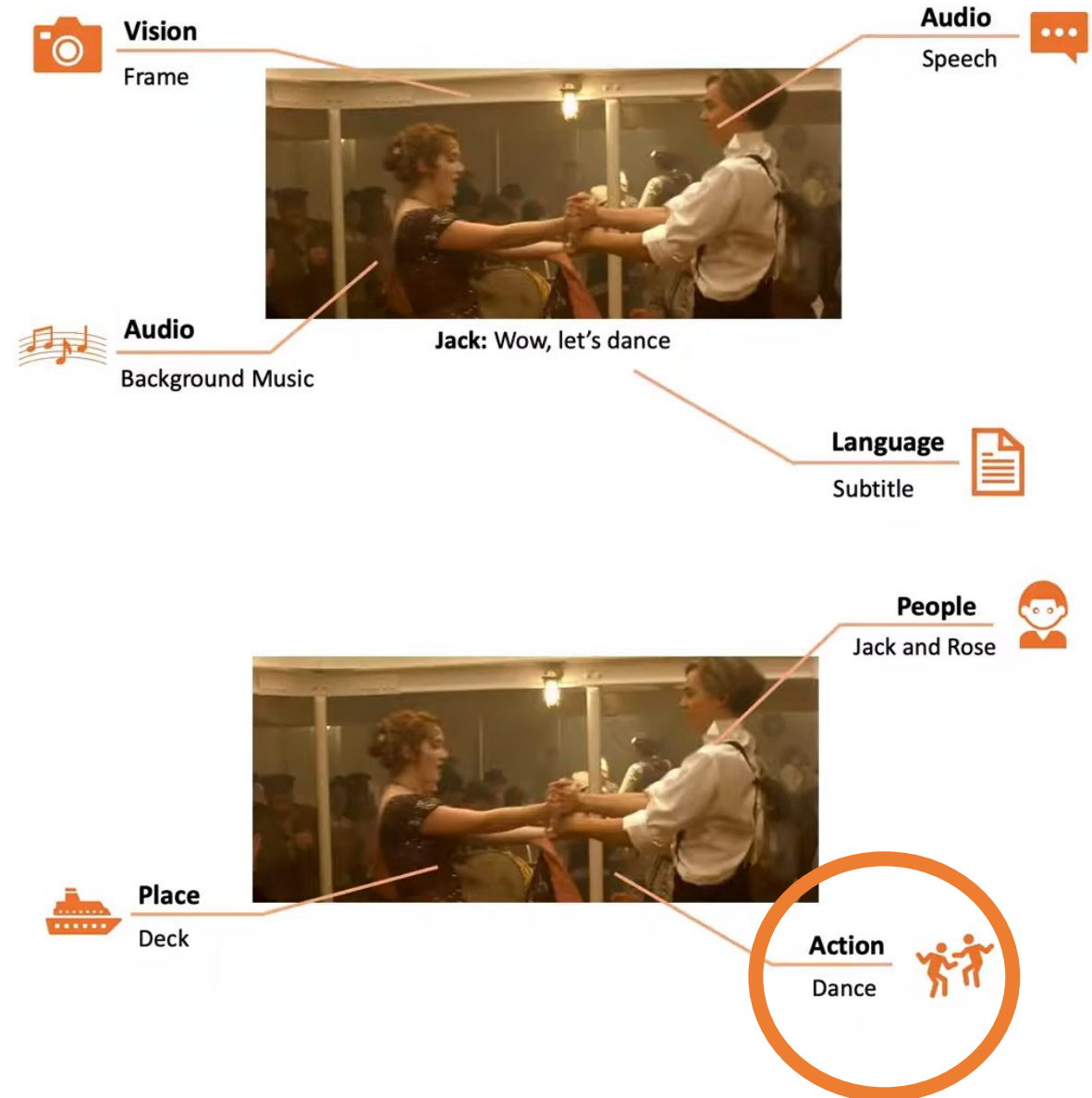
¿Cómo crear modelos artificiales que aprendan a reconocer las acciones?



1.Semantic Video Understanding

Multi Semantic Elements

Multi Modality



2. Image Understanding

¿Aplicamos las mismas técnicas
en los videos?

Image Classification



Object Detection



Semantic Segmentation

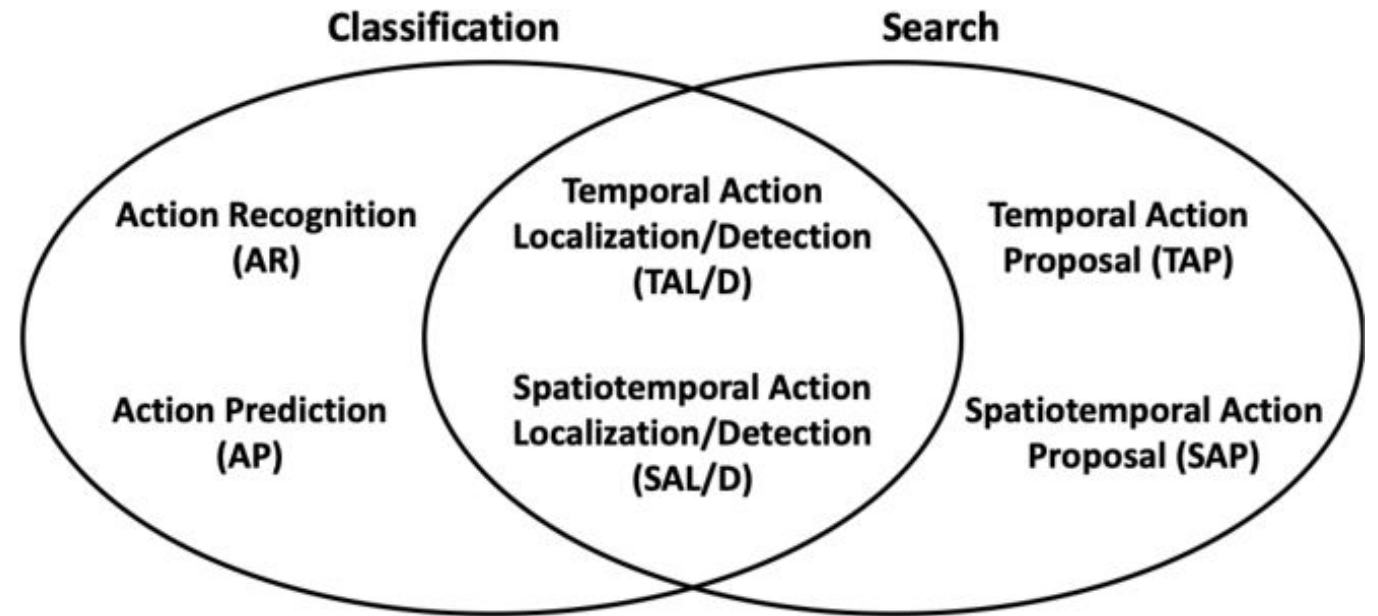


Instance Segmentation

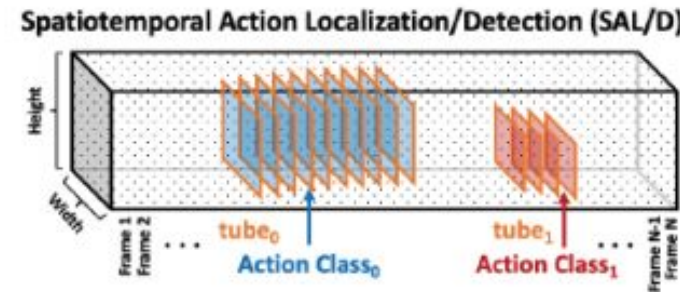
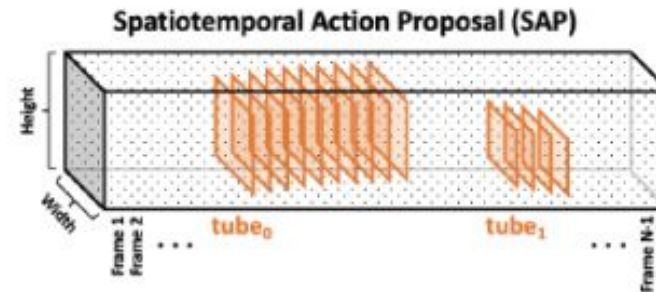
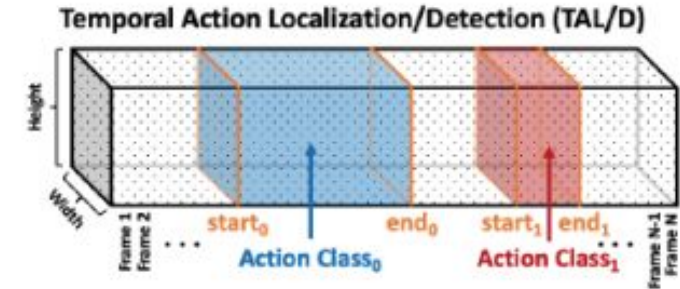
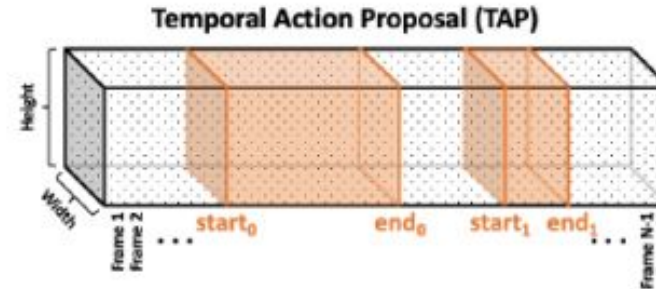
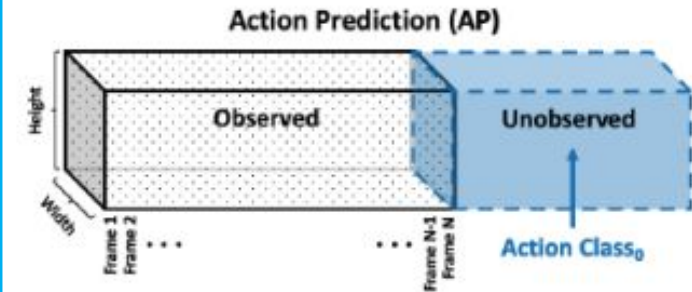
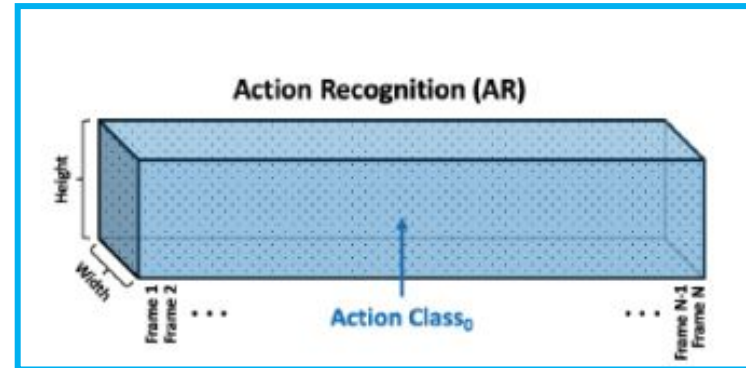


3. ¿Qué es una acción?

Video action understanding



4.Video Action Understanding

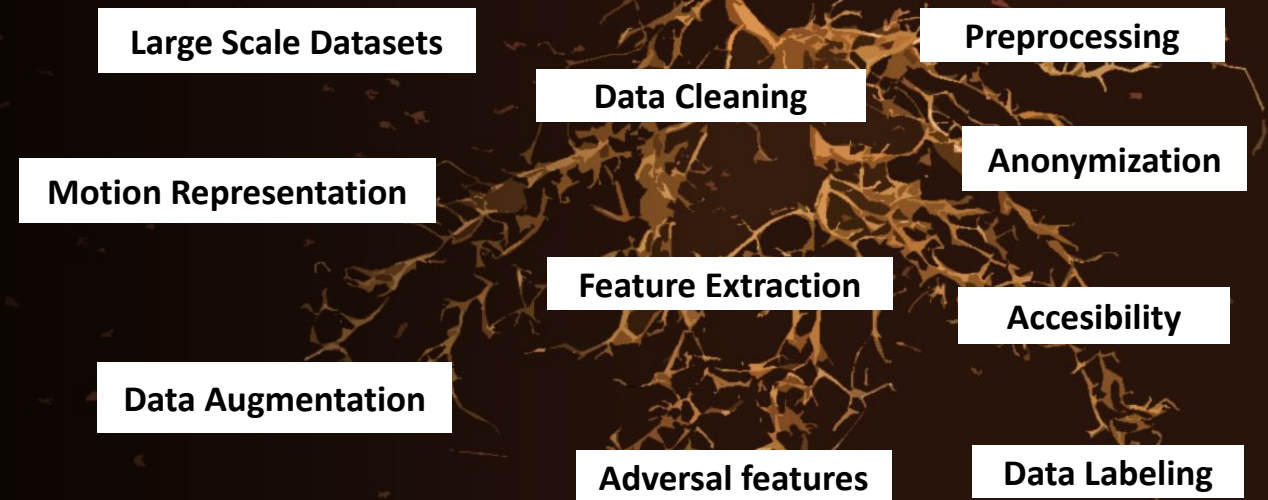


Video Human Action Recognition

Fine-grained actions
(Spatiotemporal data)

“Wild videos”

**Factores relevantes en
escenarios realistas**



Deep Neural Networks

Precisión

Tamaño

Latencia

**¿Qué tipos son adecuadas para
reconocer acciones en videos?**

Emular capacidades

Action Recognition API

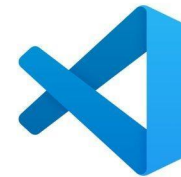


- *Ballet*
- *Break*
- *Flamenco*
- *Waltz*

Let's Dance: Learning From Online Dance Videos.

Let's do it!

Let's Dance



Líneas de Investigación

1. Extensión clasificación de imágenes
2. Representación del movimiento

1. MobileNet

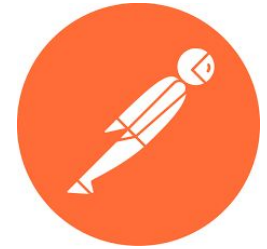
- Enfoque I : Single Frame
- Características Espaciales
- Data Agumentation

2. CNN

- Enfoque II: Secuencia de Frames
- Características espaciotemporales
- Capa Convolutonal filtro 3D

Evaluación

Escenario de Entrenamiento
Escenario de Inferencia



	MODELO 1	MODELO 2
DIMENSIONES FRAME	224 x 224 x 3	64 x 64 x 3
TIPOLOGÍA	MobileNet	CNN filtros 3D
TRAIN / VALIDATION /TEST	3600/1200/1800	4722/1574/2385
DATA AUGMENTATION	Sí (Batch size 32)	No
SECUENCIAS	No	Sí (5 frames)
OPTIMIZADOR	ADAM	ADAM
GPU	No	No

Tabla 1. *Diseño y configuración de los modelos*

	MODELO 1	MODELO 2
LOSS / ACCURACY	0.0018 / 1,00	0.0542 / 0.9900
VAL LOSS / VAL ACCURACY	0.45 / 0.85	1.0706 / 0.7103
ENTORNO DE ENTRENAMIENTO	INTEL(R) CORE(TM) I5-6300U CPU	INTEL(R) CORE(TM) I5-6300U CPU
TOTAL PARAMETROS	3,232,964	986,852
TRAINABLE PARAMS	1,867,780	986,852
EPOCHS	10	10
AVG TIME PER EPOCH	223S	213S
TEST LOSS / TEST ACCURACY	0.556 / 0.837	1.143 / 0.687

Tabla 2. Métricas de los modelos en la Fase de entrenamiento y test



E3) <http://127.0.0.1:5000/api/v1/predict>



E3) <http://127.0.0.1:5000/api/v2/predict>

```
{
  "predicted_label": "flamenco",
  "probs": [
    "ballet: 20.21%",
    "break: 14.50%",
    "flamenco: 62.84%",
    "waltz: 2.44%"
  ],
  "url": "https://www.youtube.com/watch?v=Wz_f9B4pPtg"
}
```

M1

```
{
  "predicted_label": "ballet",
  "probs": [
    "ballet: 78.36%",
    "break: 19.48%",
    "flamenco: 1.95%",
    "waltz: 0.21%"
  ],
  "url": "https://www.youtube.com/watch?v=Wz_f9B4pPtg"
}
```

M2



E1) <http://127.0.0.1:5000/api/v1/predict>



E1) <http://127.0.0.1:5000/api/v2/predict>

```
{
  "predicted_label": "break",
  "probs": [
    "ballet: 0.23%",
    "break: 93.72%",
    "flamenco: 0.62%",
    "waltz: 5.44%"
  ],
  "url": "https://www.youtube.com/watch?v=-pJcLfc0Cw"
}
```

M1

```
{
  "predicted_label": "flamenco",
  "probs": [
    "ballet: 0.49%",
    "break: 16.83%",
    "flamenco: 56.17%",
    "waltz: 26.51%"
  ],
  "url": "https://www.youtube.com/watch?v=-pJcLfc0Cw"
}
```

M2

Conclusiones

Líneas de trabajo futuro

- Complejidad de las acciones de grano fino
 - Multi-modal & Multi-label Video Datasets
 - Soluciones Innovadoras
 - Integración dispositivos embebidos
-
- Uso de diseños alternativos
 - Ampliar los servicios de la API

Gracias

