Reconociendo la actividad humana en videos

Trabajo Fin de Máster Universitario en Inteligencia Artificial

Presentado por: Madariaga Lasala, Pilar

Director: Gallego Gómez, Jenaro

unir
LA UNIVERSIDAD
EN INTERNET

Teoría del aprendizaje por observación

¿Cómo aprende el ser humano?

unir
LA UNIVERSIDAD
EN INTERNET

Inteligencia Artificial

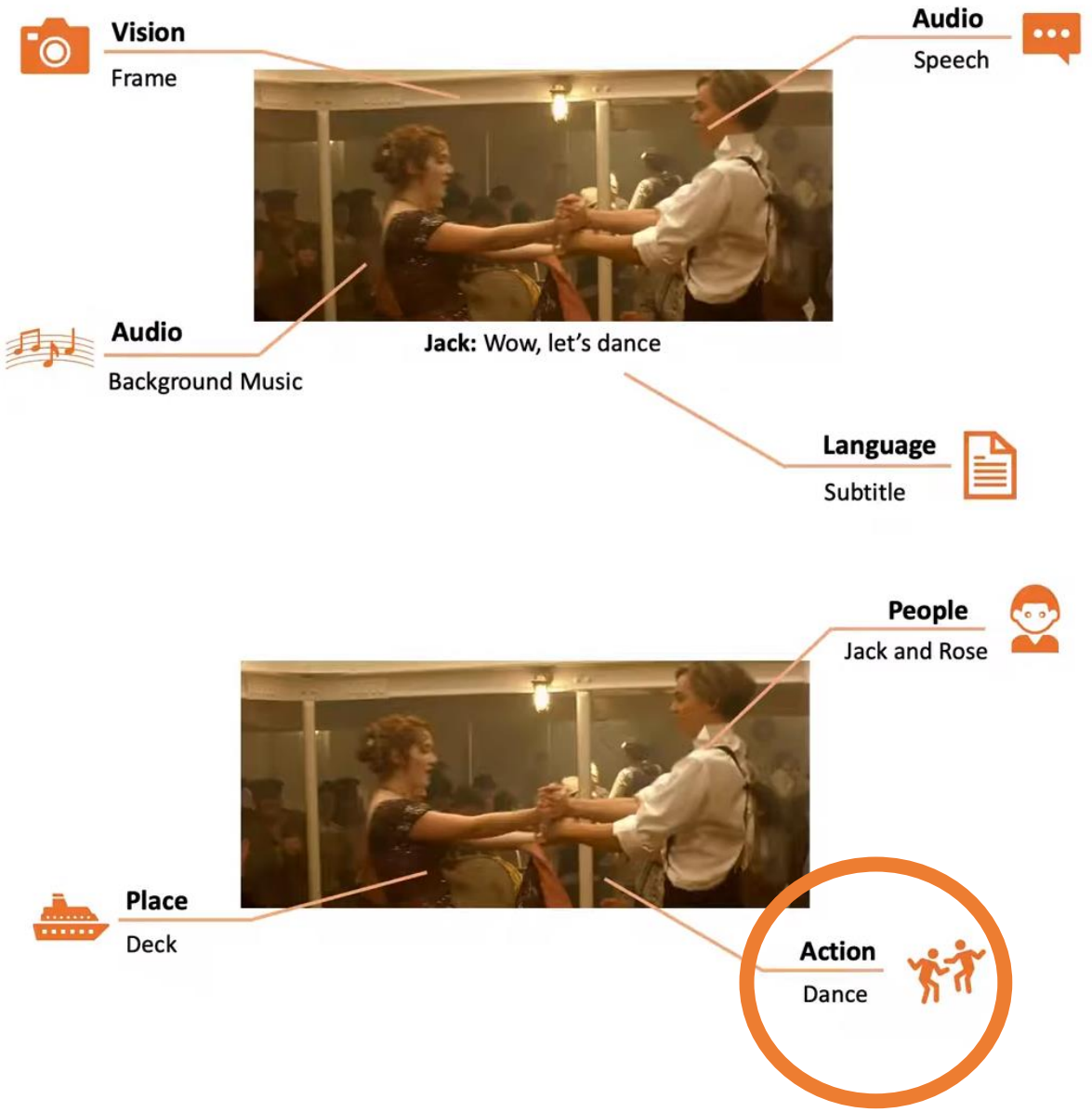
¿Cómo crear modelos artificiales que aprendan a reconocer las acciones?

unir
LA UNIVERSIDAD
EN INTERNET

1.Semantic Video Understanding

Multi Semantic Elements

Multi Modality

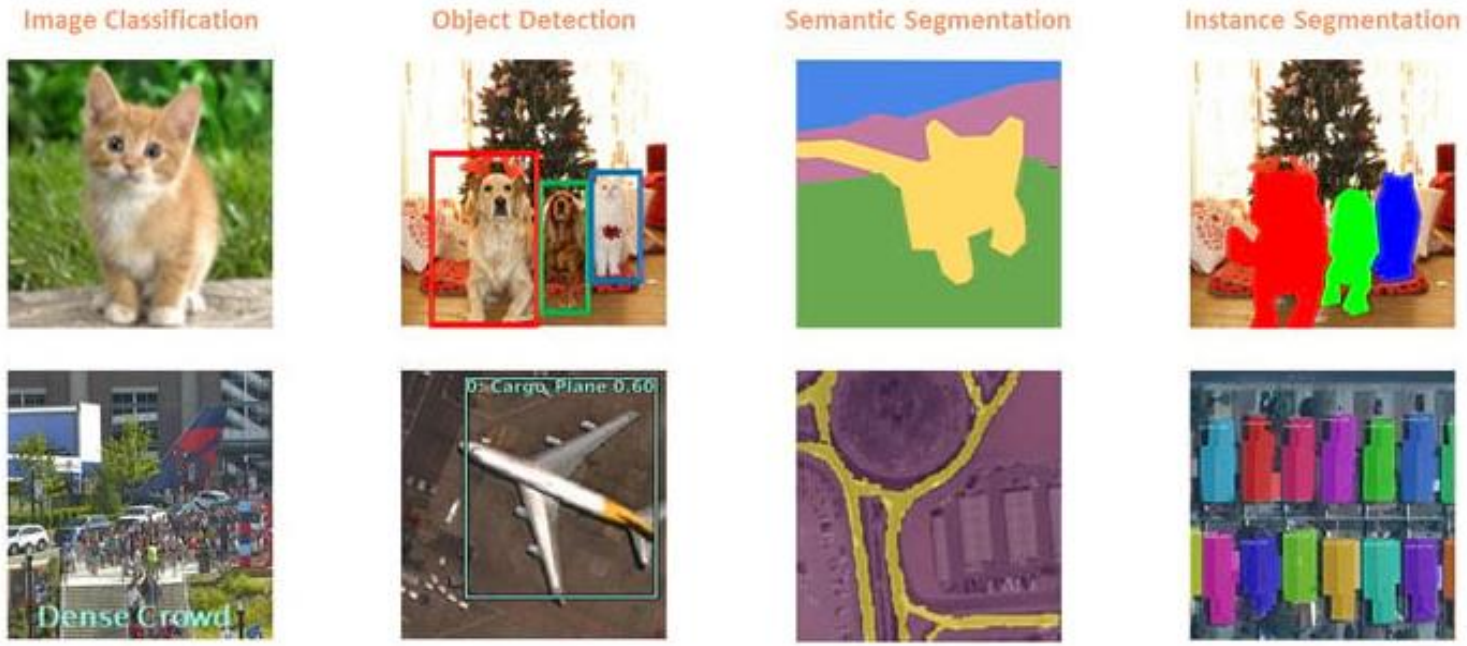


unir
LA UNIVERSIDAD
EN INTERNET

2.Image Understanding

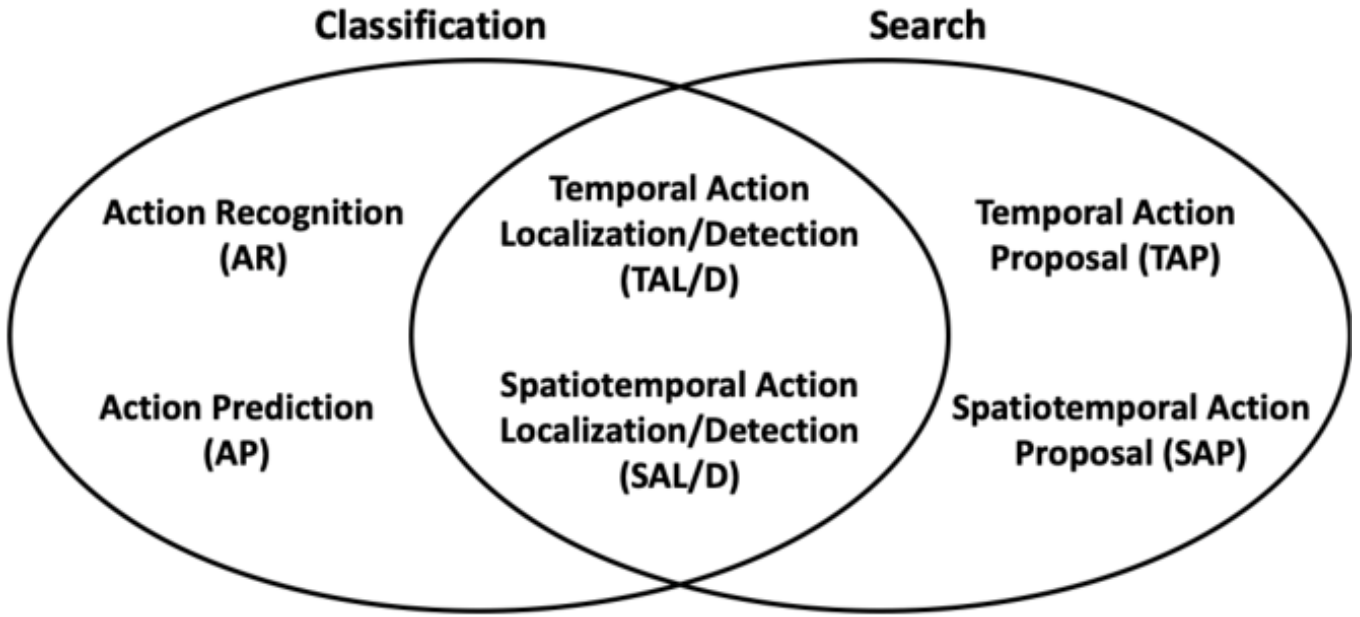
¿Aplicamos las mismas técnicas

en los videos?



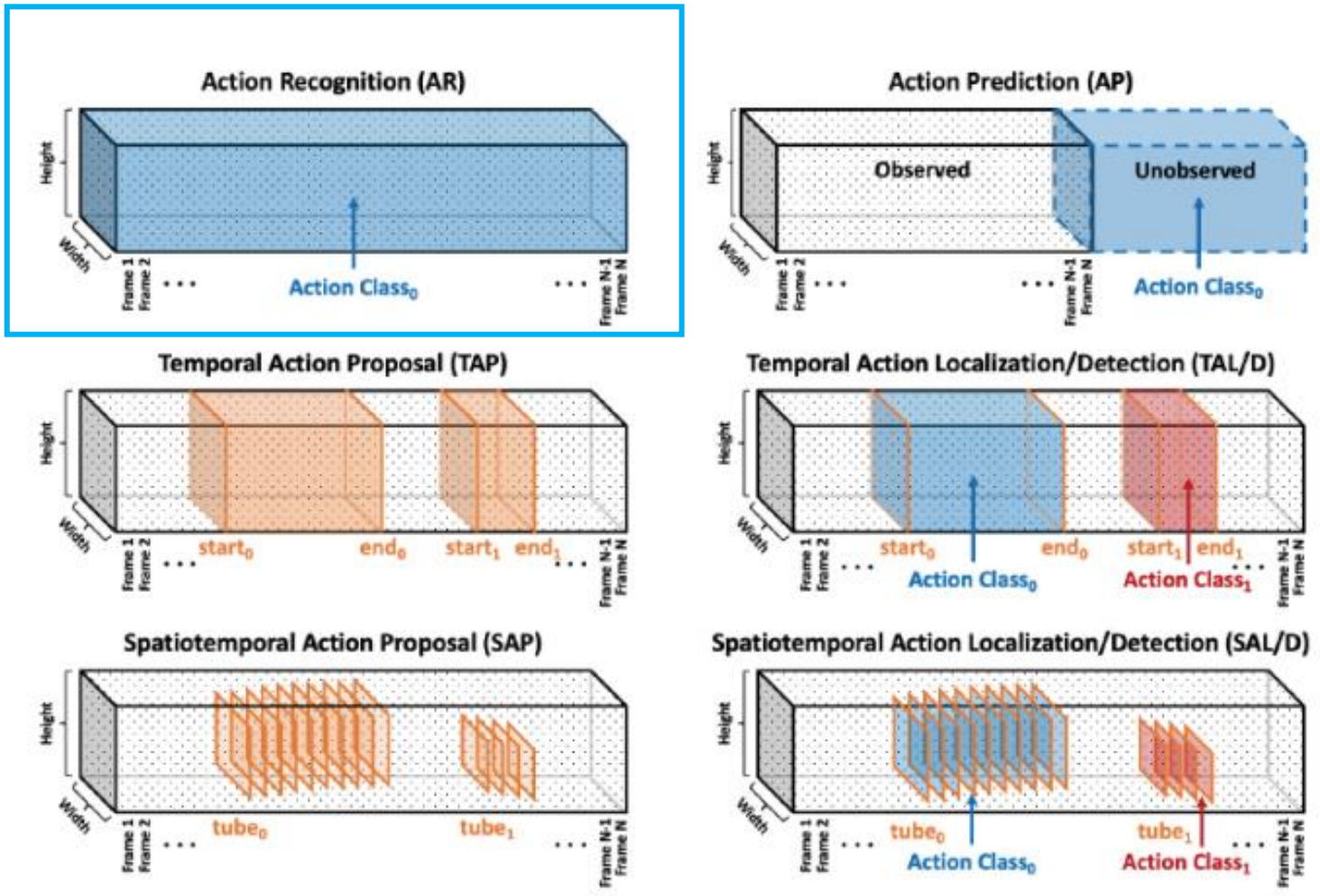
3. ¿Qué es una acción?

Video action understanding



unir
LA UNIVERSIDAD
EN INTERNET

4.Video Action Understanding



Video Human Action Recognition

Fine-grained actions

(Spatiotemporal data)

unir
LA UNIVERSIDAD
EN INTERNET

“Wild videos”

Factores relevantes en escenarios realístas

unir
LA UNIVERSIDAD EN INTERNET

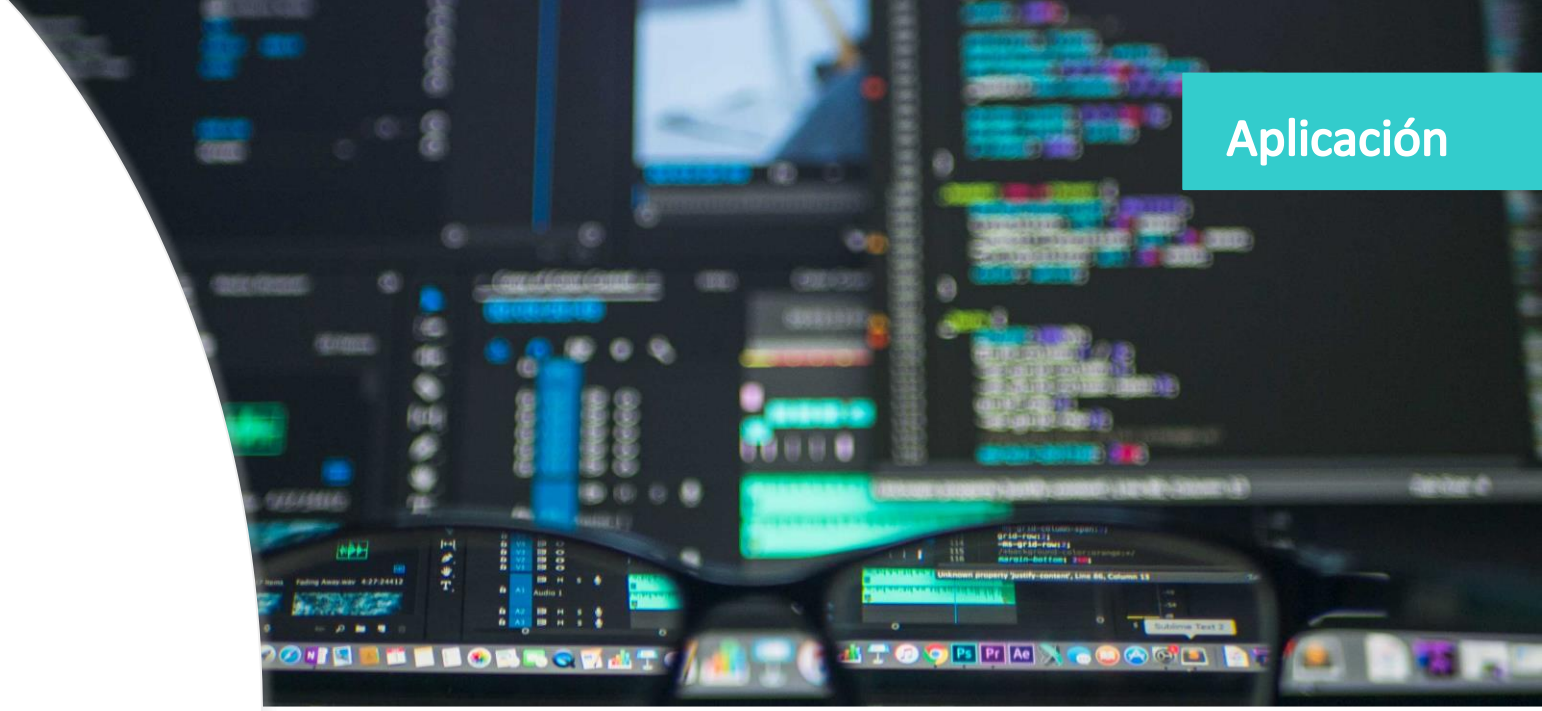
Deep Neural Networks

- Precisión
- Tamaño
- Latencia

¿Qué tipos son adecuadas para reconocer acciones en videos?

unir
LA UNIVERSIDAD
EN INTERNET

Aplicación



Emular capacidades

Action Recognition API

- *Ballet*
- *Break*
- *Flameco*
- *Waltz*

Let’s Dance: Learning From Online Dance Videos.

unir
LA UNIVERSIDAD
EN INTERNET

Let’s do it!






Let’s Dance



Líneas de Investigación

1. Extensión clasificación de imágenes
2. Representación del movimiento

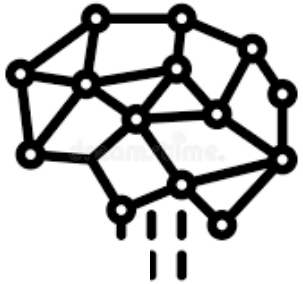
1. MobileNet

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

2. CNN

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

unir
LA UNIVERSIDAD
EN INTERNET



YouTube

Evaluación

Escenario de Entrenamiento
Escenario de Inferencia

Escenario de Configuración

	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*

Escenario de Entrenamiento y Test

	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*

Escenario de Inferencia

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

M1


```
{
    "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"
}
```


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

M2


```
{
    "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"
}
```


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

M1


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


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

M2


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

Conclusiones

- Complejidad de las acciones de grano fino
- Multi-modal & Multi-label Video Datasets
- Soluciones Innovadoras
- Integración dispositivos embebidos

Líneas de trabajo futuro

- Uso de diseños alternativos
- Ampliar los servicios de la API

unir
LA UNIVERSIDAD
EN INTERNET

Gracias