

1. ¿Cómo se elige el número óptimo de neuronas en una capa oculta?

La elección del número óptimo de neuronas en una capa oculta no tiene una fórmula exacta, pero existen principios prácticos y fundamentos teóricos que guían esta decisión. En general, el objetivo es encontrar un punto de equilibrio entre **capacidad de representación** y **generalización**, evitando tanto el *underfitting* como el *overfitting*.

Los criterios más utilizados incluyen:

a) Complejidad del problema

Problemas con relaciones no lineales fuertes (por ejemplo, actividades humanas en video) requieren más neuronas para poder capturar patrones complejos. Problemas simples requieren menos.

b) Dimensionalidad de los datos

Si las características de entrada tienen alta dimensionalidad o contienen estructura temporal (como secuencias de esqueletos), se suele incrementar el número de neuronas o incluso usar arquitecturas especializadas (GCN, RNN, Transformers).

c) Experimentos sistemáticos

El número óptimo generalmente se determina mediante:

- **Validación cruzada**
- **Grid search**
- **Pruebas incrementales**, aumentando o disminuyendo neuronas y observando impacto en accuracy, pérdida y sobreajuste.

d) Regularización y restricciones computacionales

Si el modelo comienza a sobreajustarse, aun teniendo buena capacidad, se reduce el número de neuronas o se aplican técnicas como dropout.

Si los recursos computacionales son limitados, también condicionan el tamaño de la red.

En conclusión, el número óptimo se elige con base en experimentación guiada, evitando redes demasiado pequeñas (que no aprenden) o demasiado grandes (que memorizan). No existe una fórmula universal, pero sí un proceso iterativo justificado por desempeño y teoría.

2. Aplicar técnicas de procesamiento de imágenes a la solución del reto: ¿son necesarias? ¿qué técnicas se usaron o por qué no son necesarias?

En el contexto del reto, aplicar técnicas tradicionales de procesamiento de imágenes **no era estrictamente necesario**, y esto se debe al enfoque elegido para resolver el problema.

a) ¿Por qué no son necesarias?

La solución de nuestro equipo se basa en un pipeline que transforma cada video en **representaciones de esqueletos**, utilizando:

- **YOLOv8** para detectar personas en cada frame.
- **MediaPipe Pose** para obtener 17 puntos clave normalizados por persona.
- Posteriormente, estos esqueletos alimentan el modelo **MP-GCN**, que clasifica actividades mediante grafos espaciales-temporales.

Dado este flujo, **el modelo nunca trabaja directamente con imágenes crudas**, sino con información estructurada (coordenadas de articulaciones). Por lo tanto:

- No necesitamos filtros, segmentación, detección de bordes o técnicas similares.
- No se requiere mejorar la calidad visual del frame, ya que los modelos de detección y pose trabajan adecuadamente con los videos originales.
- Las variaciones de luz, ruido o fondo son manejadas automáticamente por YOLO y MediaPipe, sin procesamiento adicionales.

b) ¿Qué técnicas sí se usaron?

Aunque no aplicamos técnicas clásicas de procesamiento digital de imágenes, sí usamos **técnicas modernas basadas en visión por computadora**, esenciales para la solución:

1. **Detección de objetos (YOLOv8):**
Permite identificar a las personas en cada frame y generar bounding boxes confiables.

2. Estimación de pose 2D (MediaPipe Pose):

Se utiliza para generar las coordenadas de los 17 puntos clave del cuerpo humano, que son la entrada del modelo de grafos.

3. Normalización espacial-temporal:

- Normalización de puntos clave.
- Estándar de 12 fps.
- Ventanas deslizantes de 48 frames.

Estas técnicas sustituyen por completo el procesamiento clásico de imágenes, ya que permiten obtener información estructural más relevante para el modelo MP-GCN.

c) Conclusión

En nuestro proyecto **no fue necesario aplicar técnicas tradicionales de procesamiento de imágenes**, porque el enfoque basado en detección y estimación de pose ya proporciona información mucho más útil y robusta para el análisis de actividades. Las técnicas que sí se usaron pertenecen a visión computacional moderna y son adecuadas para el tipo de datos (esqueletos) que utilizamos.