Informe Técnico: Métodos Alternativos de Pose y Segmentación para Estimación de Dimensiones de Salmones

Ernesto Gamero, Gustavo Venegas, Martín Ortega, Lucas Petit, Alonso Castillo
Asignatura: Procesamiento Digital de Imágenes [TEL328]

November 6, 2025

Abstract

La estimación precisa de las dimensiones de salmones mediante visión artificial es un desafío operativo en la acuicultura moderna. Este informe técnico evalúa tres arquitecturas de redes neuronales que complementan a YOLOv8, cada una orientada a resolver una limitación específica: HRNet-W48 para la estimación de pose de alta fidelidad, DeepLabCut para el seguimiento robusto de múltiples individuos en condiciones de oclusión, y PointRend para la segmentación de contornos con precisión sub-milimétrica. El análisis se centra en los fundamentos conceptuales de cada método, sus resultados empíricos y los criterios prácticos para su implementación en sistemas de monitoreo acuícola.

Contents

1 Introducción: El Dilema entre Velocidad y Precisión

La monitorización automática del crecimiento de salmones es fundamental para la acuicultura moderna. Sin embargo, los sistemas de visión artificial enfrentan un entorno complejo: los peces se mueven constantemente, se ocultan unos a otros y las condiciones de luz submarina son impredecibles. Aunque modelos como YOLOv8 ofrecen una solución rápida para la detección en tiempo real, su eficiencia se logra a costa de una menor precisión en las mediciones. Este informe explora arquitecturas alternativas diseñadas para escenarios donde la exactitud es un factor crítico, priorizando la calidad de los datos sobre la velocidad de procesamiento.

2 Paradigmas Complementarios a la Detección en Tiempo Real

El enfoque de YOLOv8, conocido como estrategia *top-down*, consiste en reducir progresivamente la resolución de la imagen para ganar velocidad. Este método es eficaz para la detección general, pero sacrifica detalles finos que son cruciales en tres escenarios específicos:

1. **Mediciones de alta precisión**: Cuando se requiere una exactitud sub-milimétrica para la clasificación comercial.
2. **Oclusiones severas**: En situaciones donde los peces se superponen en más de un 30% de su cuerpo, dificultando la identificación individual.
3. **Geometrías complejas**: Para la delineación precisa de estructuras anatómicas irregulares, como las aletas.

Las arquitecturas alternativas que se presentan a continuación operan bajo un principio distinto: mantienen representaciones de la imagen a múltiples resoluciones de forma paralela, preservando así la información espacial necesaria para resolver estos desafíos.

3 HRNet-W48: Preservando el Detalle Fino

3.1 El Principio de la Multi-Resolución Paralela

Las arquitecturas de visión convencionales operan de manera similar a un resumen progresivo: a medida que analizan una imagen, reducen su resolución (e.g., $1920\times1080 \rightarrow 960\times540 \rightarrow 480\times270$), perdiendo detalles finos en cada paso. Aunque este proceso es eficiente, la información espacial perdida es irrecuperable.

HRNet adopta un enfoque radicalmente distinto. En lugar de reducir la resolución secuencialmente, mantiene cuatro "ramas" de procesamiento que operan en paralelo, cada una a una escala diferente:

- **Alta resolución (1/1)**: Conserva cada detalle de la imagen original, crucial para identificar texturas y contornos precisos.

- **Resoluciones intermedias (1/4 y 1/8)**: Capturan el contexto local y las relaciones entre estructuras anatómicas cercanas.
- **Baja resolución (1/16)**: Ofrece una vista global que permite comprender la orientación y posición general del pez.

La clave de HRNet reside en sus **módulos de fusión bidireccional**, que actúan como un sistema de comunicación constante entre las distintas ramas. De este modo, la información de contexto global ayuda a interpretar los detalles finos, y a su vez, los detalles precisos enriquecen la comprensión general de la escena.

3.2 Potenciando la Arquitectura

Para mejorar aún más su rendimiento, HRNet se complementa con dos módulos avanzados:

- **CBAM (Módulo de Atención Convolucional)**: Actúa como un mecanismo de "atención selectiva", permitiendo que la red se concentre en las características más relevantes de la imagen (tanto a nivel de canales como de regiones espaciales) y ignore la información superflua.
- **Convoluciones Dilatadas**: Permiten a la red ampliar su "campo de visión" para capturar relaciones espaciales a mayor distancia, sin añadir una carga computacional significativa.

3.3 Impacto en los Resultados

Estudios aplicados a la estimación de pose en peces (*Oplegnathus punctatus*) han demostrado que una versión mejorada de HRNet (HPFPE) supera al modelo base:

Table 1: Comparación de precisión en estimación de pose

Métrica	**HRNet-W48 Base**	**HPFPE (Mejorado)**
Average Precision (AP)	72.84%	74.12%
AP_{75} (IoU$\geq$0.75)	80.74%	81.99%

Un incremento de +1.28 puntos porcentuales en la precisión promedio (AP) puede parecer modesto, pero en un entorno comercial, se traduce en aproximadamente 13 detecciones correctas adicionales por cada 1000 peces analizados. Esta mejora tiene un impacto directo en la rentabilidad cuando las mediciones se utilizan para la toma de decisiones económicas.

4 DeepLabCut: Seguimiento de Múltiples Individuos

4.1 Asociación de Partes mediante Campos de Afinidad

El principal desafío en el seguimiento de múltiples animales es mantener la identidad de cada individuo a lo largo del tiempo, especialmente cuando se cruzan y ocultan. DeepLab-

Cut aborda este problema mediante los **Part Affinity Fields (PAFs)**, una técnica que codifica las conexiones espaciales entre las distintas partes del cuerpo de un animal.

En lugar de simplemente detectar puntos clave (como la cabeza o la cola), los PAFs generan un campo vectorial que "une" las partes que pertenecen al mismo individuo. Matemáticamente, este campo se define como:

$$\mathbf{L}_{c,k}(\mathbf{p}) = \begin{cases} \frac{\mathbf{p}_2 - \mathbf{p}_1}{||\mathbf{p}_2 - \mathbf{p}_1||} & \text{si } d < \sigma \\ \mathbf{0} & \text{en otro caso} \end{cases}$$

donde $\mathbf{p}_1$ y $\mathbf{p}_2$ son las coordenadas de dos partes anatómicas, d es la distancia de un píxel a la línea que las une, y σ es un umbral de tolerancia.

En la práctica, esto permite al sistema resolver ambigüedades: incluso si dos peces se cruzan, los PAFs indicarán con alta probabilidad qué cabeza corresponde a qué cola, basándose en la coherencia de los campos vectoriales.

4.2 Re-Identificación tras Oclusiones

Para mantener la identidad de un pez después de que ha sido completamente ocultado, DeepLabCut extrae un "descriptor visual" único para cada individuo, una especie de firma basada en sus patrones de pigmentación, proporciones y texturas. Cuando un pez reaparece, su nuevo descriptor se compara con los almacenados en una memoria temporal mediante una métrica de similitud coseno:

$$sim(\mathbf{f}_{new}, \mathbf{f}_i) = \frac{\mathbf{f}_{new} \cdot \mathbf{f}_i}{||\mathbf{f}_{new}|| \cdot ||\mathbf{f}_i||}$$

De este modo, el sistema puede reasignar la identidad correcta, garantizando la continuidad del seguimiento a largo plazo.

5 PointRend: Segmentación de Alta Fidelidad

5.1 El Desafío de los Bordes

La segmentación precisa de los contornos de un objeto es un problema complejo para las redes neuronales. Debido a que operan en resoluciones reducidas, los métodos convencionales tienden a "suavizar" los bordes al momento de reconstruir la imagen a su tamaño original. Este efecto, producto de la interpolación bilineal, introduce errores sistemáticos que son inaceptables cuando se requieren mediciones milimétricas.

5.2 Un Enfoque Inspirado en Gráficos por Computadora

PointRend aborda este problema desde una perspectiva novedosa, inspirada en las técnicas de renderizado de gráficos por computadora. Su hipótesis es que no todos los píxeles de una imagen requieren el mismo nivel de atención: mientras que las regiones interiores y exteriores de un objeto son fáciles de clasificar, los bordes son inherentemente ambiguos y demandan un análisis más detallado.

El algoritmo de PointRend funciona en cuatro fases:

1. **Predicción inicial**: Se genera una máscara de segmentación de baja resolución, computacionalmente económica.

2. **Identificación de puntos inciertos**: Se seleccionan los píxeles del borde donde la red tiene mayor incertidumbre, utilizando la entropía de la predicción como métrica:

$$U(i) = -\sum_{c} p_i^{(c)} \log p_i^{(c)}$$

3. **Refinamiento localizado**: Para cada punto incierto, se extraen características de la imagen original a alta resolución y se utiliza una pequeña red neuronal (MLP) para recalcular la predicción con mayor precisión.

4. **Proceso iterativo**: Este proceso de selección y refinamiento se repite varias veces, aumentando progresivamente la resolución hasta alcanzar la calidad deseada.

5.3 Resultados en la Práctica

Table 2: Precisión en la detección de contornos

Métrica	**YOLOv8-Seg**	**PointRend**
Error promedio en borde (píxeles)	2.3	0.8
Precisión del contorno (IoU)	∼87%	∼89%
Latencia adicional (ms)	0	+20-50

Considerando una cámara que captura imágenes a una escala de 0.3 mm por píxel, la diferencia en el error de medición es notable:

- **Error con YOLOv8-Seg**: 2.3 píxeles × 0.3 mm/píxel = ±0.69 mm
- **Error con PointRend**: 0.8 píxeles × 0.3 mm/píxel = ±0.24 mm

Esta reducción del error es crucial para aplicaciones comerciales donde la clasificación por tamaño determina el valor del producto.

6 Síntesis y Recomendaciones

6.1 Criterios para la Selección del Modelo Adecuado

La elección de la arquitectura más apropiada depende de las prioridades de cada aplicación. La siguiente tabla resume los escenarios de uso para cada modelo:

6.2 Hacia una Arquitectura Híbrida

En un entorno de producción, la solución más eficaz es una arquitectura híbrida que combine las fortalezas de cada modelo. Se recomienda una implementación en dos fases:

Fase 1 (Implementación Base): Utilizar YOLOv8 para el monitoreo continuo, obteniendo mediciones con una precisión estándar (±5mm).

Fase 2 (Implementación Selectiva): Integrar los modelos especializados para activarlos según las condiciones: HRNet para mediciones de alta precisión, DeepLabCut en escenarios de alta densidad, y PointRend cuando la clasificación comercial exija una exactitud inferior a 2mm.

Table 3: Matriz de decisión por escenario operacional

Escenario	**Algoritmo Recomendado**	**Justificación**
Monitoreo general	YOLOv8	Balance óptimo entre velocidad y precisión
Mediciones de alta precisión	HRNet-W48	Preservación de detalles finos en multi-resolución
Alta densidad de población (¿5 peces/m^2)	DeepLabCut	Seguimiento robusto ante oclusiones severas
Contornos críticos (e.g., aletas)	PointRend	Refinamiento adaptativo de bordes

7 Conclusión

La elección de una arquitectura de visión artificial para la medición de salmones no se reduce a una única solución, sino a un balance estratégico entre velocidad, precisión y robustez. Mientras que YOLOv8 se destaca por su eficiencia en tiempo real, modelos como HRNet-W48, DeepLabCut y PointRend ofrecen herramientas especializadas para superar los desafíos de la estimación de pose de alta fidelidad, el seguimiento en condiciones de oclusión y la segmentación precisa de contornos.

El enfoque más prometedor para la industria acuícola reside en el desarrollo de sistemas híbridos e inteligentes, capaces de seleccionar dinámicamente el algoritmo más adecuado en función de las condiciones de cada captura y la criticidad de la medición requerida. Esta adaptabilidad es clave para optimizar tanto el monitoreo general de la biomasa como las decisiones comerciales que dependen de una precisión sub-milimétrica.

References

[1] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* https://doi.org/10.1109/CVPR.2019.00584

[2] Lauer, J., Zhou, M., Ye, S., et al. (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods, 19*(4), 496–504. https://doi.org/10.1038/s41592-022-01443-0

[3] Kirillov, A., Wu, Y., He, K., & Girshick, R. (2020). PointRend: Image Segmentation as Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* https://doi.org/10.1109/CVPR.2020.01152

[4] Gamero, E., Venegas, G., Ortega, M., Petit, L., & Castillo, A. (2025). *Documento de Análisis y Diseño de Salmo Metrics*. Procesamiento Digital de Imágenes [TEL328].

A Apéndice: Requerimientos de Hardware

Table 4: Hardware mínimo recomendado por tecnología

Tecnología	**GPU Mínima**	**FPS Esperado**
YOLOv8	Jetson Nano 4GB	15-20
HRNet-W48	Jetson TX2	8-12
DeepLabCut	Jetson AGX Xavier	5-8
PointRend	Jetson AGX Xavier	8-10