

Reidentificación de Personas Aplicando Aprendizaje Autosupervisado

Harold Canto
Juan Torres

Asesor: Edward Jorge Yuri Cayllahua Cahuina

Universidad de Ingeniería y Tecnología (UTEC)

2025

Contenido

1

Motivación

2

Problema Computacional

3

Justificación

4

Objetivos

5

Pipeline Metodológico

6

Dataset y Preprocesamiento

7

Resultados

8

Conclusiones y Recomendaciones

9

Trabajo Futuro

10

Referencias

Motivación

- Person Re-ID identifica personas entre cámaras sin depender del rostro.
- Supervisado: excelente rendimiento, alto costo de etiquetado.
- PersonViT: robusto a occlusiones pero muy costoso computacionalmente.
- ISR: autosupervisado con contraste inter-frames (sin etiquetas).
- Autosupervisado reduce costos y mejora generalización.

Fuente: Person Re-Identification - an overview | ScienceDirect Topics

Problema Computacional

¿Cómo lograr que el aprendizaje autosupervisado aprenda representaciones de identidad en datos no etiquetados que funcionen en distintas cámaras, condiciones de iluminación, pose y ocultación, alcanzando rendimiento cercano al supervisado pero con menor costo computacional?

Justificación

Técnica:

- El supervisado requiere grandes datasets etiquetados.
- Técnicas autosupervisadas han demostrado resultados competitivos.

Práctica:

- Reduce costos operativos.
- Escalable para vigilancia real en entornos urbanos.
- Menor dependencia de anotaciones manuales.

Objetivo General

Proponer un método autosupervisado para reidentificación de personas, evaluando su capacidad de reducir dependencia de datos etiquetados y mejorar la generalización en condiciones reales.

Objetivos Específicos

- Seleccionar datasets adecuados para entrenamiento y evaluación.
- Explorar técnicas autosupervisadas aplicables a Re-ID.
- Comparar enfoque autosupervisado vs supervisado.
- Definir métricas relevantes: Rank-1, mAP.
- Evaluar escalabilidad y desafíos prácticos.

Pipeline Metodológico



Fuente: Elaboración propia.

Motivación del Modelo Híbrido

- El aprendizaje autosupervisado captura patrones robustos sin necesidad de etiquetas.
- El aprendizaje supervisado refina las representaciones para la tarea específica de Re-ID.
- El modelo híbrido combina ambos enfoques para mejorar generalización y rendimiento.
- Esta estrategia reduce la dependencia de datos etiquetados sin sacrificar exactitud.

Dataset

CASIA-B ofrece una secuencia completa de imágenes para el estudio de *gait recognition*.

- 124 sujetos, 3 condiciones (normal, bolso, abrigo).
- 11 cámaras desde distintos ángulos.
- Secuencias de marcha ideales para SSL temporal.

Arquitectura General del Modelo Híbrido

Etapas del híbrido:

- ① Preentrenamiento SSL: el backbone aprende representaciones invariantes.
- ② Transferencia: se reutilizan pesos aprendidos como inicialización.
- ③ Fine-tuning Supervisado: se optimiza con Triplet Loss + softmax.

Componente autosupervisado (SSL)

- Basado en contraste temporal entre frames de la misma secuencia.
- Utiliza proyección SimCLR y pérdida InfoNCE.
- Aprende representaciones invariantes a iluminación y perspectiva.
- Actúa como base robusta para la etapa supervisada.

Componente supervisado

- Se añade una cabeza de clasificación sobre el embedding del SSL.
- Se usa Triplet Loss para maximizar separación entre identidades.
- Se emplea **P-K Sampler** para mejorar el aprendizaje métrico.
- Ajuste fino del backbone para adaptar las características al problema Re-ID.

Módulo de integración híbrida

El modelo híbrido combina:

- Representaciones autosupervisadas (*global invariances*).
- Ajuste supervisado para la identidad (*local discriminative features*).

Algunas ventajas del modelo híbrido:

- Mejora Rank-1 y mAP sin necesidad de más datos.
- Obtiene embeddings más consistentes entre cámaras distintas.

Resultados Cuantitativos

Modelo	Rank-1	mAP
Supervisado desde cero	62.1	55.3
SSL + Supervisado (propuesto)	91.29	44.83

Conclusiones

- El modelo híbrido alcanzó una precisión Rank-1 del 91.29%, validando que el preentrenamiento autosupervisado reduce eficazmente la dependencia de grandes volúmenes de datos etiquetados.
- La discrepancia observada en el mAP (44.83%) se atribuye a la pérdida de información dinámica en ángulos de visión extremos (0° y 180°), donde la silueta pierde discriminabilidad.
- El sistema demostró estabilidad algorítmica al escalar la población de prueba, manteniendo el rendimiento en el rango del 88-91% a pesar del incremento de distractores.
- Se confirmó que la fase autosupervisada actúa como un catalizador eficiente, aunque la inicialización de la capa de clasificación requiere estabilización para evitar óptimos locales.

Trabajo Futuro y Recomendaciones

- Se recomienda implementar *Stochastic Weight Averaging* (SWA) para mitigar la sensibilidad a la inicialización y reducir la varianza durante el ajuste fino.
- Es necesario integrar mecanismos de atención espacial (View-Aware) para ponderar dinámicamente las regiones visibles y mejorar la recuperación en vistas frontales.
- Se sugiere transitar hacia arquitecturas *Vision Transformers* (ViT) para capturar dependencias espacio-temporales globales superiores a las CNNs tradicionales.
- Se propone investigar la fusión multimodal con datos RGB o esqueletos para compensar la falta de información de la silueta en condiciones de oclusión severa.

Referencias

- C. Joshua et al., “Using optical flow consistency for self-supervised person Re-ID”, 2025.
- X. Liu et al., “UCM-VeID V2: Multi-View End-to-End Re-ID”, CVPR 2025.
- M. Varenyk et al., “Self-supervised low-FPS tracking for Re-ID”, 2025.
- H. Rao et al., “Self-supervised gait encoding for Re-ID”, IJCAI 2020.
- Z. Dou et al., “Identity-seeking self-supervised Re-ID”, ICCV 2023.

Gracias