

Apellidos: \_\_\_\_\_

Nombres: \_\_\_\_\_

Fecha: 18/02/2025

Nota:

--

Indicaciones:

La Duración es de **40 minutos**.

La evaluación consta de **10 preguntas**.

Las preguntas de opción múltiple valen 1.5 puntos.

1. En un proyecto de procesamiento de imágenes, Lauriel emplea DeepLab v3+. Para mejorar su capacidad de detección, Lauriel decide cambiar el backbone por uno basado en Transformers. Considerando que solo se podrán hacer cambios mínimos a DeepLab v3+ (cambiar el número de canales, por ejemplo), ¿cuál sería la mejor opción como backbone?
  - a) ViT.
  - b) CrossViT.
  - c) SwimTransformer.
  - d) Cualquiera.
  - e) Ninguna opción es compatible.
2. ¿Cuál de las siguientes alternativas explica mejor la diferencia entre BYOL y SimCLR en como previene el colapso?
  - a) BYOL utiliza un mecanismo de negative sampling similar al de SimCLR, pero con mayor cantidad de muestras negativas.
  - b) SimCLR depende del uso de numerosos pares negativos para forzar la diversidad en las representaciones, mientras que BYOL evita el colapso mediante una red target con gradientes detenidos, sin emplear muestras negativas.
  - c) SimCLR previene el colapso mediante un contraste explícito entre pares positivos y negativos, mientras que BYOL lo logra actualizando asimétricamente una red online y una target, prescindiendo de negativas.
  - d) La diferencia principal es que SimCLR adopta un enfoque simétrico de contraste, mientras que BYOL basa su mecanismo en auto-supervisión sin necesidad de augmentations.
  - e) Ambos evitan el colapso mediante la normalización de las representaciones, pero BYOL añade un loss basada en KL-divergence.
3. ¿Cuál de las siguientes afirmaciones refleja de manera más precisa el rol del teacher en Knowledge Distillation, DeiT y Noisy Student Training?
  - a) En todas las estrategias, el teacher actúa únicamente como inicializador del modelo student, sin influencia directa en la función de pérdida durante el entrenamiento.
  - b) Knowledge Distillation y Noisy Student Training dependen de un teacher preentrenado para transferir conocimiento, mientras que en DeiT el teacher se usa para generar pseudo-etiquetas en un entorno supervisado.
  - c) Knowledge Distillation utiliza un teacher preentrenado para transferir información mediante logits y características; Noisy Student Training mejora la robustez del student introduciendo ruido en las predicciones del teacher; y DeiT utiliza el teacher como guía para la atención.
  - d) El teacher en Noisy Student Training genera predicciones robustas en presencia de ruido, mientras que en Knowledge Distillation guía la optimización del student a través de logits suavizados, y en DeiT se implementa un mecanismo de atención basado en distillation.
  - e) DeiT se diferencia eliminando la necesidad de un teacher mediante técnicas de auto-regularización, mientras que los otros enfoques requieren una arquitectura teacher-student explícita.

4. Demuestre las relaciones entre  $H(x)$ ,  $H(x, y)$ ,  $H(x|y)$ ,  $I(x, y)$ ,  $D_{KL}(p(x)||p(y))$ .

[5 pts]

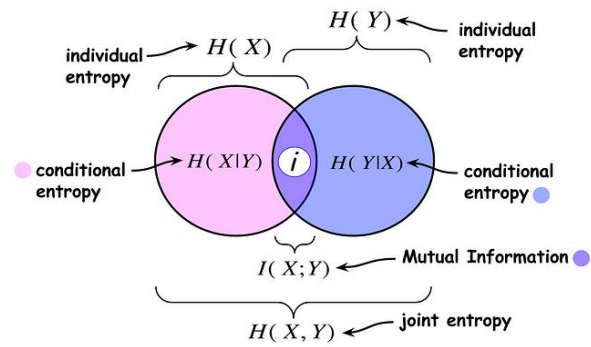
$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

$$H(X|Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x|y)$$

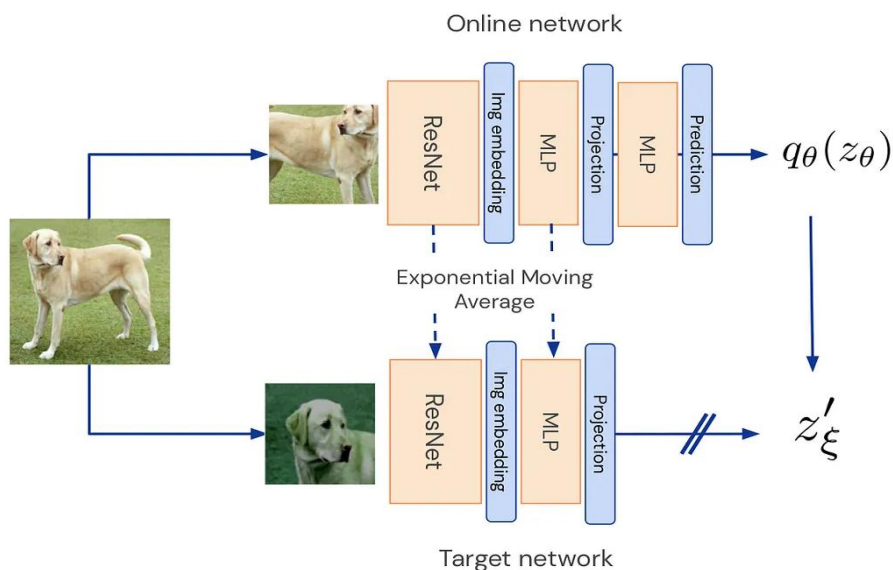
$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)q(y)}$$

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(y)}$$



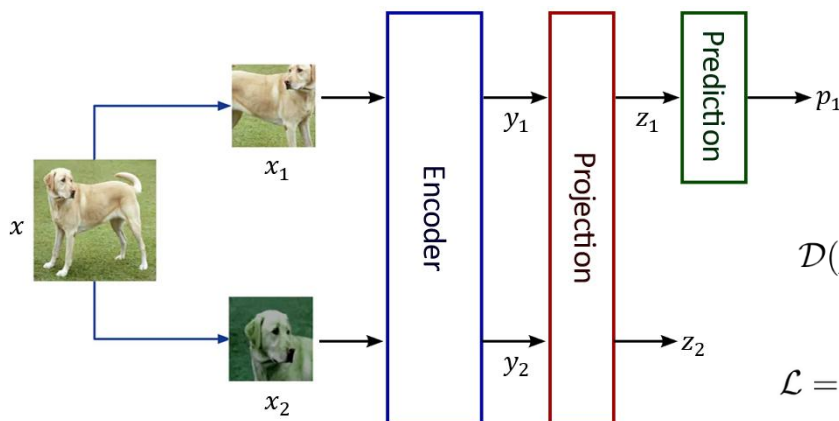
5. Crea un pseudocódigo para el entrenamiento de Bootstrap Your Own Latent (BYOL).  
Asuma que puede importar el backbone de su elección.

[5 pts]



$$\mathcal{L}_{\theta, \xi} \triangleq \|\overline{q_{\theta}}(z_{\theta}) - \overline{z'_{\xi}}\|_2^2 = 2 - 2 \cdot \frac{\langle q_{\theta}(z_{\theta}), z'_{\xi} \rangle}{\|q_{\theta}(z_{\theta})\|_2 \cdot \|z'_{\xi}\|_2}$$

6. Crea un pseudocódigo para el entrenamiento de SimSiam. [5 pts]  
 Asuma que puede importar el backbone de su elección.



$$\mathcal{D}(p_1, z_2) = -\frac{p_1}{\|p_1\|_2} \cdot \frac{z_2}{\|z_2\|_2}$$

$$\mathcal{L} = \frac{1}{2}\mathcal{D}(p_1, z_2) + \frac{1}{2}\mathcal{D}(p_2, z_1).$$

7. Crea un pseudocódigo para entrenar ResNet-34 con los datos ImageNet-21K mediante Knowledge distillation, donde el teacher es ViT. [5 pts]