

3 Modelos pre-entrenados: DINOv2

En esta sección se familiarizarán con la familia de modelos open source DINOv2¹ [9]. Este método de entrenamiento, desarrollado por Meta AI, ha dado lugar a modelos fundacionales de visión aplicables a una amplia variedad de tareas (clasificación, segmentación, estimación de profundidad, entre otras) y que logran generalizar a distribuciones de datos significativamente distintas a las utilizadas durante el entrenamiento, sin necesidad de realizar fine-tuning.

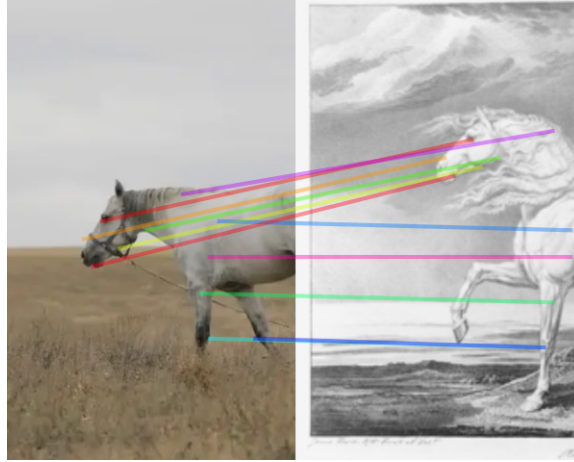


Figura 2: Ejemplo de sparse matching usando DINOv2.

Actividad 8

Gran parte de la eficacia de DINOv2 se debe a la alta calidad de los datos de entrenamiento. En el artículo donde se introduce el método, se mencionan varias medidas tomadas para mejorar la calidad de los datos utilizados.

Investigue y explique cómo los autores “limpian” sus datos de preentrenamiento.

Actividad 9

Investigue y explique el concepto de sparse matching.

Utilice el modelo DINOv2 para extraer descriptores de parches de dos imágenes de un mismo animal. Luego, visualice al menos 10 coincidencias usando líneas entre puntos correspondientes, utilizando el código entregado. Haga el mismo experimento rotando una de las imágenes.

¿Qué observa en la precisión del máatching luego de rotar?

Actividad 10

Cree un MLP simple utilizando PyTorch para clasificar imágenes del conjunto de datos CIFAR-10 [5]. Compare los resultados de entrenamiento utilizando directamente los píxeles de las imágenes como entrada, versus el uso de features extraídas con DINOv2 (utilice el código de extracción de features presente en el notebook).

¹Para ejemplos, casos de uso y demostraciones, puede visitar: <https://dinov2.metademolab.com/>