

NOMBRE: Benjamín Farías Valdés

N.ALUMNO: 22102671



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC3692 — Tópicos Avanzados en Inteligencia Artificial — 2' 2022

Lectura 15

Crítica

A Simple Framework for Contrastive Learning of Visual Representations

El paper presenta un *framework* para aprender a representar imágenes, utilizando *contrastive learning* como estrategia de entrenamiento.

Este tipo de modelos me recuerda a las redes siamesas, que justamente hacen uso de un mecanismo que contrasta las representaciones de dos entradas. De esta forma la red aprende a juntar en su espacio de características a estos pares de ejemplos que se refieren a lo mismo. La función de pérdida utilizada me parece adecuada, dado que se basa en entropía cruzada (que es ampliamente utilizada para estos problemas), pero agregando hiper-parámetros que permiten ajustar mejor la medida de similitud entre ejemplos.

Un componente sumamente importante de este trabajo es el aumento de datos seleccionado. Encuentro inteligente la decisión de abarcar el problema desde un enfoque más simple, lo que le permitió a los autores darse cuenta de que usando técnicas típicas de *cropping* y *resizing* es posible emular trabajos previos que utilizaban arquitecturas especializadas para entrenar ciertas habilidades, tales como reconocer partes locales y globales de una misma imagen. La gracia es que ahora es posible aprender estas mismas habilidades utilizando cualquier arquitectura, ya que surgen de forma natural al aumentar los datos con *cropping*.

Otro aspecto a destacar en el aumento de datos fue la necesidad de componer distintas transformaciones geométricas con las de cambio de color, dado que la uniformidad de algunas imágenes puede causar que la red aprenda a reconocerlas simplemente mirando los colores e ignorando el contenido geométrico. Este tipo de problemas se han visto anteriormente en literatura de *shortcut learning*, y en este caso es posible evitarlos directamente al forzar a la red a fijarse tanto en la geometría como en los colores con un mismo ejemplo aumentado.

Un último aspecto que encontré interesante respecto al aumento de datos fue el hecho de que, según los experimentos, la estrategia de *contrastive learning* aprovecha mejor el aumento de datos en comparación al enfoque supervisado clásico. En mi opinión esto es un resultado importante, dado que implica que ciertas técnicas de aumento que no son útiles en estrategias supervisadas de entrenamiento, ahora podrían ser perfectamente viables al cambiar a aprendizaje contrastivo.

Respecto a la comparación con trabajos relacionados, es claro que la propuesta tiene gran valor, dado que en muchos *datasets* supera o es competitiva con el estado del arte, en donde este último pertenecía

a redes entrenadas de forma completamente supervisada. Esto nos dice que, diseñando un buen aumento de datos y estrategia de entrenamiento, es posible competir con modelos supervisados utilizando enfoques completamente automáticos, los que son mucho más eficientes al no depender de humanos.

Finalmente, lo que me gustaría rescatar de este trabajo es la idea de que mejorar arquitecturas no es necesariamente la mejor forma de avanzar siempre en esta área de investigación, en ocasiones la mayor diferencia podría ocurrir debido a un cambio en los enfoques de entrenamiento, lo que va de la mano con el término de *coaching* mencionado en clases (que se está popularizando hoy en día).