

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 1

Crítica

Deep Visual-Semantic Alignments for Generating Image Descriptions

El paper está muy bien estructurado y completo. Pensando en el contexto del año 2015, donde hace poco había ocurrido el resurgimiento de las redes neuronales gracias a la aparición de *GPUs* potentes y modelos de *Deep Learning*, resulta de gran valor lo presentado en el artículo, referente a la generación de descripciones textuales sobre imágenes. Se presentan de forma ordenada los trabajos que componían el estado del arte en el área (principalmente clasificadores de imágenes o generadores de descripciones simples basadas en categorías pre-existentes para las imágenes). Al entrar al primer modelo propuesto, a pesar de que se apoya en varias formalizaciones matemáticas que requieren revisión extensiva para comprender en su totalidad (para alguien que no es un experto en el área), sí lo encontré bastante claro respecto a expresar las ideas e intuiciones detrás de su funcionamiento. Me pareció muy interesante la idea de combinar la información visual y semántica mediante un *mapping* a un espacio multidimensional común, para lograr así la alineación entre fragmentos de texto y regiones de la imagen. Por otro lado, también presentan buenos resultados en el segundo modelo, que se encarga de generar las descripciones a partir de imágenes. Existe una buena aplicación del primer modelo para generar datos que alimenten al segundo, y se propone una mejora a futuro que consiste en intentar unificar ambos modelos en una sola red neuronal compuesta. Si tuviera que cambiar algo sobre el paper, sería el disminuir un poco la cantidad de sub-secciones dentro de algunos de los capítulos, ya que a veces sentía que cambiaba entre distintos temas demasiado rápido, a pesar de que realmente hablaban de cosas similares (que podrían estar conjuntas o resumidas de forma más ordenada, principalmente en el capítulo de resultados). Finalmente, cabe mencionar que una gran cantidad de resultados fueron reservados para el apéndice del artículo, dejando así más espacio para explicar los puntos clave (lo que me parece una excelente idea).