Aprendizaje Profundo (DL) – Semestre: Primavera 2020
Instituto Tecnológico Autónomo de México, ITAM.

Proyecto Final: Text-based image generation

Objetivo:
Poner en práctica las técnicas estudiadas durante el curso, combinando tantas como sean necesarias.

Reto:
Crear modelos capaces de generar imágenes a partir de entradas en forma de texto. Proponer soluciones "interesantes" dadas las limitaciones de tiempo, datos y hardware.

El modelo debe ser robusto en términos de generación de: 1) imágenes de buena calidad; 2) imágenes consistentes (contenido visual acorde con la descripción textual); y 3) espacio latente apropiado (explicado a continuación). Adicionalmente, el modelo debe ser eficiente en términos de su tamaño en memoria (número de parámetros) y tiempo requerido para predecir (el tiempo de entrenamiento también debe ser reportado).

Espacio latente: La definición de un espacio latente apropiado es ambigua, y depende de la tarea a resolver. En algunos casos queremos que sean ralos y en otros no, en algunos casos queremos que sean *over* o *under-complete*, en algunos casos queremos que nos permitan muestrear, y en algunos casos queremos que sus componentes tengan distribuciones de probabilidad específicas, etc. Ustedes deberán elegir las características que requiera su espacio latente y justificarlas, así como reportar en qué grado lograron obtenerlas y cuáles fueron las limitaciones.

Datos:
Usaremos los datos de *Flick8k dataset*[1]. Este data set fue creado para la tarea contraría[2] de la que estamos resolviendo, es decir, para generar etiquetas de texto a partir de información visual. Los datos pueden ser obtenidos desde la siguientes ligas:
https://github.com/jbrownlee/Datasets/releases/download/Flickr8k/Flickr8k_Dataset.zip
https://github.com/jbrownlee/Datasets/releases/download/Flickr8k/Flickr8k_text.zip

El dataset consta de más de 8,000 imágenes de personas realizando actividades, y de 5 descripciones en formato texto para cada imagen. Es decir, la misma salida tendrá hasta 5 posibles entradas. En total existen 40,455 pares (x=texto, y=imagen), aunque físicamente sólo se proporcionan 8,091 imágenes. Usteden deberán diseñar la forma de desenvolver las asignaciones para generar los 40,455 pares 1-a-1. Pueden usar el codigo de python que les adjunto en un notebook como punto de partida.

Organización de los datos: dividan los datos en 80% para entrenar y 20% para prueba final (test). Concrétamente, por cada imágen, seleccionen 4 de sus correspondientes textos para entrenar, y dejen uno para probar. La división para validar la dejo a su elección, pero debe haber alguna.

Resolución: Las imágenes tienen tamaños distintos, ustedes deberán idea alguna manera para dárselas a la red. Más abajo dejo unas notas que, en combinación con el notebook, puede ayudarles. Sugiero que las reescalen a 64x64 pixeles (llamémoslo resolución *stage I*[3]) para que el procesamiento no sea tan pesado, a costa de perder calidad en la resolución de salida. Llamemos *stage II* = 256x256 pixeles.

1 https://www.ijcai.org/Proceedings/15/Papers/593.pdf
2 https://data-flair.training/blogs/python-based-project-image-caption-generator-cnn/
3 https://medium.com/@mrgarg.rajat/implementing-stackgan-using-keras-a0a1b381125e

Evaluación:
Ya que este es un problema complejo, y tanto los datos como el tiempo y los recursos son limitados, no se espera que lo resuelvan como lo hubiera hecho Yahoo o Google. Es decir, se harán las consideraciones pertinentes en la evaluación, principalmente ante evidencia cuantitativa (listado de recursos de cómputo, tiempos de evaluación, etc.) que justifique las limitaciones y los resultados obtenidos. Sin embargo, para efectos de estas consideraciones se excluyen justificaciones de falta de tiempo, falta de hardware, enfermedad (salvo casos excepcionales), mala elección de compañeros de equipo, y principalmente reportes con resultados que parecieran carecer de toda estructura.

Concretamente, la evaluación considerará:

- 20% Representación latente apropiada (justificada).
- 20% Generación de imágenes (*test*) de calidad aceptable (comparables con *stage I*).
- 20% Generación sintética de imágenes (*random*) de calidad aceptable (*stage I*).
- 30% Calidad del reporte y presentación en clase.
- 10% Voto anónimo de los compañeros de clase.
- Existe la posibilidad de obtener 10% extra si las imágenes generadas tienen calidad comparable con *stage II*.

Entrega:
Por equipos de mínimo 2 personas y máximo 3. Deberán reportarme, a más tardar el día 7 de mayo, los nombres de los integrantes de cada equipo por correo electrónico.

Deadline: Reporte escrito – jueves 14 de mayo, 4:00 pm.
Presentación en clase – jueves 14 de mayo, 4:00-7:00 pm.

Reporte:
El reporte deberá ser enviado por correo electrónico, en **formato PDF**. Deberá tener de 6 a 8 páginas (páginas extras son válidas como apéndices), e incluir imágenes y tablas que ayuden a validar los resultados. Mirar detalles de contenido en la siguiente página.

Notas:
- Debido a que las imágenes pueden variar en tamaño y orientación (*portrait* vs *landscape*), se sugiere inscribir cada una dentro de una imágene genérica de 256x256 pixeles negros, para así procesarlas todas con el mismo tamaño (ver notebook).
- 256x256 pixeles corresponde a una calidad de visualización deseable (stage II). Sin embargo, dadas ciertas limitaciones potenciales de procesamiento, se considerará aceptable el tamaño de 64x64 pixeles (stage I).
- Se provee un notebook de Jupyter que puede ser usado para dar ese formato a las imágenes.
- Si se ocupa el notebook proporcionado, notarán que se generan 40,455 pares de descripciones-imágenes (imágenes repetidas). Y que 5 descripciones no contienen imágenes asociadas, las cuales simplemente descartaremos.

Referencias

- https://www.ijcai.org/Proceedings/15/Papers/593.pdf
- https://arxiv.org/pdf/1612.03242.pdf
- https://arxiv.org/abs/1809.01110
- https://vision.ece.vt.edu/clipart/

Título (Creen un título)
Autores: Nombres, CU’s y afiliaciones (programa).

Resumen
En un párrafo, de 6 a 10 líneas, mencionar el problema, indicar la solución propuesta, y hacer un comentario breve de la calidad de los resultados que se obtuvieron con su propuesta.

1. Introducción (expandir el resumen)
   Mencionar en qué consiste el problema.
   Qué tipo de soluciones existen actualmente. Qué nivel de calidad obtiene, cómo lo obtienen, cuáles son sus limitaciones.
   Qué proponen ustedes y por qué. Qué nivel de calidad obtienen ustedes.
   Organización del resto del documento.

2. Descripción detallada del problema
   Incluir detalles de la base de datos.
   Qué retos encuentran en cuanto a los datos, diseño de la solución, técnicos, de implementación.

3. Descripción detallada de la solución presentada
   Describan su mejor método, ese será el que proponen.
   La descripción puede incluir intuiciones, así como motivaciones y detalles técnicos.
   Justifiquen las decisiones de su diseño.
   Apoyense de diagramas, tablas, etc.
   Pueden mencionar, algunas otras propuestas evaluadas que no hayan sido tan exitosas como la final, pero sólo a manera de mención. No lo cuenten como historia progresiva.

4. Experimentos y resultados
   Presenten tablas, curvas, ejemplos, etc.
   Comparen valores de parámetros y variantes de decisiones que hayan tomado.

5. Conclusiones
   Un resumen (recap) breve de lo que presentaron y los resultados obtenidos.
   Incluyan lecciones aprendidas.
   Incluyan limitaciones actuales y posibles experimentos futuros (no que los deban hacer, sino qué más se podría intentar para mejorar los resultados actuales).

Referencias
Listen las referencias que haya usado para documentarse, y que tengan cita en el documento.