

Aprendizaje Automático para Datos en Grafos
Propuesta de Proyecto Final
Redes Neuronales en Grafos para visión por computadora

Graciana Castro
4.808.848-2
gcastro@fing.edu.uy

Julian O’Flaherty
6.285.986-9
julian.o.flaherty@fing.edu.uy

1. Introducción

Las imágenes son un tipo de datos con estructura espacial, la cual podemos representar con un grafo. Por ejemplo, existen métodos clásicos de segmentación de imágenes que se basan en construir un grafo y aplicarle algoritmos para obtener el corte mínimo del mismo [6]. Dados los avances en el aprendizaje profundo y los desarrollos en las redes neuronales en grafos, es natural postular la utilidad de estas para tareas de visión por computadora. En este trabajo se realizará un pequeño relevamiento del estado del arte del uso de GNNs para tareas de visión por computadora, enfocándose en tareas de clasificación y segmentación de imágenes.

El objetivo de este proyecto será responder la pregunta **¿Tienen lugar las GNNs en la tarea de Visión por computadora?**

2. Propuesta de trabajo

2.1. Relevamiento del estado del arte

La primera etapa del proyecto consistirá en un relevamiento del estado del arte, analizando trabajos que hayan explorado esta idea. Por lo que hallamos preliminarmente, el desafío se divide en dos etapas:

- Diseño y construcción del grafo
- Arquitectura y entrenamiento de la GNN

En el relevamiento se apuntará a cubrir varias estrategias de construcción del grafo y arquitecturas, buscando analizar que combinación obtiene los mejores resultados.

Entre los trabajos que encontramos, destacan los siguientes:

- **VisionGNN** [4]: este paper propone una estrategia de creación de grafo tomando patches de la imagen como nodos y conectándolos en base a su contenido.
- **Superpixel Image Classification with Graph Attention Networks** [1]: este paper propone una estrategia de creación de grafo tomando superpixels de la imagen como nodos y definiendo una estrategia de conexión basada en el contenido de los mismos.

2.2. Implementación y datasets

Una vez seleccionados los modelos y estrategias a usar, se recopilarán las implementaciones en un repositorio que permita la rápida evaluación de los métodos. Se desarrollará en Python, utilizando la librería torch-geometric para los modelos de aprendizaje profundo. El repositorio

se implementará tal que los bloques de creación de grafos y los modelos sean intercambiables y combinables, permitiendo la rápida iteración y comparación.

Para trabajar con un nivel de computo manejable, se trabajara sobre los siguientes datasets:

- MNIST [?]: dataset de dígitos escritos a mano. Actuara como dataset de debugging, contando con imágenes pequeñas y simples de clasificar, permitiendo la rápida iteración y validación de modelos.
- STL-10 [2]: dataset de imágenes de objetos de 10 clases (500 imagenes de entrenamiento por clase).
- ImageNet-100 [5]: subset de ImageNet con 100 clases (en vez de las 10k).
- Pascal VOC0712 [3]: dataset de clasificación, detection y segmentación de objetos. Cuenta con 20 clases.

3. Entregables

El entregable del proyecto será un informe con los resultados obtenidos, haciendo una comparativa de los resultados obtenidos con cada estrategia. Se sumará a la comparación, resultados obtenidos para CNNs del mismo tamaño¹. También se hara una comparativa con la construcción del grafo básica, conectando simplemente patches vecinos, buscando medir el impacto de la construcción del grafo en el modelo final. Se hara también un analisis computacional, tratando de medir la escalabilidad de estos métodos.

También se entregará el código de investigación utilizado, donde se implementen los modelos y estrategias de construcción de grafo.

Referencias

- [1] Pedro H. C. Avelar, Anderson R. Tavares, Thiago L. T. da Silveira, Cláudio R. Jung, and Luís C. Lamb. Superpixel image classification with graph attention networks, 2020.
- [2] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 215–223, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.
- [3] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge, 2010.
- [4] Kai Han, Yunhe Wang, Jianyuan Guo, Yehui Tang, and Enhua Wu. Vision gnn: An image is worth graph of nodes, 2022.
- [5] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

¹Se excluyen los VITs, dado que su arquitectura destaca en grandes volúmenes de datos

- [6] Wikipedia contributors. Graph cuts in computer vision — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Graph_cuts_in_computer_vision, 2025. [Online; accessed 21-November-2025].