**A Survey on Graph Neural Networks and Graph Transformers in Computer Vision: A Task-Oriented Perspective**

Graph Neural Networks (GNNs) have accelerated the state of the art in a field called graph representation learning and a wide range of fields, including natural language processing (e.g., relation extraction and sequence learning), computer vision (e.g., object identification and point cloud learning), and data mining (e.g., social network analysis and recommender systems), to mention a few. Using the Emergence of Transformers in Computer Vision and Natural Language Processing, Graph Transformers include a graph structure into its architecture in order to get over the drawbacks of local neighbourhood aggregation and stay away from rigid structural inductive biases. In this study, we provide a thorough task-oriented evaluation of GNNs and graph Transformers in computer vision.

According to the type of input data, we specifically classify their computer vision applications into five categories: 2D natural photos, videos, 3D data, vision + language, and medical images. We then segment the applications under each category into several other vision-related tasks. We may investigate how each task is handled by various GNN-based approaches and how effective these approaches are using such a task-oriented taxonomy. We outline the definitions and difficulties of the tasks, in-depth coverage of the representative techniques, debates regarding insights, limitations, and prospects, all based on the necessary preliminaries.

Despite the revolutionary advances in perception, giving deep learning models the ability to reason still poses a significant barrier for contemporary computer vision systems. Regarding this, GNN and graph transformers have proven to be significantly more adaptable and superior in handling'relational' jobs. In order to achieve this, we have provided the first thorough examination of GNN and graph Transformers in computer vision from a task-oriented viewpoint. According to the type of input data, such as picture, video, and point cloud, a range of traditional and modern algorithms are divided into five groups. We expect that by methodically classifying the approaches for each work, this study will reveal further advancement in the future. By outlining significant innovations, restrictions, and prospective future research areas,