Superpixel-based Domain-Knowledge Infusion in Computer Vision

Gunjan Chhablani[1] and Abheesht Sharma[1] and Harshit Pandey[3] and Tirtharaj Dash[1,2]

1- Dept. of CS & IS, BITS Pilani, Goa Campus, Goa, India
2- APPCAIR, BITS Pilani, Goa Campus, Goa, India
3- Dept. of CS, Savitribai Phule Pune University, Pune, India
(First 3 authors: Equal contribution; Random order)

Abstract. Superpixels are higher-order perceptual groups of pixels in an image, often carrying much more information than raw pixels. There is an inherent relational structure to the relationship among different superpixels of an image. This relational information can convey some form of domain information about the image, e.g. relationship between superpixels representing two eyes in a cat image. Our interest in this paper is to construct computer vision models, specifically those based on Deep Neural Networks (DNNs) to incorporate these superpixels information. We propose a methodology to construct a hybrid model that leverages (a) Convolutional Neural Network (CNN) to deal with spatial information in an image, and (b) Graph Neural Network (GNN) to deal with relational superpixel information in the image. The proposed deep model is learned using a generic hybrid loss function that we call a 'hybrid' loss. We evaluate the predictive performance of our proposed hybrid vision model on four popular image classification datasets: MNIST, FMNIST, CIFAR-10 and CIFAR-100. Moreover, we evaluate our method on three real-world classification tasks: COVID-19 X-Ray Detection, LFW Face Recognition, and SOCOFing Fingerprint Identification. The results demonstrate that the relational superpixel information provided via a GNN could improve the performance of standard CNN-based vision systems.

1 Introduction

Deep learning, especially in the task of image classification and recognition has taken centerstage, mostly due to the introduction of ILSVRC challenge[1]. There have been significant architectural innovations for convolutional neural networks (CNNs). In the last few years, the core approach of basic convolution has been adopted to graph-structured data via the introduction of graph neural networks (GNNs)

A graph is a representation of binary relations and GNNs can deal with these relational data. In the case of tasks involving images, binary relations can be easily seen at the level of 'image superpixels'. Superpixels are a higher-order perceptual groups of pixels in an image, often conveying much more information than low-level raw pixels and sharing some common characteristics such as intensity levels.

---

[1] https://image-net.org/challenges/LSVRC/