CS – 549

Paper Review – Learning to Compose Domain-Specific Transformations for Data Augmentation

Akhil Sai Chintala (934409906)

- **What is the problem discussed in the paper?**
  Data augmentation is a method that is used for increasing the size of labeled training sets by leveraging task-specific data transformations that preserve class labels. While it is easy to specify individual transformations, construction and tuning the more sophisticated compositions typically need to achieve state-of-the-art results is taking a lot of time than expected.

- **Why is it important?**
  Data augmentation has quickly become a critical and effective tool for combatting this labelled data scarcity problem. This can be seen as a form of weak supervision, providing a way for practitioners to leverage their knowledge of invariances in a task or domain. Hence, it is important to directly and flexibly leverage domain experts knowledge of invariances as a valuable form of weak supervision in real-world situations where training data is limited.

- **What are the main ideas of the proposed solution for the problem?**
  The authors have proposed a new idea for the problem by creating a new method for data augmentation that directly leverages user domain knowledge in the form of transformation operations, and automates the difficult process of composing and parameterizing them.

- **What are the shortcomings of the proposed solution?**
  In the paper, for the training sets with labelled data, sometimes but not more often as the percentage of labelled training data increases for few datasets, the best performance is not shown compared to the other sets with different amount of data.