Efficient Transformers for Long-Range Dependencies

Abstract. A novel attention mechanism that scales linearly with sequence length.

1. Introduction

Transformers have revolutionized NLP.

2. Method

Our approach combines local and global attention.

3. Conclusion

Efficient transformer architecture demonstrated.