Discussion Questions

1. How do you rate the paper given its limited technical contribution?

2. What is the factor causing the biggest increase in performance?

3. Are the inductive biases of CNNs beneficial or harmful?

4. Given the results from the last ~4 papers + today's paper, is self-attention truly needed?

5. If ConvNeXts perform just as well as Vision Transformers, are there any advantages of using Vision Transformers?

6. Where do we go from here? Are we going to be using Transformers for visual recognition tasks in 2, 5, 10 years?

Discussion Questions

1. How do you rate the paper given its limited technical contribution?

Discussion Questions

2. What is the factor causing the biggest increase in performance?

Discussion Questions

3. Are the inductive biases of CNNs beneficial or harmful?

Discussion Questions

4. Given the results from the last ~4 papers + today's paper, is self-attention truly needed?

Discussion Questions

5. If ConvNeXts perform just as well as Vision Transformers, are there any advantages of using Vision Transformers?

Discussion Questions

6. Where do we go from here? Are we going to be using Transformers for visual recognition tasks in 2, 5, 10 years?

Key Takeaways

Transformers:

- Attention is NOT all you need (but it can still be useful).
- The generality of Transformer architecture is helpful when considering multimodal data.
- Inductive biases of CNNs might not be as harmful as previously claimed (even in big-data regimes), and they might even benefit Transformers (e.g., Swin).

General:

- Don’t give in to the hype but instead critically evaluate each paper based on the empirical evidence.
- Pay attention to hidden implementation details (e.g., optimization, training schedule, data augmentation, etc.).
- Learn to appreciate simple yet effective ideas.
- Focus on the big picture (e.g., potential impact of the paper).