Discussion Questions

1. Is Swin capable of learning global long-range dependencies in the visual data like ViT does?

2. Can we conclude that the inductive biases of CNNs adopted by Swin transformer (locality, translation invariance, etc.) are fundamentally needed for vision tasks?

3. Do you agree with the authors on why previous visual models struggle with scalability?

4. What are the advantages and disadvantages of hierarchical feature learning using local attention?

5. The authors claim that their architecture unifies CV and NLP. How would we use Swin to jointly model visual and textual signals?

6. What would you say is the biggest contribution of this paper? How could the paper be improved?

7. Why does Swin work better on medium-sized datasets than the original ViT?

8. Can we conclude the proposed Swin architecture is better than CNNs in terms of performance? efficiency?

9. Would it be easier to scale Swin transformer to 22B parameters than the standard ViT?

10. What could be the potential value of incorporating dynamic window sizes into the Swin Transformer architecture?

Discussion Questions

1. Is Swin capable of learning global long-range dependencies in the visual data like ViT does?

Discussion Questions

2. Can we conclude that the inductive biases of CNNs adopted by Swin transformer (locality, translation invariance, etc.) are fundamentally needed for vision tasks?

Discussion Questions

3. Do you agree with the authors’ on why previous visual models struggle with scalability?

Discussion Questions

4. What are the advantages and disadvantages of hierarchical feature learning design using local attention?

Discussion Questions

5. The authors claim that their architecture unifies CV and NLP. How would we use Swin to jointly model visual and textual signals?

Discussion Questions

6. What would you say is the biggest contribution of this paper? How could the paper be improved?

Discussion Questions

7. Why does Swin work better on medium-sized datasets than the original ViT?

Discussion Questions

8. Can we conclude the proposed Swin architecture is better than CNNs in terms of performance? efficiency?

Discussion Questions

9. Would it be easier to scale Swin transformer to 22B parameters than the standard ViT?

Discussion Questions

10. What could be the potential value of incorporating dynamic window sizes into the Swin Transformer architecture?