Discussion Questions

1. Can TimeSformer recognize actions that involve fast-moving objects?

2. Why does TimeSformer struggle with temporally-heavy datasets such as SSv2?

3. Why does divided space-time attention perform better than joint attention?

4. Should video transformers replace 2D/3D CNNs or are there still some advantages of CNNs? Is it useful to combine CNNs and Transformers?

5. Pros/Cons of CNNs vs. Transformers for multi-modal learning?

6. What is the biggest advantage of video transformers?

7. What is the biggest disadvantage of video transformers?

8. Does the emerging trend of transformers widen the gap between academia and industry If so, what is the path forward for academia?

9. Is (space-time) attention all you need?

10. Are we going to be using Transformers in 2, 5, 10 years or is it going to be some other type of model?

Discussion Questions

1. Can TimeSformer recognize actions that involve fast-moving objects?

Discussion Questions

2. Why does TimeSformer struggle with temporally-heavy datasets such as SSv2?

Discussion Questions

3. Why does divided space-time attention perform better than joint attention?

Discussion Questions

4. Should video transformers replace 2D/3D CNNs or are there still some advantages of CNNs? Is it useful to combine CNNs and Transformers?

Discussion Questions

5. Pros/Cons of CNNs vs. Transformers for multi-modal learning?

Discussion Questions

6. What is the biggest advantage of video transformers?

Discussion Questions

7. What is the biggest disadvantage of video transformers?

Discussion Questions

8. Does the emerging trend of transformers widen the gap between academia and industry If so, what is the path forward for academia?

Discussion Questions

9. Is (space-time) attention all you need?

Discussion Questions

10. Are we going to be using Transformers in 2, 5, 10 years or is it going to be some other type of model?