Discussion Questions

1. Conceptually, why does masked auto-encoding work well for visual representation learning? Is random masking the best strategy?

2. How does MAE compare to other SSL methods (e.g. DINO) in terms of semantic understanding? Which paradigm should be used moving forward?

3. Compared to BERT in NLP that works best with ~15% masking ratio, why does MAE work well with very high mask ratios (e.g., ~75%)? How would varying the mask ratio size impact the performance of the model?

4. Does MAE scale to large models/datasets better than prior contrastive SSL approaches?

5. Can this method be applied to models with inductive biases (CNNs, Swin)?

6. Are the comparisons to other approaches fair considering that MAE was trained for 1600 epochs?

7. Why is the decoder design important for learning the semantic level of the learned latent representations?

8. Will self-supervised learning methods surpass supervised approaches in CV in the future?

9. How can we extend this framework to other modalities (video, audio, etc.)?

Discussion Questions

1. Conceptually, why does masked auto-encoding work well for visual representation learning? Is random masking the best strategy?

Discussion Questions

2. How does MAE compare to other SSL methods (e.g. DINO) in terms of semantic understanding? Which paradigm should be used moving forward?

Discussion Questions

3. Compared to BERT in NLP that works best with ~15% masking ratio, why does MAE work well with very high mask ratios (e.g., ~75%)? How would varying the mask ratio size impact the performance of the model?

Discussion Questions

4. Does MAE scale to large models/datasets better than prior contrastive SSL approaches?

Discussion Questions

5. Can this method be applied to models with inductive biases (CNNs, Swin)?

Discussion Questions

6. Are the comparisons to other approaches fair considering that MAE was trained for 1600 epochs?

Discussion Questions

7. Why is the decoder design important for learning the semantic level of the learned latent representations?

Discussion Questions

8. Will self-supervised learning methods surpass supervised approaches in CV in the future?

Discussion Questions

9. How can we extend this framework to other modalities (video, audio, etc.)?