Residuals and Normalizations - Summary

Let's train really deep networks

- Vanishing gradients and activations are normal
  - Better than explosions
- Residual connections and normalization deal with vanishing gradients