

Attention Augmented Conv. Net

Summary

1. Argues that conv. operator is limited because it does not capture global context
2. Self-attention allows for capturing context without increasing the no. of parameters.
3. They introduce a new "two-dimensional relative self-attention" mechanism that prove competitive in replacing convolutions as a stand-alone component for image classification.
4. Conv + self-attention performs best.