Spatially Adaptive Computation Time for Residual Networks

Michael Figurnov¹  Maxwell D. Collins²  Yukun Zhu²  Li Zhang²
Jonathan Huang³  Dmitry Vetrov²,³  Ruslan Salakhutdinov⁴
¹National Research University Higher School of Economics  ²Google Inc.  ³Yandex  ⁴Carnegie Mellon University

Motivation
Do we really need to put these blue pixels through 100 layers to detect the bird?

Can we make networks faster and more interpretable by adapting the amount of computation?

Contribution
A novel mechanism that
- adapts the computation spatially
- is end-to-end trainable
- provides introspection
- is problem-agnostic

Spatially Adaptive Computation Time (SACT)
Apply ACT "convolutionally" to every spatial position of residual network’s block

Residual Network (ResNet)
image conv+pool residual units pool + fc
block 1 block 2 block 3 block 4

Adaptive Computation Time (ACT)

Input: halting score

activation
block of residual units

Output: remainder

Spatially Adaptive Computation Time (SACT)

Residual units are evaluated only in the active positions!

Image classification (ImageNet dataset)

Object detection (COCO dataset)

Visual saliency (cat2000 dataset)

Code: github.com/mfigurnov/sact