# Feature-Dependent Cross-Connections in Multi-Path Neural Networks

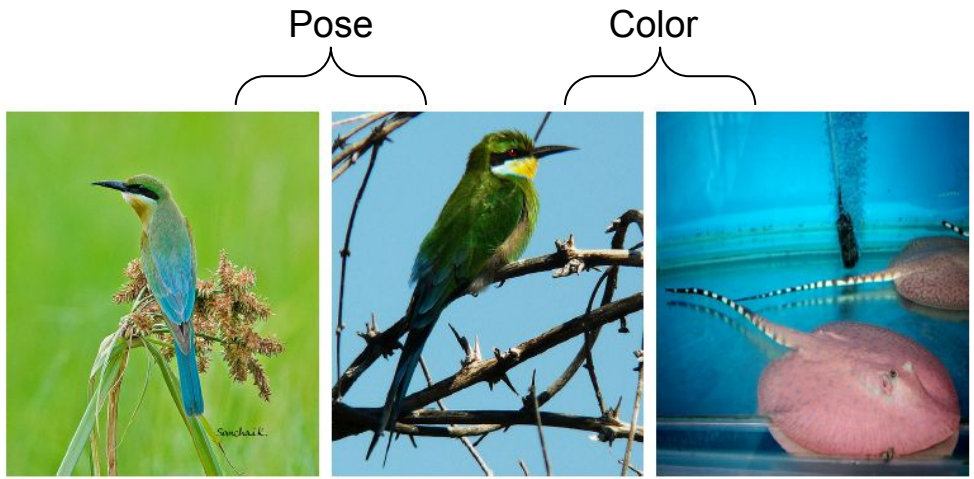
**Dumindu Tissera**[*+], **Kasun Vithanage**[*+], **Rukshan Wijesinghe**[*+], **Kumara Kahatapitiya**[*], **Subha Fernando**[*+] and **Ranga Rodrigo**[*]

[*]University of Moratuwa, Sri Lanka

[+]CODEGEN QBITS Lab

# Rich Layer-wise Feature Extraction by Multi-paths

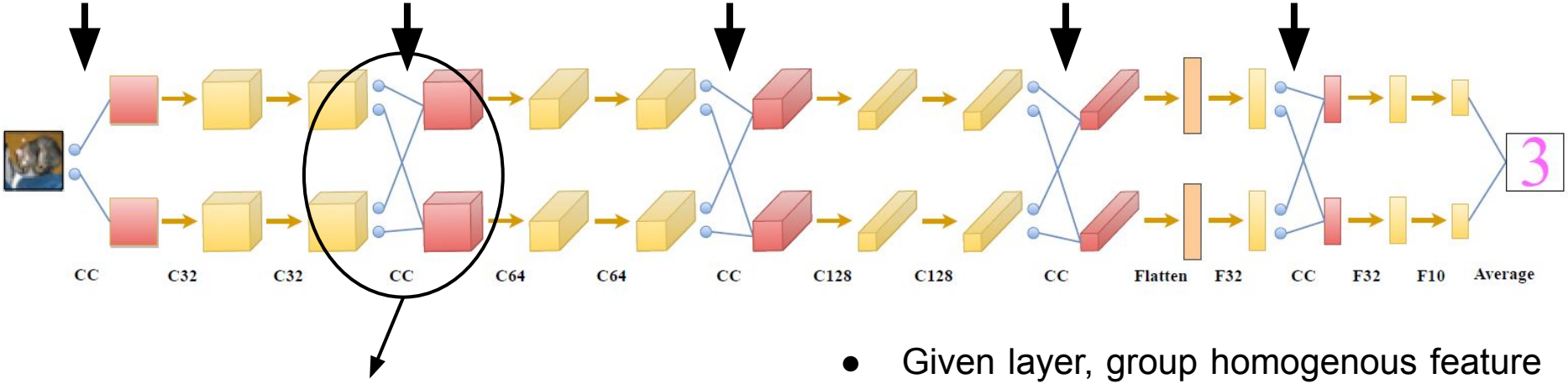
- Neural network deepening is well established
- Powerful feature extraction within layers?
- ~~Conventional Widening~~ → Parallel computations in a layer
- No context-dependent allocation of resources in a layer

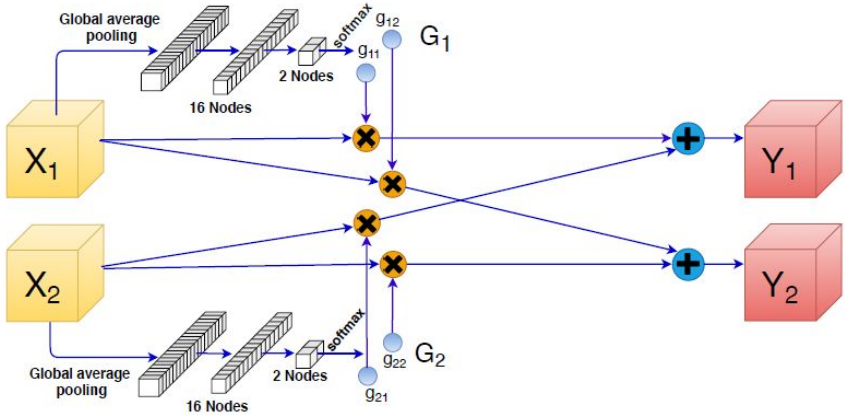



a) Hummingbird b) Hummingbird c) Electric Eel

- Image context is distributed along the depth of NN
- In a multi-path network, the nature of resource allocation may change with the depth
- It is intuitive to learn the resource allocation separately, layer-wise.

# Feature-Dependent Cross-Connections

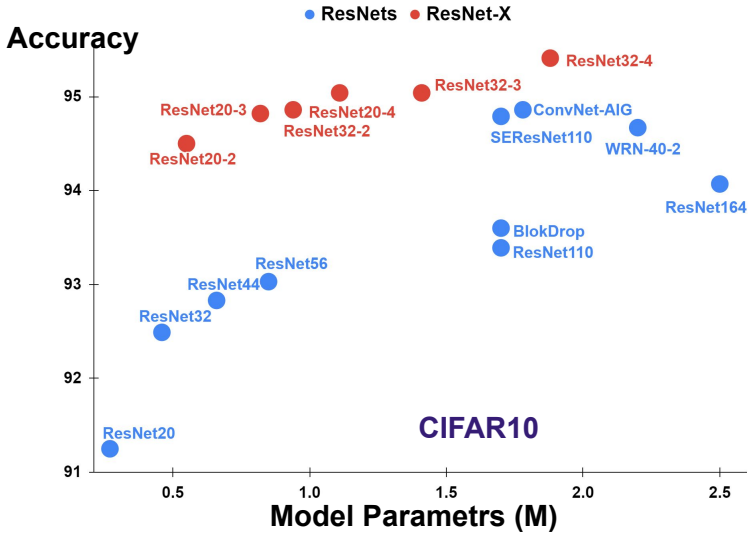


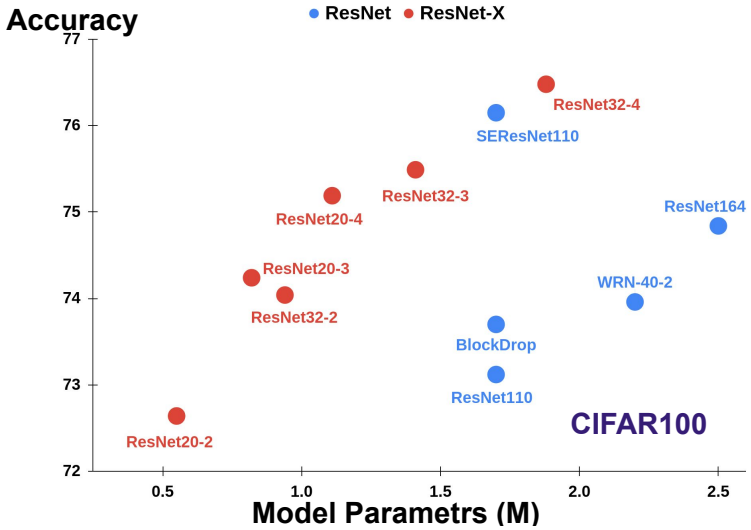



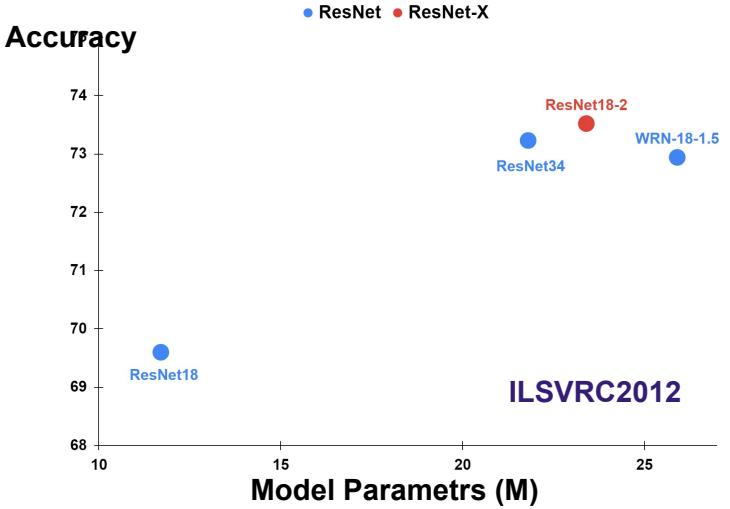

- Given layer, group homogenous feature maps to parallel paths
- Route the input, layer-wise, end-to-end through such paths

- X → Global Average Pooling → non-linear computation → gates for X
- Cross-weight the connections and add to output Y

# Image Recognition Domain

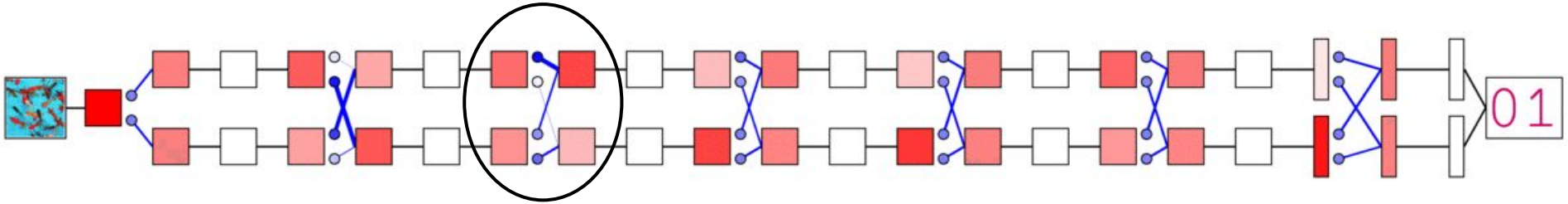


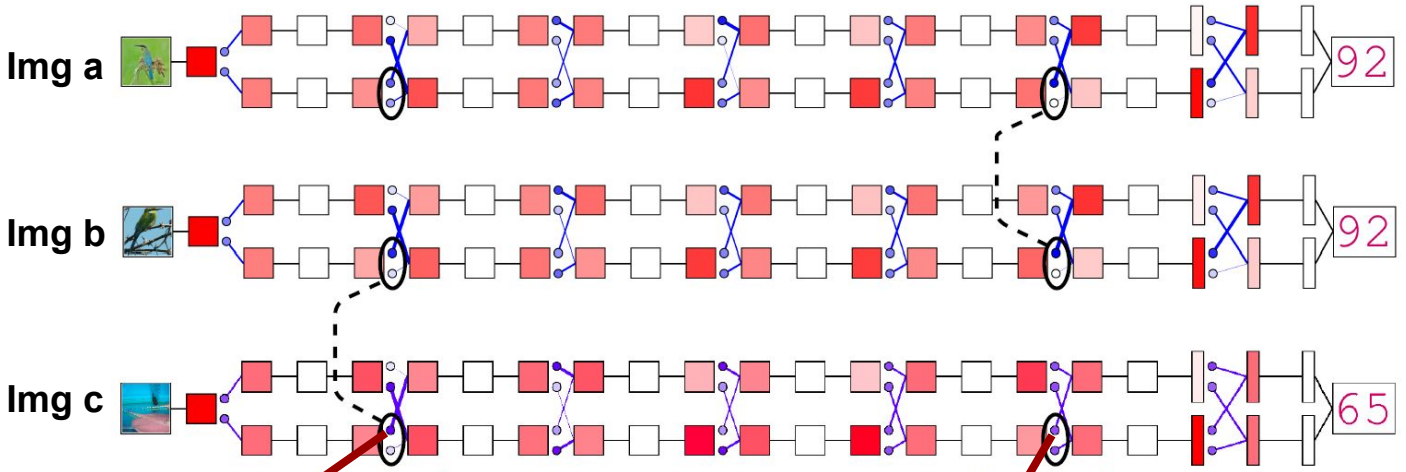

- ResNet-X → ResNet with X paths and cross-connections
- Our multi-path nets surpass
  - Conventional widening
  - Existing adaptive feature extraction methods
  - Deeper networks
- With similar or less complexity

# Routing Visualization




- Input-output activation strengths (red intensities)
- Gate strengths (blue intensities and connection thickness)
- White boxes show the layer stack where no cross-connections are inserted
- Slightly different soft dynamic routing can be observed

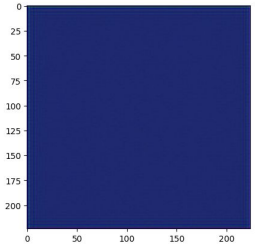
# Gating Patterns




**Shallow Gate**

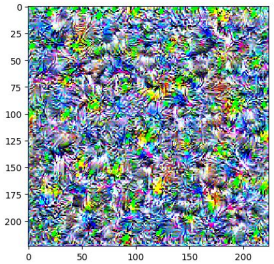
Maximized for b and c

Maximaly activated images



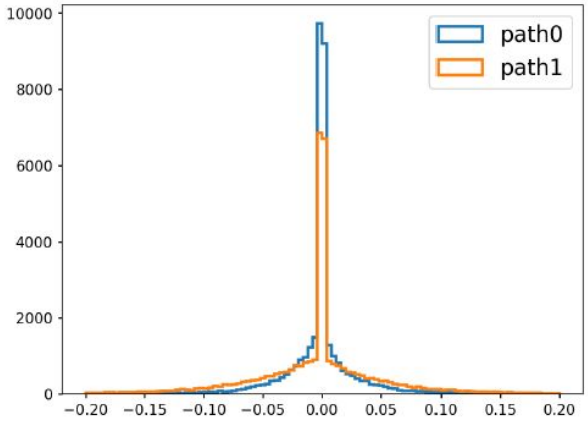
**Deeper Gate**

Maximized for a and b

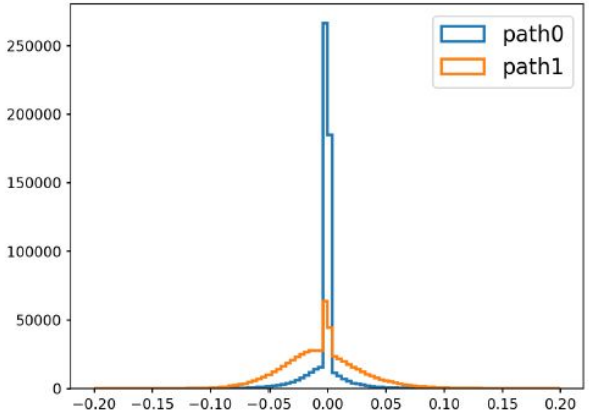


Synthesized image to maximize gate

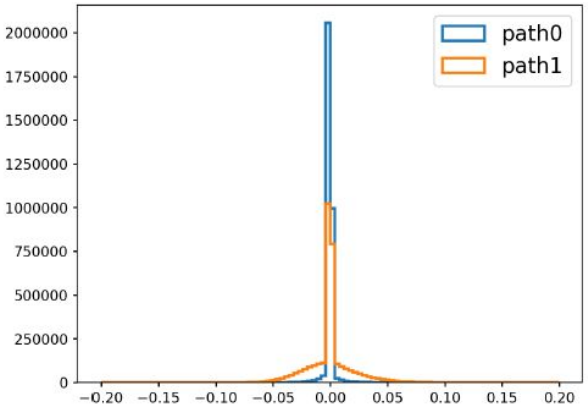
# Gate Histograms of Parallel Paths




layer 4




layer 8




layer 11

- A single layer (one graph) consists of two histograms (two paths)
- Thanks to the adaptive cross-connection based routing, the parallel computations learn distinct features.

# Thank You

**More Info:**

Join our Q&A Session or,

email: dumindutissera@gmail.com