Optimizing Convolutional Neural Networks for Cloud Detection

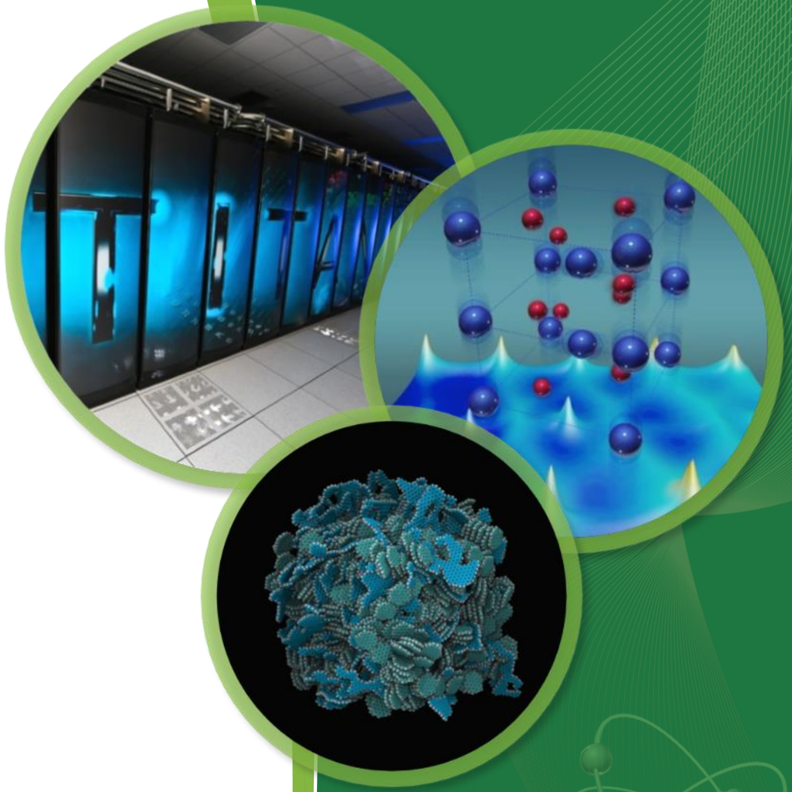
Travis Johnston,

Steven R. Young,

David Hughes,

Robert M. Patton, and

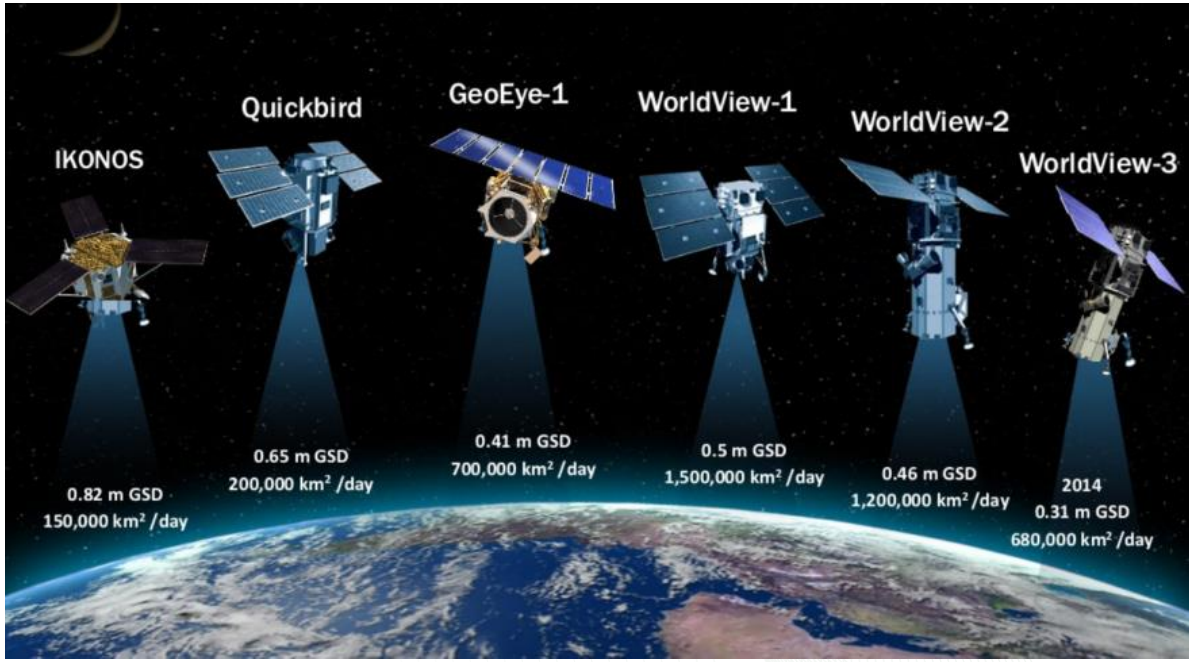
Devin White



ORNL is managed by UT-Battelle
for the US Department of Energy

Background: Abundant Overhead Imagery

- Earth imaging satellites are generating more data at a faster pace than ever.
- About 70% of the earth is covered by clouds.
- Imaging scientists care about clouds for one of two reasons:
    - they are studying clouds (e.g. climate scientists)
    - the clouds are *in the way* (e.g. geographers or cartographers)



Digitalglobe constellation as of October 2015

Background: Approaches for Cloud Detection

- Radiometric feature-based algorithms which utilize exact spectral bands that manifest *cloud characteristics*.
- Temporal images
- Methods that use spectral characteristics as well as texture and/or spatial data

These algorithms rely on hand-engineered feature extraction.

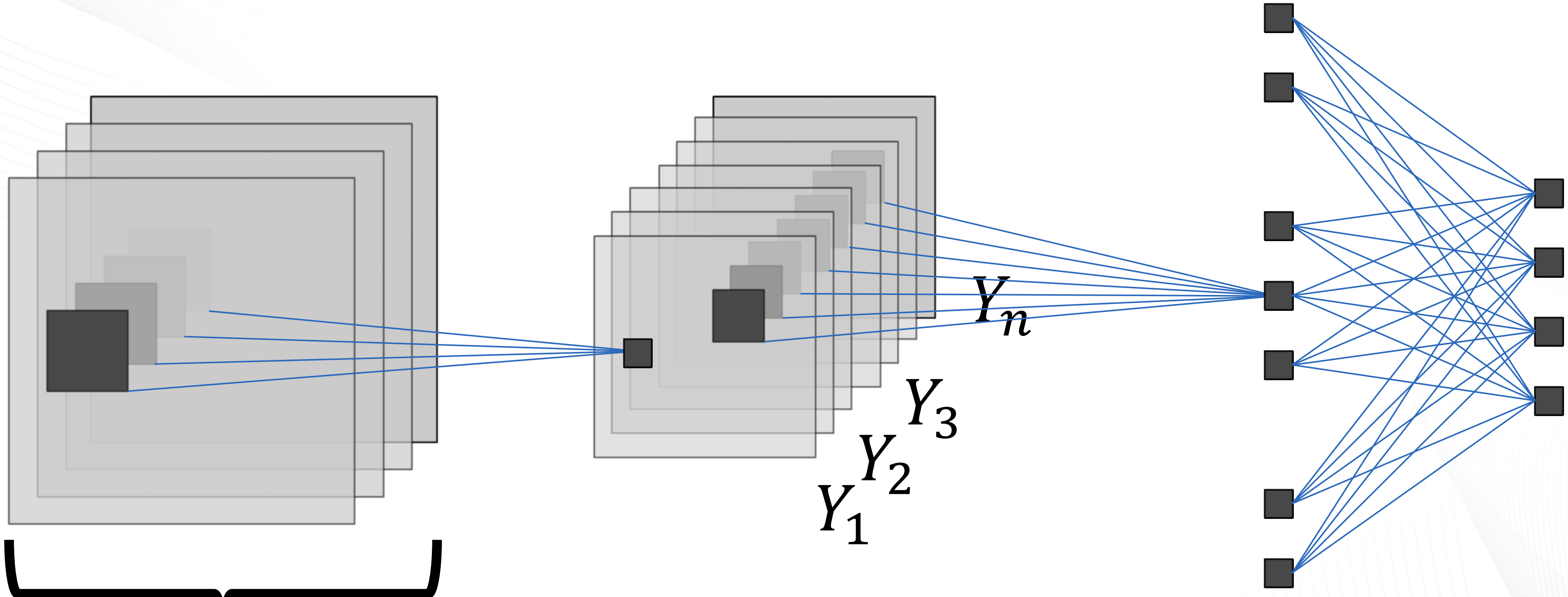
Our Approach: Deep Learning…

Hand Engineered Feature Extraction

Convolutional Neural Networks

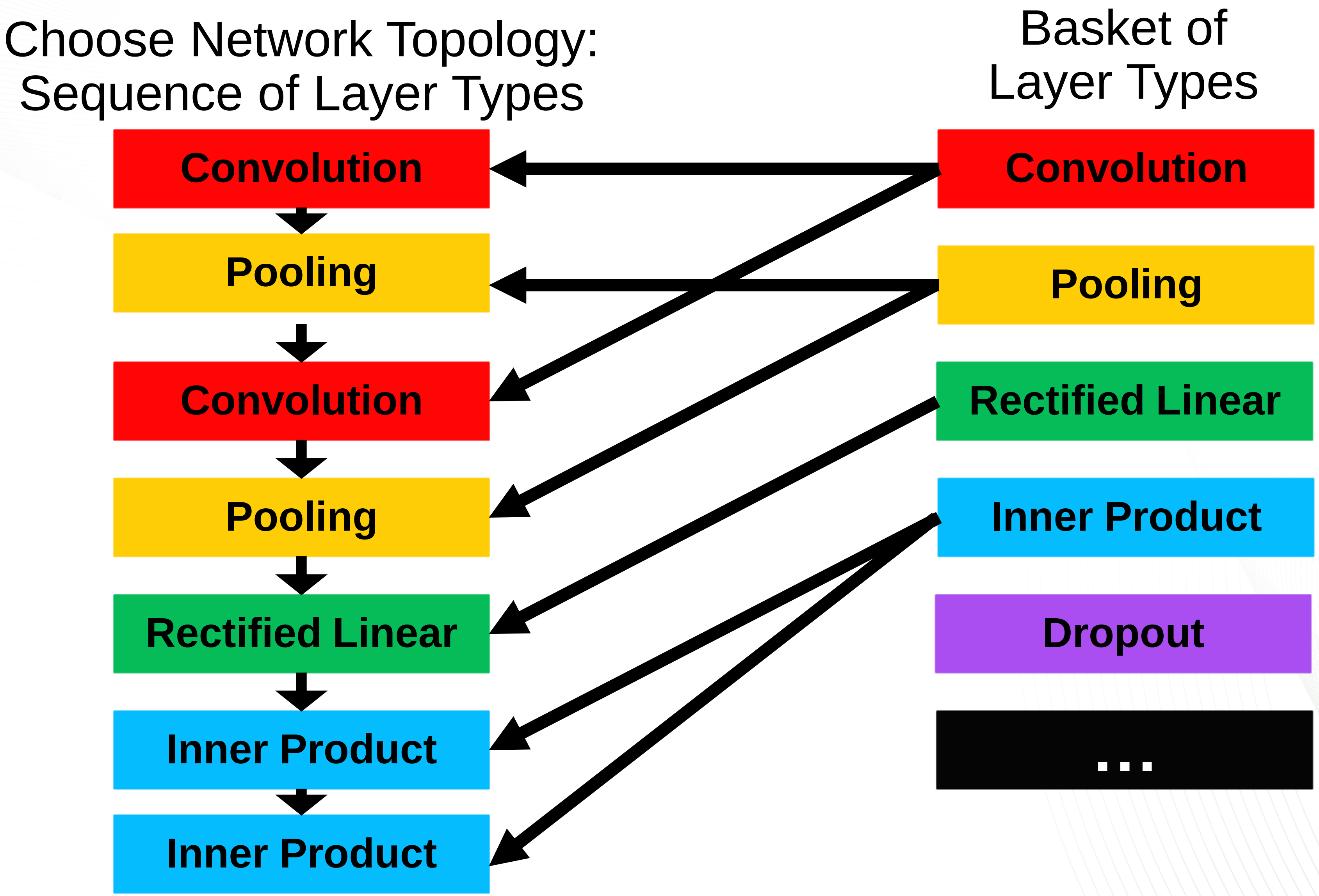
- Learn *general* features from training data (WorldView-2)
- Extensible across spectral bands (trained with RGB and IR)
- Fast inferencing with GPUs

What are convolutional neural networks (CNNs)?



$$f_i(X) = Y_i$$

What is required to deploy a CNN?



What is required to deploy a CNN?

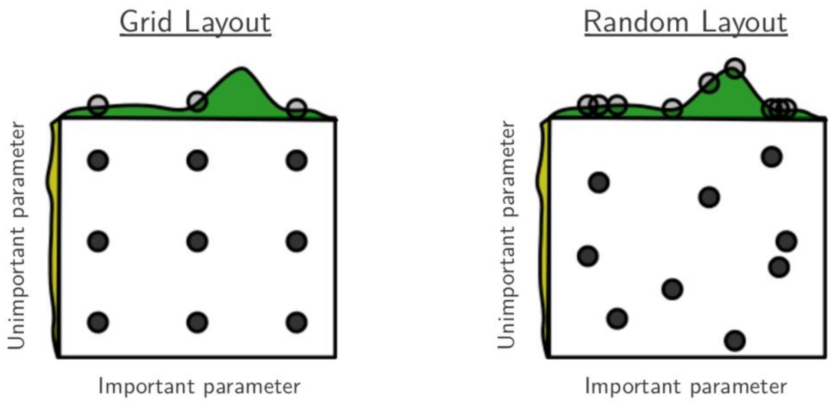
For each layer:
Assign values to layer hyper-parameters

Convolution
↓
Pooling
↓
Convolution
↓
Pooling
↓
Rectified Linear
↓
Inner Product
↓
Inner Product

- Number of Kernels (output images)
- Size of Kernels
    - Kernel width
    - Kernel height
- Pad size
    - Pad width
    - Pad height
- Stride
    - Horizontal stride
    - Vertical stride

Current Approaches to Hyper-parameter Optimization

- Use *out-of-the-box* network
  - Why spend time trying to create your own network when there are already so many good ones available? Surely, one of those networks will also solve your problem.

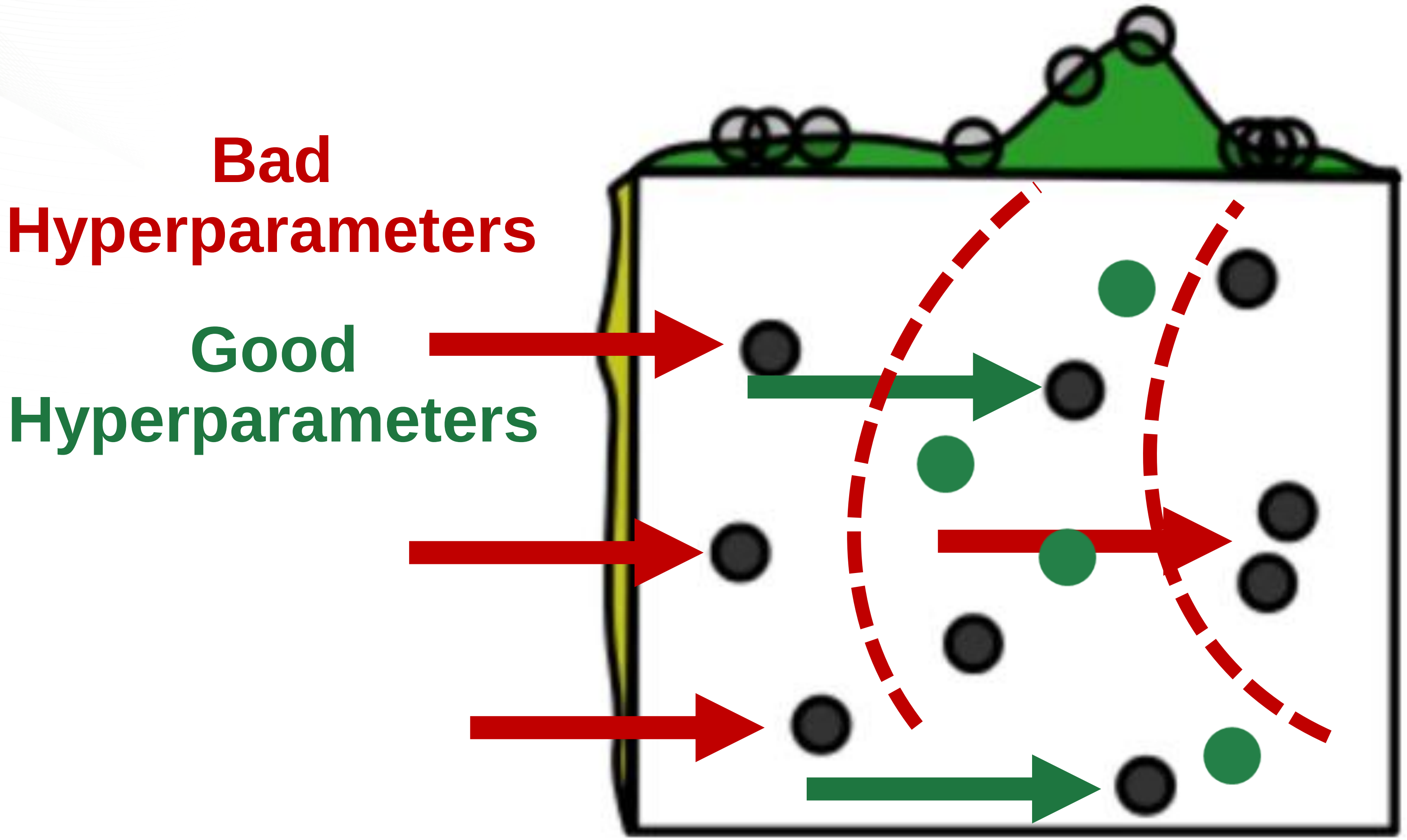


Bergstra, J, and Bengio, Y. Random Search for Hyperparameter Optimization, Journal of Machine Learning Research, Feb. 2012.

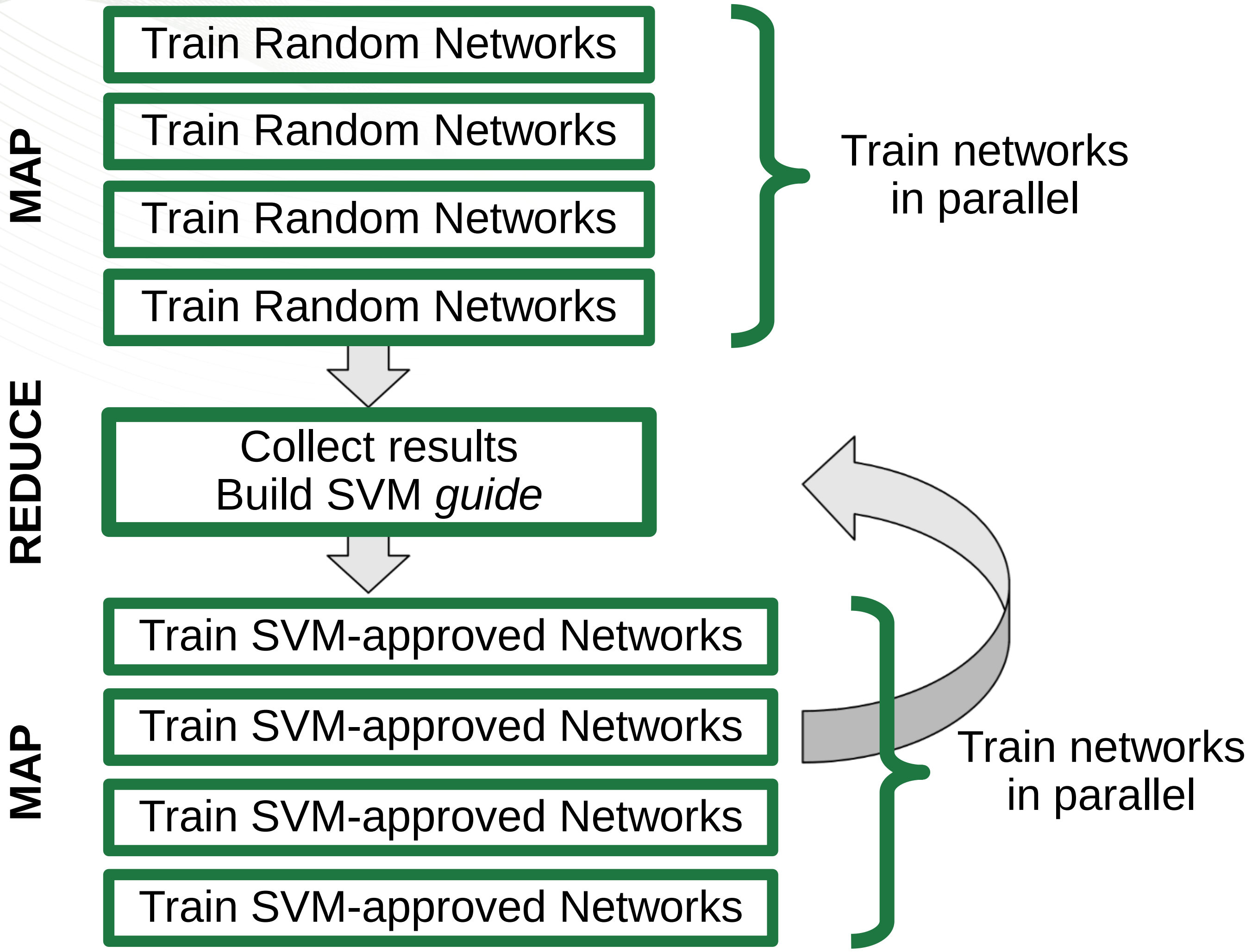
- *Tune* an out-of-the-box network
  - **Hyper-parameter sweeps**
    Assumes independence of hyper-parameters
  - **Grid search**
    Requires training an exponential number of networks (infeasible)
  - **Random search**
    Significant improvement over grid search, but doesn't make use of information learned during training.

OAK RIDGE National Laboratory

Improving random search



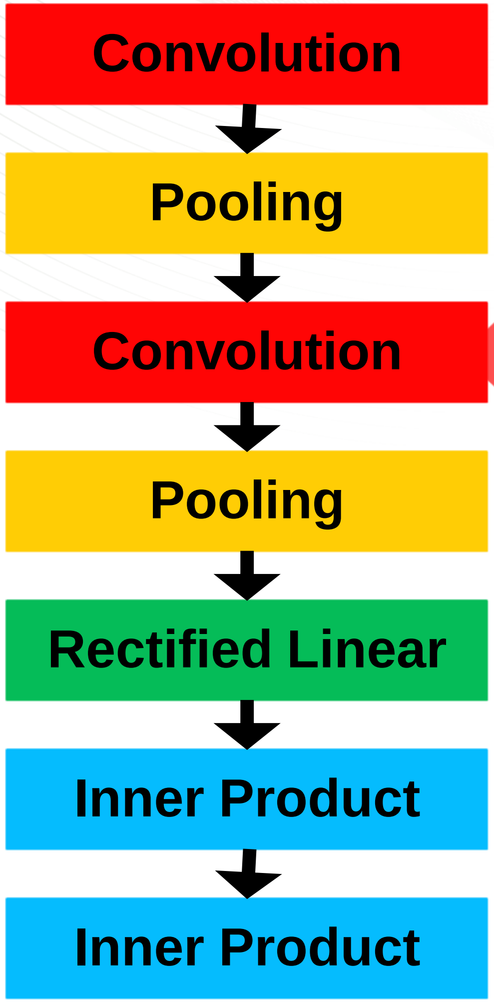
RAvENNA: RApidly Evolving Neural Network Architectures



RAvENNA:

- Implemented in Apache-Spark
  - Extremely parallel
  - Fault tolerant
  - Widely available
- Uses **caffe** to train networks
- Running on Titan
  (up to 18,000 nodes)

Optimizing LeNET for cloud detection

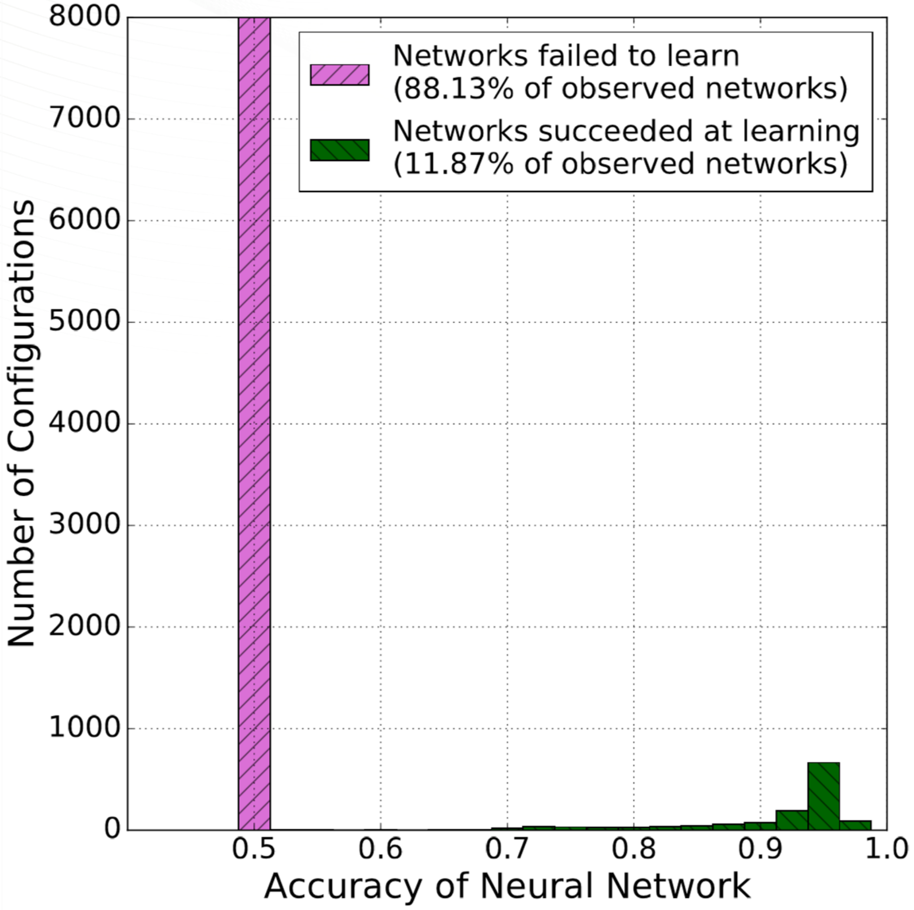


- Number of Kernels (output images)
- Size of Kernels
  - Kernel width
  - Kernel height
- Pad size
  - Pad width
  - Pad height
- Stride
  - Horizontal stride
  - Vertical stride

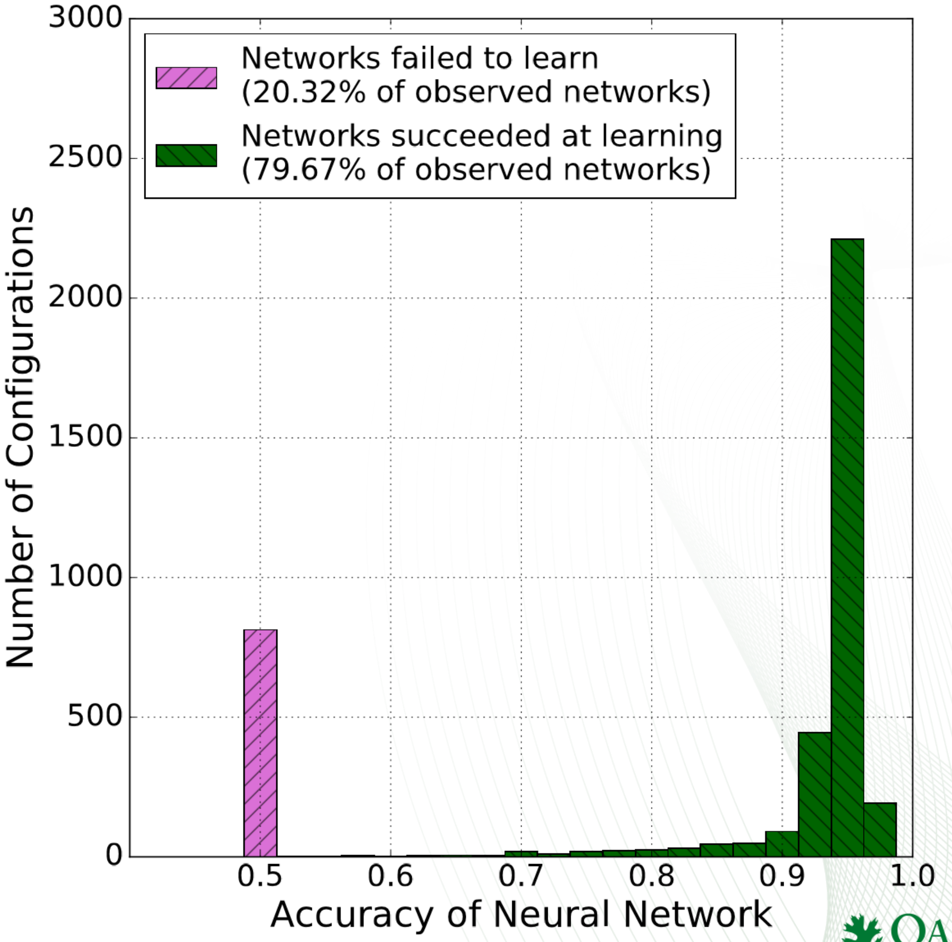
Oak Ridge National Laboratory

How much does RAvENNA improve random search?

Random Search



RAvENNA



OAK RIDGE National Laboratory

But, optimal hyper-parameters are intuitive right?

	Convolution			Pooling		Convolution			Pooling		Inner Product		
	Number of Outputs	Kernel Size	Stride	Kernel Size	Stride	Number of Outputs	Kernel Size	Stride	Kernel Size	Stride	Number of Outputs	Number of Outputs	Accuracy
LeNET	20	5	1	2	2	50	5	1	2	2	500	(2)	50.0%
Random Variations	57	120	7	4	1	94	2	2	3	1	604	(2)	50.0%
	113	38	3	3	1	103	13	1	14	2	439	(2)	50.0%
Optimized	101	6	1	9	8	64	3	3	3	2	770	(2)	97.1%

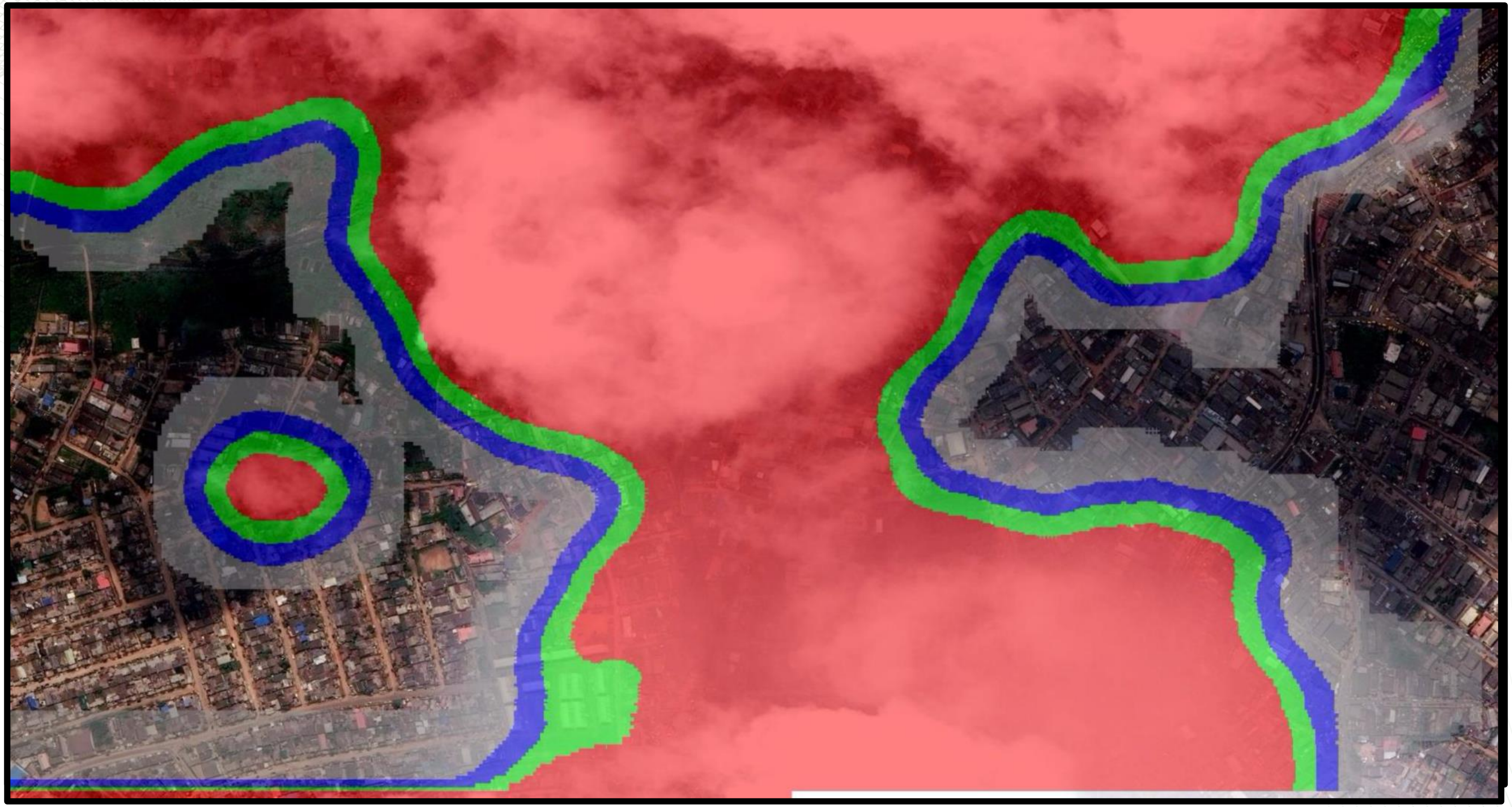
What does 97.1% accuracy mean in practice?

Network's Cloud Score (*probability*)

(0, 25]
(25, 50]
(50, 75]
(75, 100]

Oak Ridge National Laboratory

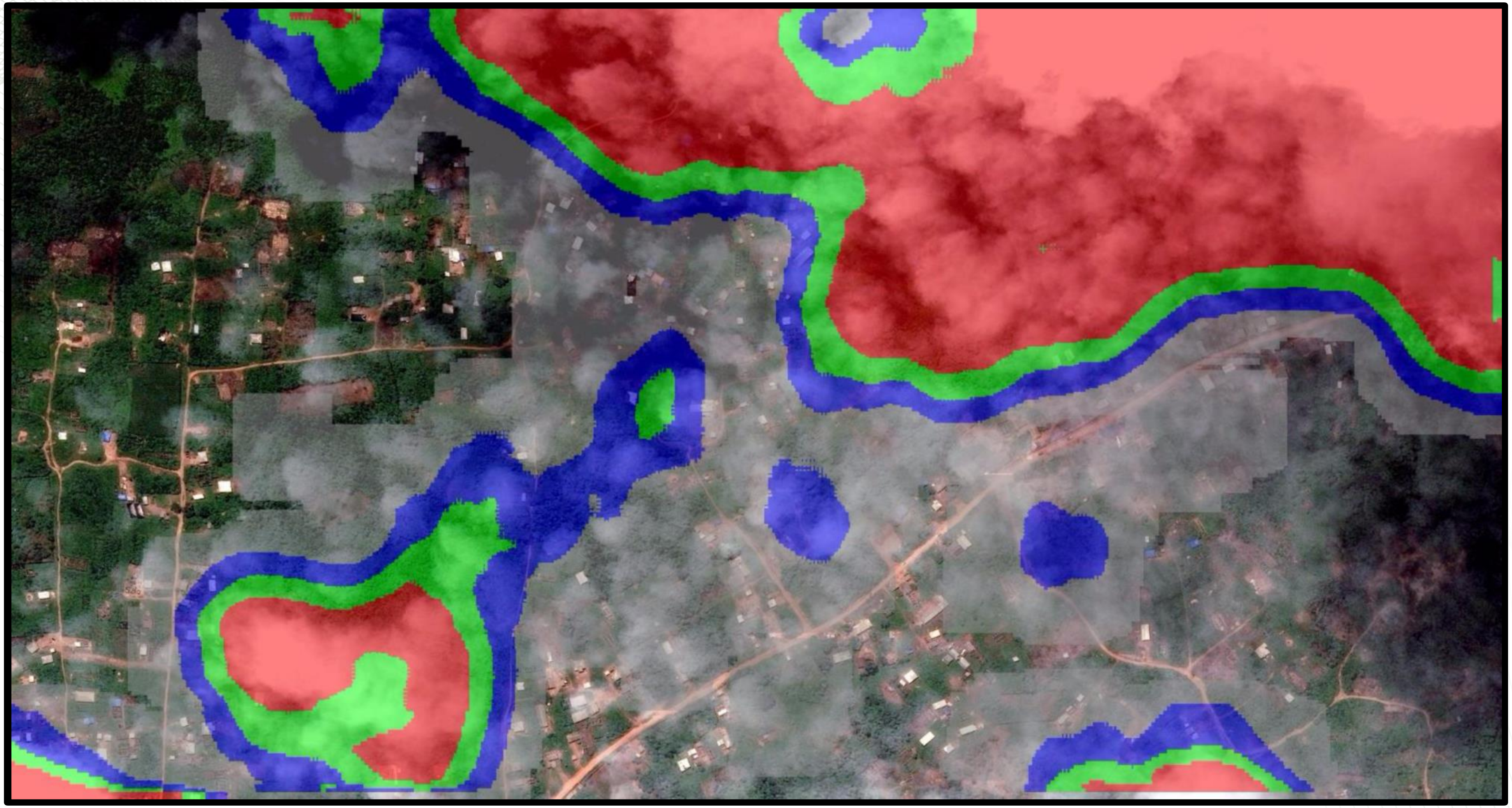
What does 97.1% accuracy mean in practice?



Network's Cloud Score (*probability*)

(0, 25]
(25, 50]
(50, 75]
(75, 100]

What does 97.1% accuracy mean in practice?



Network's Cloud Score (*probability*)

(0, 25]
(25, 50]
(50, 75]
(75, 100]

Conclusion

- We demonstrated RAvENNA's ability to optimize the hyper-parameters of a fixed network topology by applying it to the specific task of cloud detection in overhead imagery.
- RAvENNA's optimized network outperformed GoogLeNET by:
  - More than 40% reduction in error
  - More than 200x speedup in inference time
  - Memory requirements less than 1/10th that of GoogLeNET.
- RAvENNA used 4000 nodes of Titan for 1 hour for optimization.