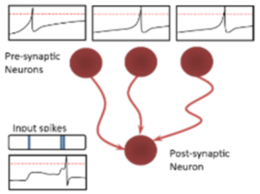
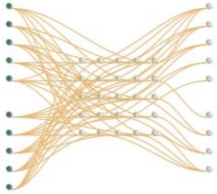
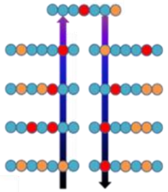
Sandia National Laboratories

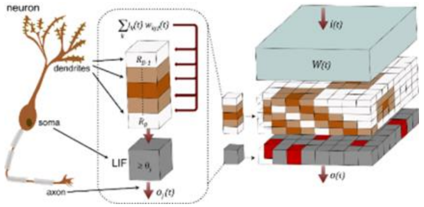
SAND2020-12681C

Neural Network Approaches for Enabling Automatic Target Recognition









PRESENTED BY

Craig M. Vineyard - cmviney@sandia.gov

Matt Kagie, Ryan Melzer, Srideep Musuvathy, John Richards, William Severa, Darby Smith

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

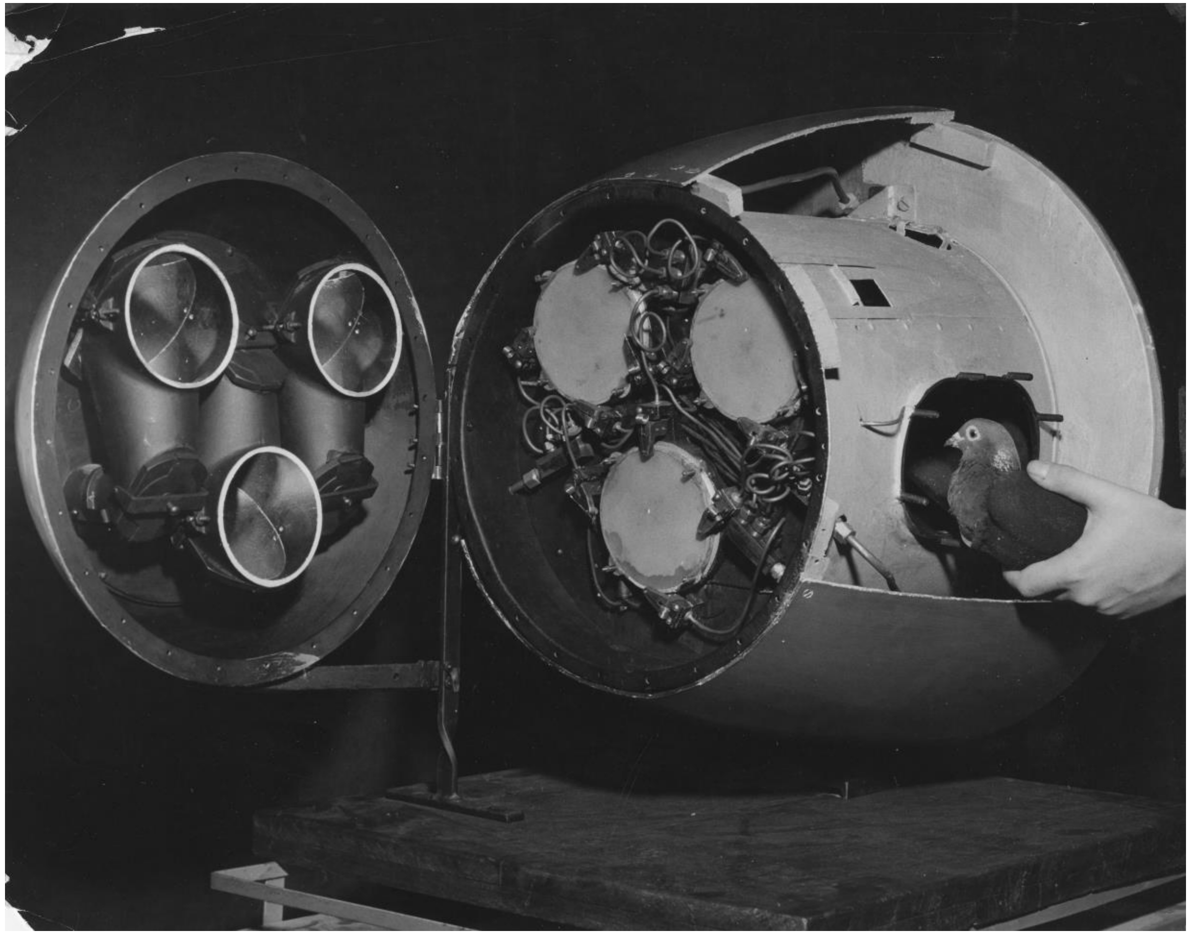
Automatic Target Recognition (ATR)

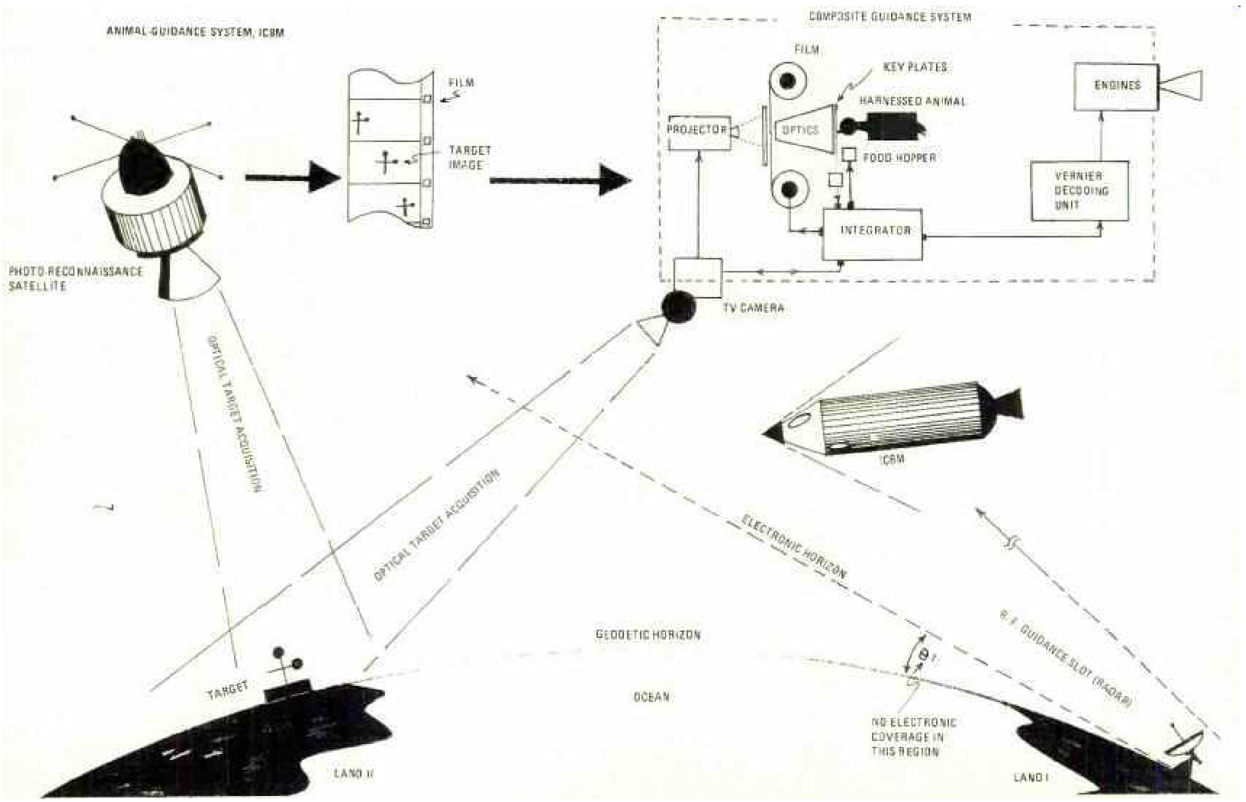
Automatic Target Recognition (ATR) - an exploitation algorithm for the detection or classification of items of interest via a remote sensor

Three-stage ATR block diagram

- Detector - first operates upon the raw sensed data to extract regions which express features or expressions that there may be a target of interest in the smaller identified sub-region
- Indexer - operates upon this reduced data to compare against the representations of known targets of interest
- Identifier - receives regions of interest (ROIs) as well as cues/hypotheses regarding the salient features (whether template or model based) which are used to determine a quantified score

Project Orcon



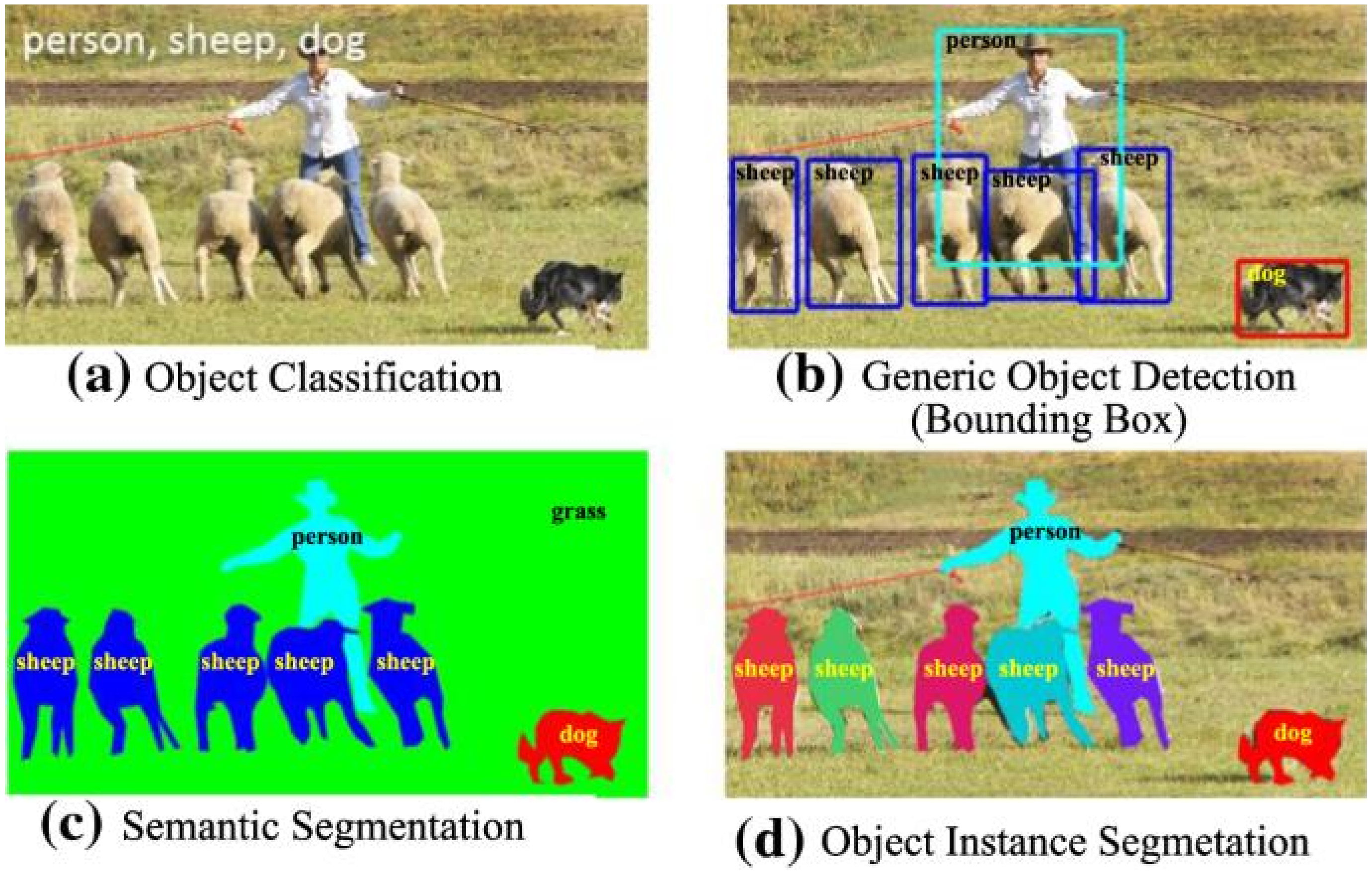


https://www.industrial-electronics.com/ew_1971-12_animal.html

Image Processing

Neural networks have enabled many breakthroughs and state-of-the-art performance for a variety of image processing tasks

- Example include – detection, classification, segmentation, tracking, generation

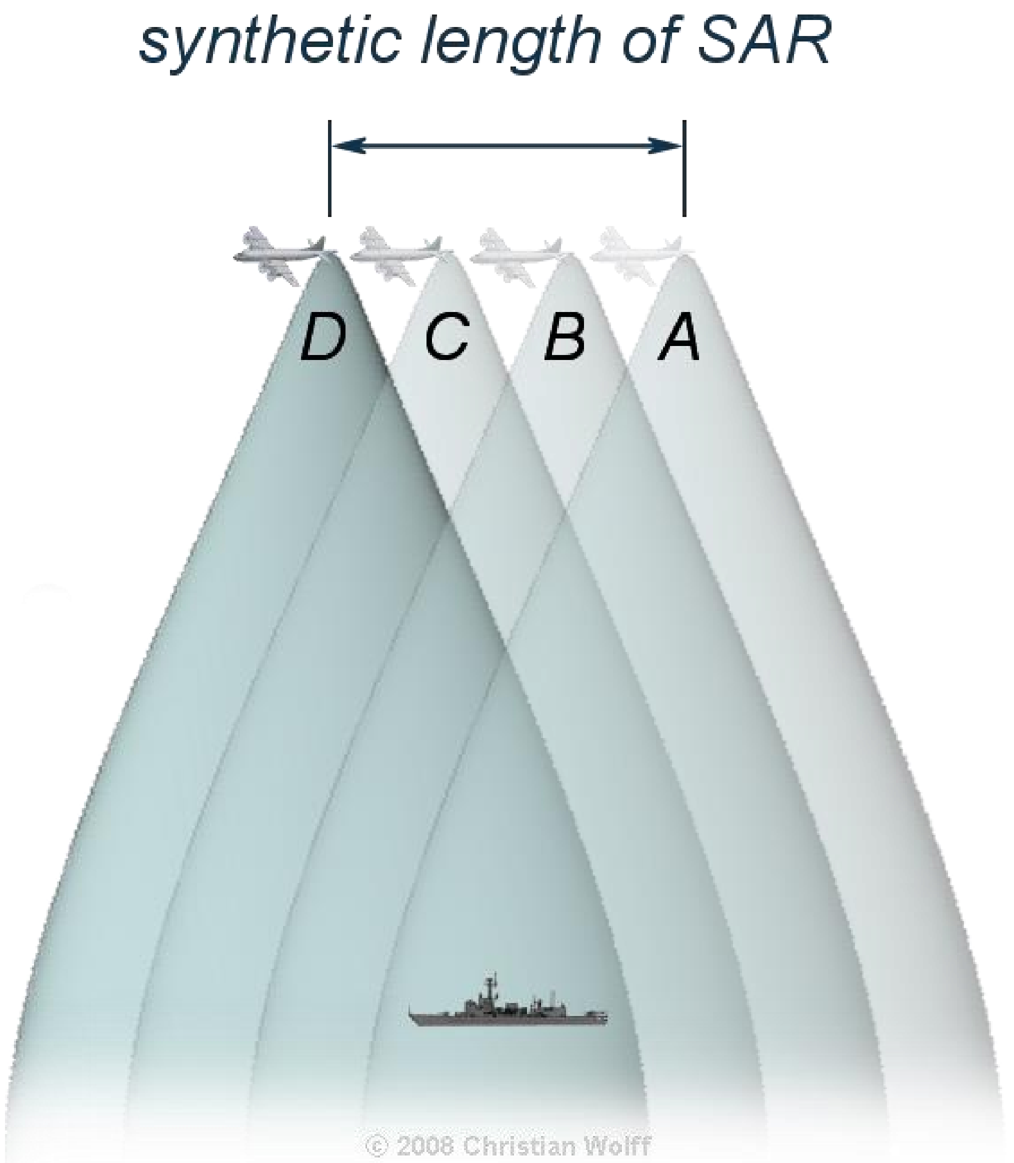


Liu, Li, et al. "Deep learning for generic object detection: A survey." *International journal of computer vision* 128.2 (2020): 261-318.

Synthetic Aperture Radar (SAR)

Radar based alternative to optical images

- Uses motion of radar antenna to create a large synthetic aperture enabling high resolution
- Measures radio frequency reflectivity of the imaged scene
- Weather robust



https://lynceans.org/all-posts/synthetic-aperture-radar-sar-and-inverse-sar-isar-enable-an-amazing-range-of-remote-sensing-applications/

Synthetic Aperture Radar (SAR)

Figure 1. SAR image of a location at Kirtland Air Force Base, Albuquerque, N.M., exhibiting 4-inch (10 centimeter) resolution. Note that the aircraft are better defined by their shadows than by their direct echo return.

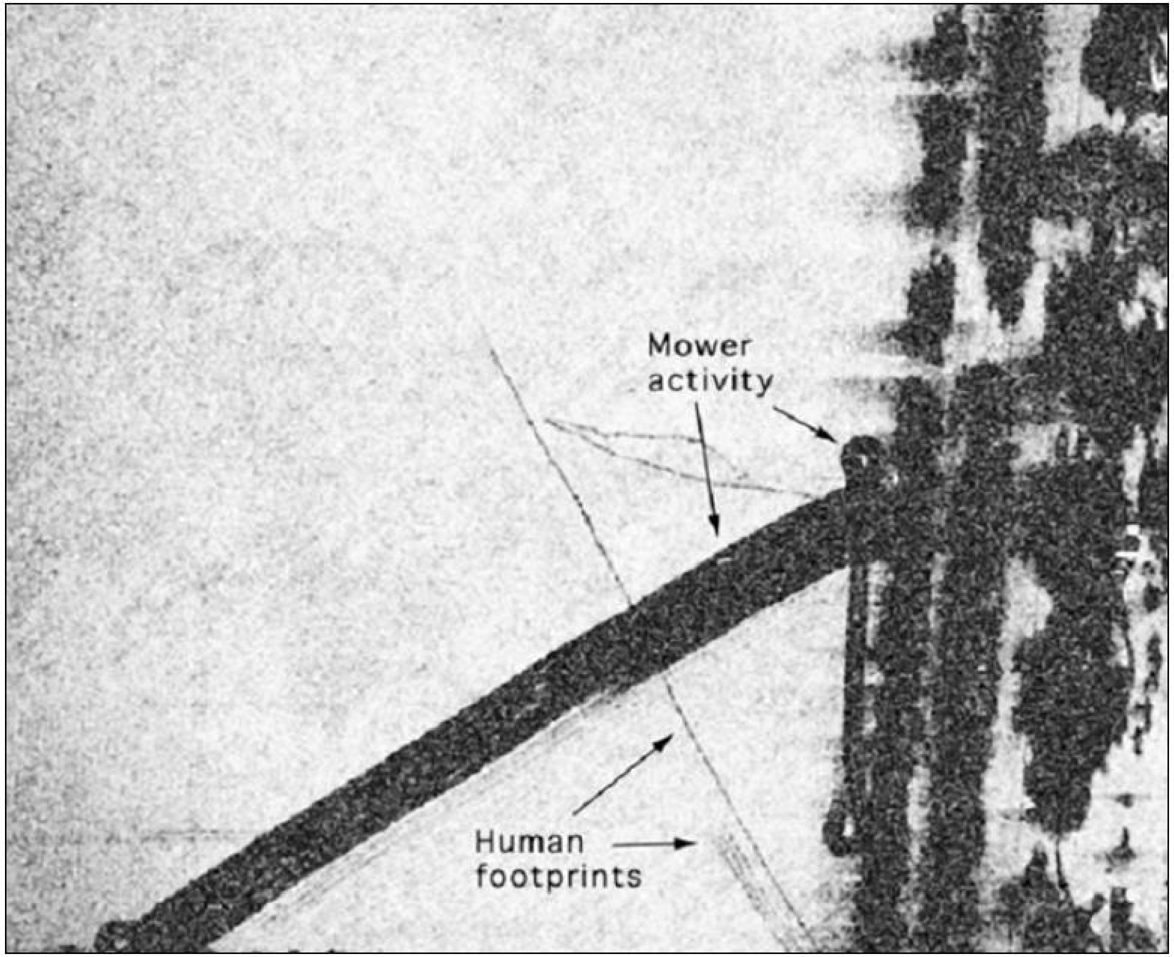


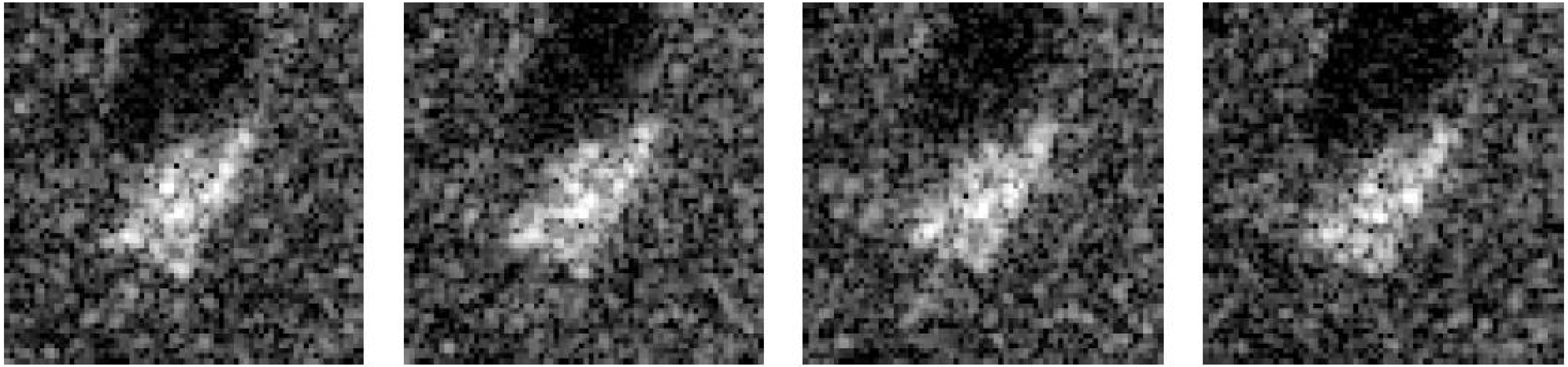
Figure 6. Coherent change detection map showing mower activity and footprints on Hardin Field Parade Ground at Kirtland Air Force Base. Dark areas denote regions of decorrelation caused by a disturbance to the clutter field; light areas denote no disturbance. The foliage along the right side of the image decorrelates because of wind disturbance.

Doerry, Armin W., and Fred M. Dickey. "Synthetic aperture radar." *Optics and photonics news* 15.11 (2004): 28-33.

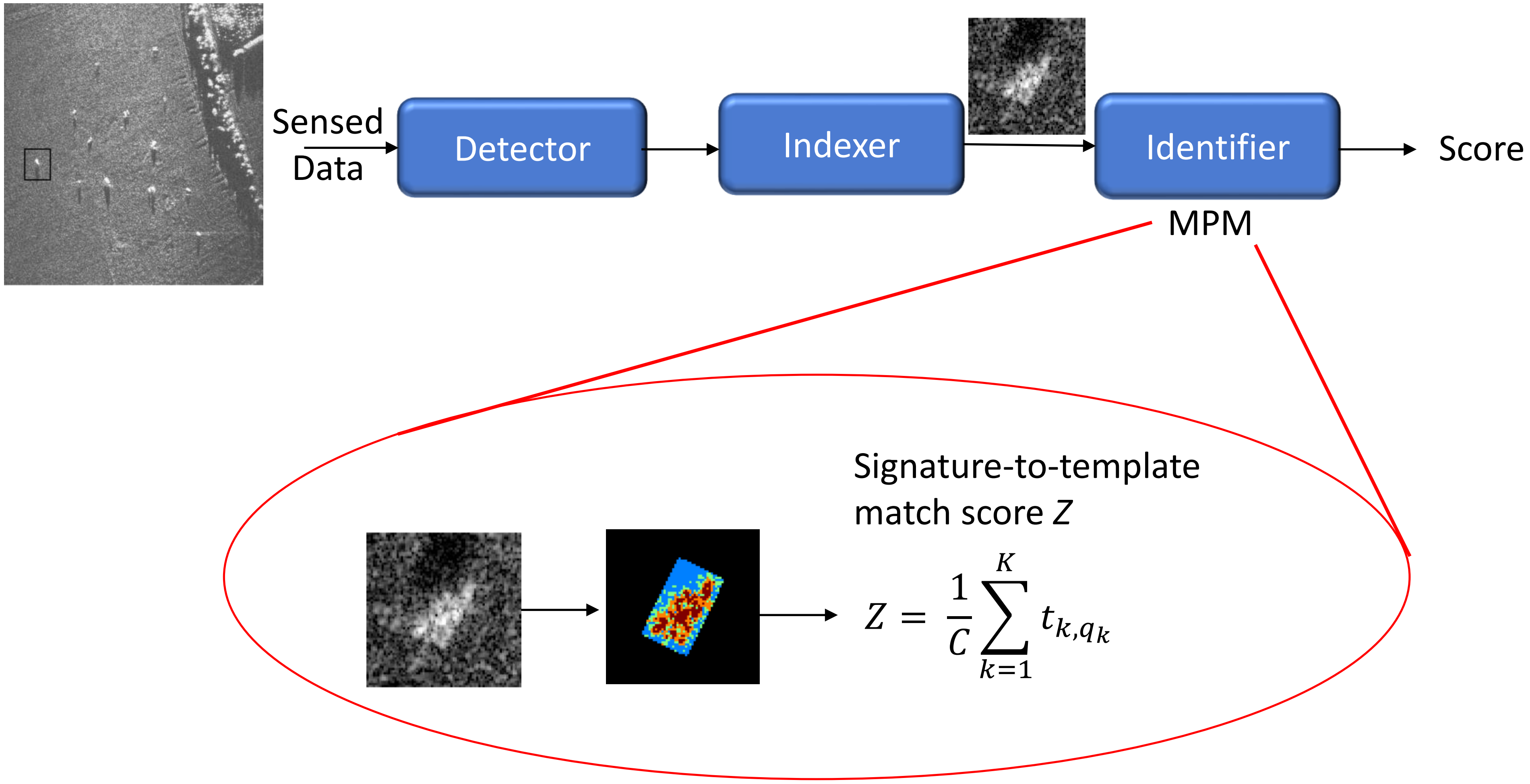
Synthetic Aperture Radar (SAR)

Advantages but also sensor induced challenges - Signal variability due to:

- Coherence: complex valued measurements encompassing magnitude and phase
- Specularity: radar energy is scattered directionally instead of diffusely as a consequence of the wavelength and size of objects
- Speckle: multiplicative noise process due to the coherent interaction between multiple scatters in individual cells



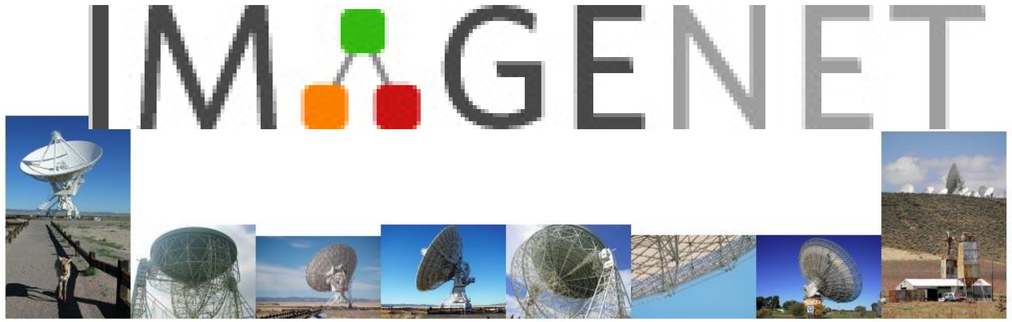
Multinomial Pattern Matching (MPM)



Data Challenge

Deep Neural Networks have greatly benefit from large, labeled data sets such as ImageNet

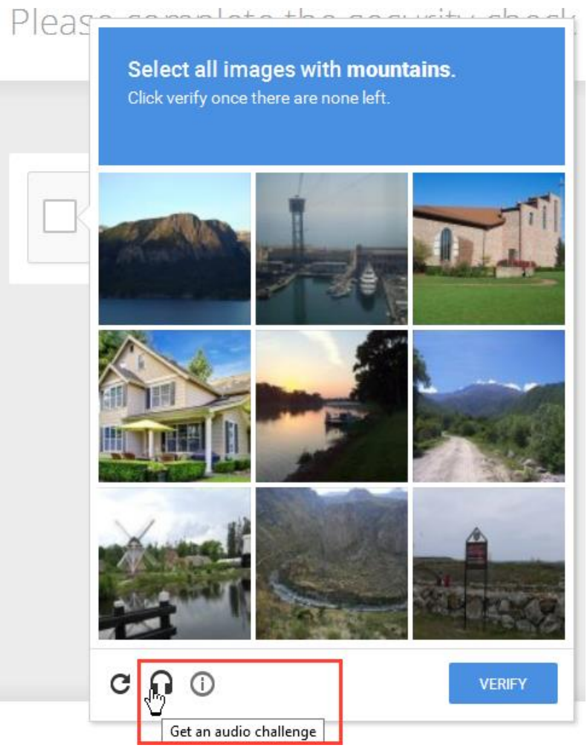
- 14,197,122 images (as of August 2020)



http://www.image-net.org/

CAPTCHA

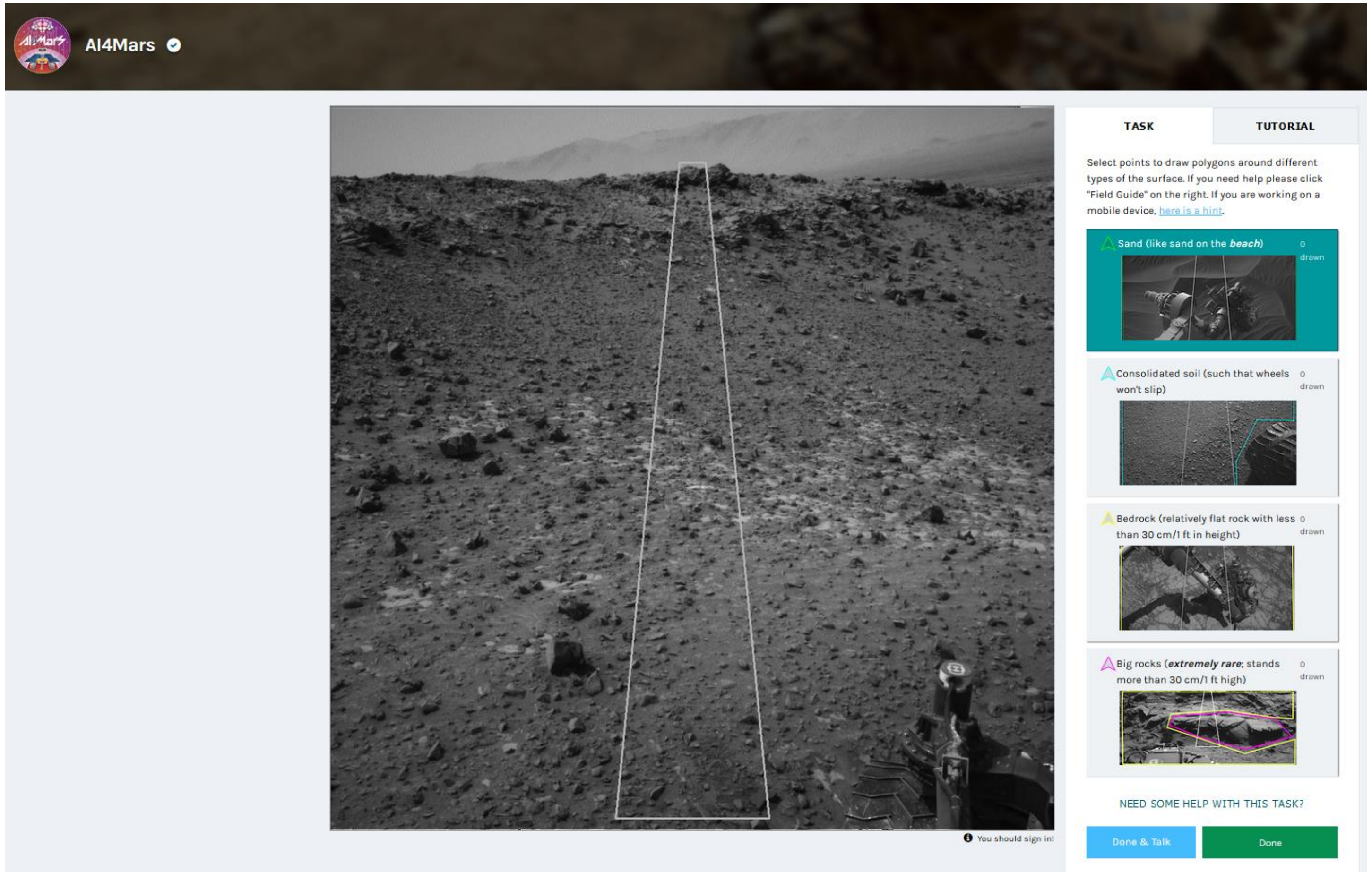
- Have provided labels for datasets



AI4Mars

There's even a program for helping to label terrain on Mars

But for ATR we do not have an abundance of data

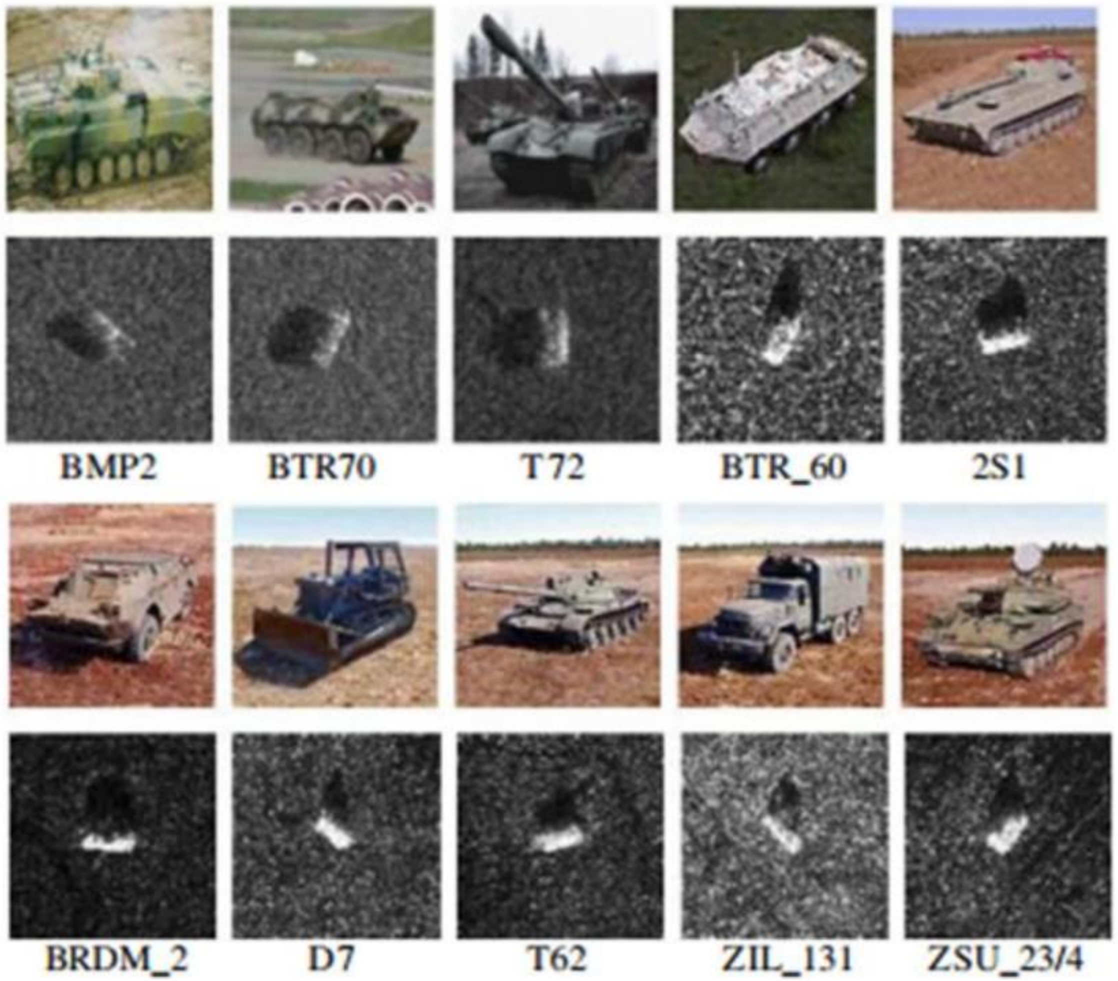


https://www.zooniverse.org/projects/hiro-ono/ai4mars

MSTAR

Moving and Stationary Target Acquisition and Recognition

- Collection of one-foot resolution SAR images
- Collected by the Air Force Research Laboratory, Sandia National Laboratory, and the Defense Advanced Research Projects Agency (DARPA) during the latter half of the 1990s

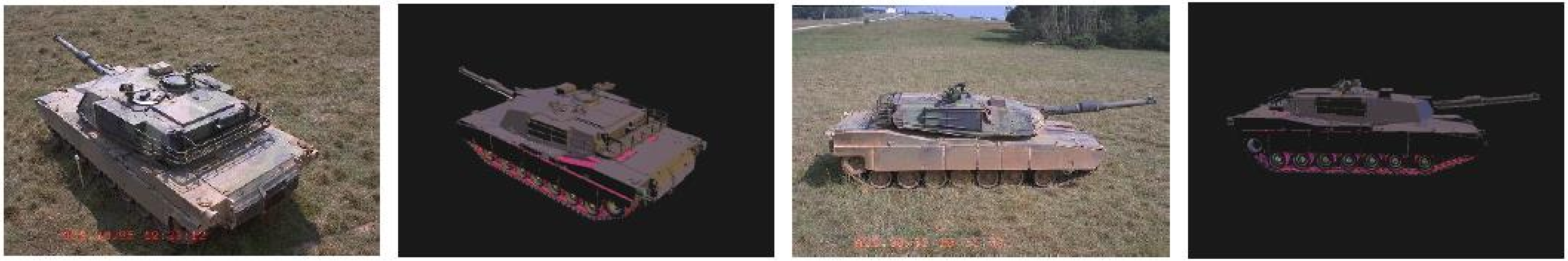


Targets	BMP2	BTR70	T72	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU234
17	233	233	232	256	299	298	299	299	299	299
15	587	196	582	195	274	274	274	273	274	274

SAMPLE

Synthetic and Measured Paired Labeled Experiment

- Uses electromagnetic computational tools to provide predictions of the radar return for highly realistic CAD models
- Includes accounting for material properties for each surface of a target (glass, pain, metal, rubber, etc.) to compute the electromagnetic property values of the radar return



(a) (b) (c) (d)

- Comparison between optical images of an M1 taken during the MSTAR data collect (a, c) and CAD models of the same vehicle (b, d) from two viewpoints

Lewis, Benjamin, et al. "A SAR dataset for ATR development: the Synthetic and Measured Paired Labeled Experiment (SAMPLE)." *Algorithms for Synthetic Aperture Radar Imagery XXVI*. Vol. 10987. International Society for Optics and Photonics, 2019.

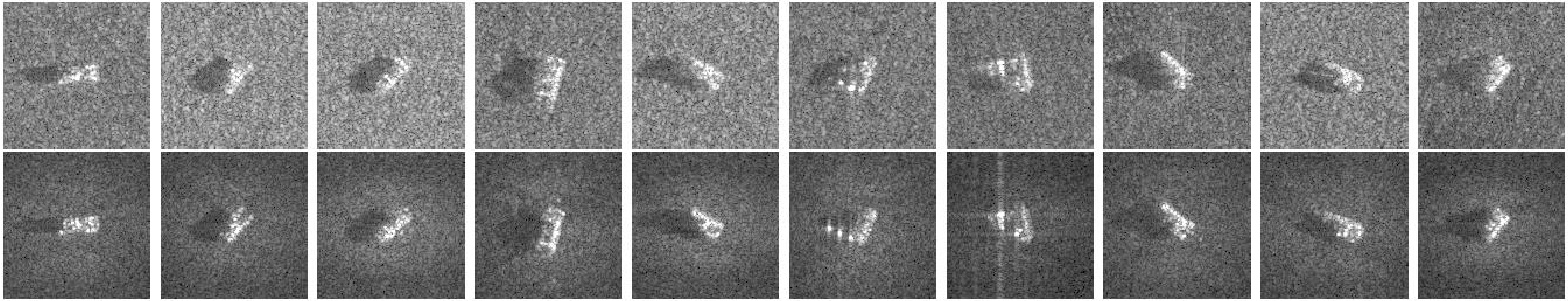
SAMPLE

Enables the development of a more operationally realistic dataset which can train on synthetic images and test on measured (real) data

Shown below measured MSTAR data is the top row and the corresponding synthetic images are the bottom row

Class	Measured	Synthetic	Total
2S1	177	177	354
BMP2	108	108	216
BTR70	96	96	192
M1	131	131	262
M2	129	129	258
M35	131	131	262
M548	129	129	258
M60	178	178	356
T72	110	110	220
ZSU23	177	177	354
Total	1366	1366	2732

Table 5: Distribution of publicly available SAMPLE data for each class.



Lewis, Benjamin, et al. "A SAR dataset for ATR development: the Synthetic and Measured Paired Labeled Experiment (SAMPLE)." *Algorithms for Synthetic Aperture Radar Imagery XXVI*. Vol. 10987. International Society for Optics and Photonics, 2019.

Neural Network Zoo

Algorithmic sweeps over many neural network architectures providing insight into impacts of different computational approaches

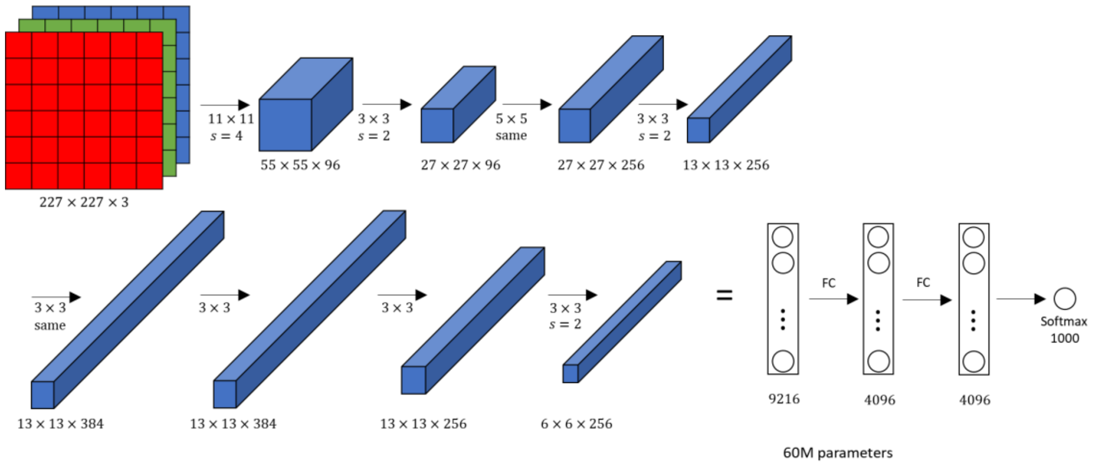
- AlexNet
- VGG
- ResNet
- DenseNet
- ResNeXt
- WideResNet
- SEResNet
- MobileNetV2
- MobileNetV3
- SEMobileNet
- ShuffleNet
- SqueezeNet
- EfficientNet
- MnasNet
- SENet

- ~15 Unique architectures
- > 50 total models (architecture and configuration e.g. ResNet34, ResNet50, Efficientnet-b0, …)

AlexNet & VGG

AlexNet

- The original record-breaking ImageNet CNN
- Hand tuned kernel sizes, number of filters, number of layers, etc.



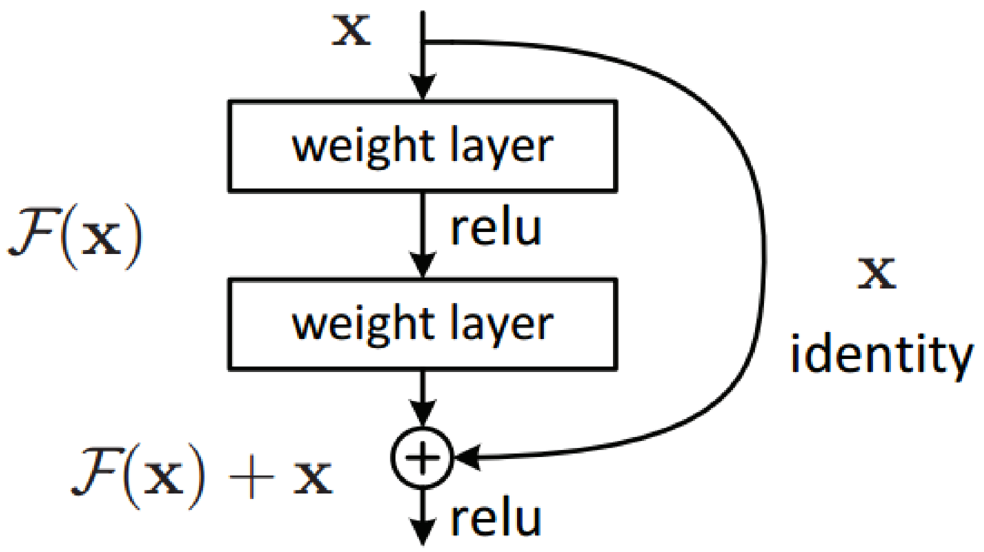
http://datahacker.rs/deep-learning-alexnet-architecture/

VGG

- Improvement on AlexNet
- All 3x3 kernels
- All layers have same configuration (conv, batch norm, relu, max pool)
- Number of channels per set of layers increases by powers of two.
- Deeper and wider than AlexNet
- 11, 13, 16, and 19 layer configs

ResNet

- Neural Nets get better as they get deeper but information gets harder to propagate when the model is forced to learn too many sequential non-linear transformations
- Want to keep making nets deeper
- Solution: use residual connections to help propagate information
- Residual connection: add output of layer L-1 to layer L
- ResNets much deeper than AlexNet/VGGs and more effective. Up to 200 layers
- Every model after this point uses residuals in some form

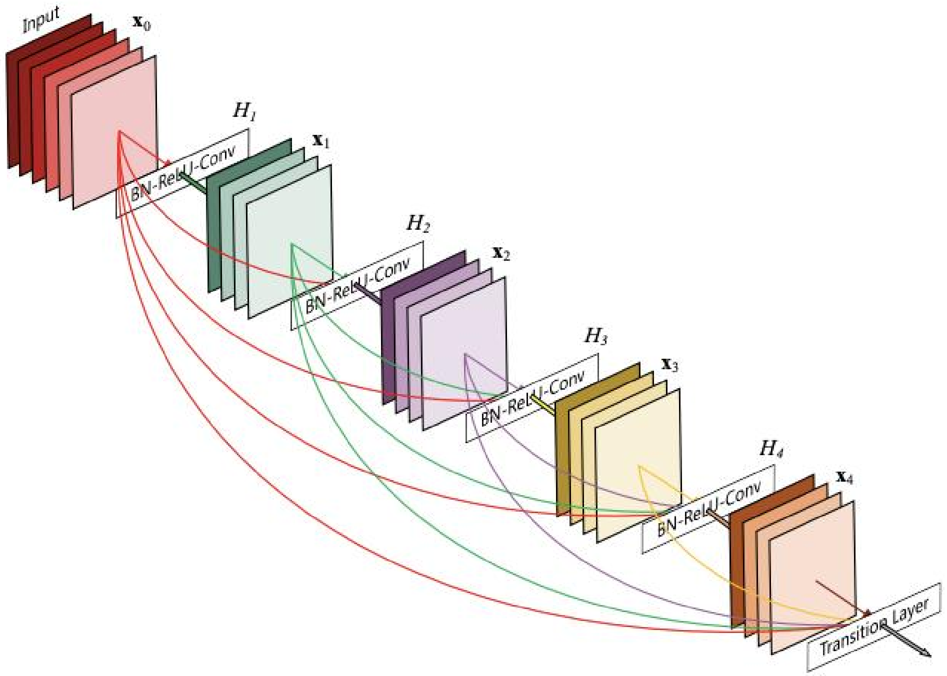


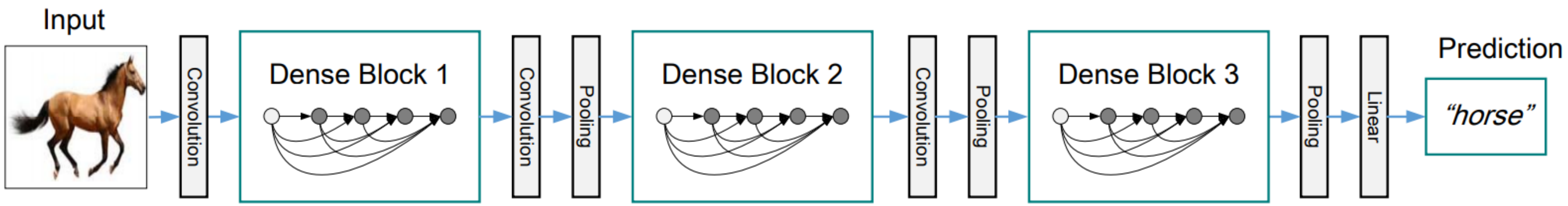
A residual 'block'

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

DenseNet

- Idea: instead of using a single residual connection from the previous layer, use residuals from all previous layers.
- Uses concatenation instead of addition.
- Utilizes fewer number of filters per layer and much more depth.

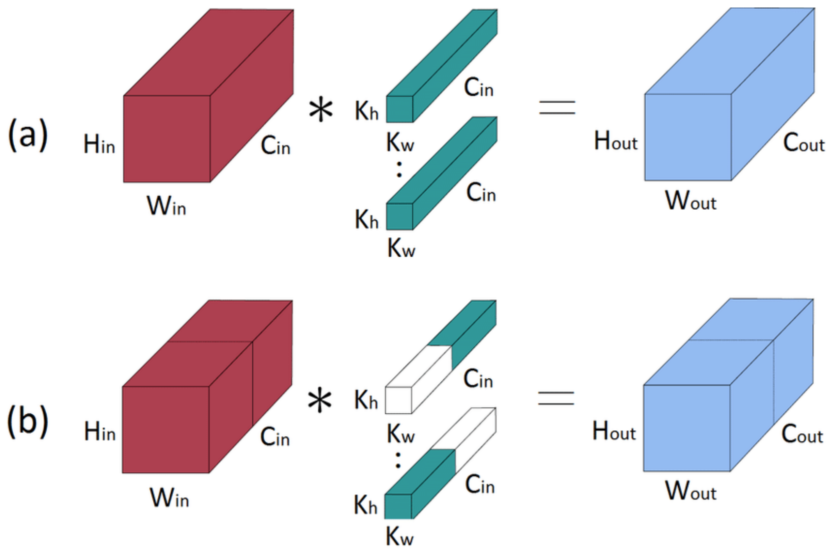




Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

ResNeXt

ResNeXt = ResNet + ***grouped convolutions***



Gibson, Perry, et al. "Optimizing Grouped Convolutions on Edge Devices." *2020 IEEE 31st International Conference on Application-specific Systems, Architectures and Processors (ASAP)*. IEEE, 2020.

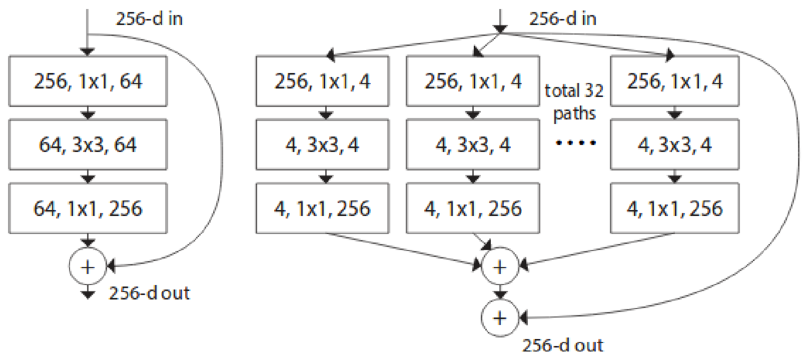


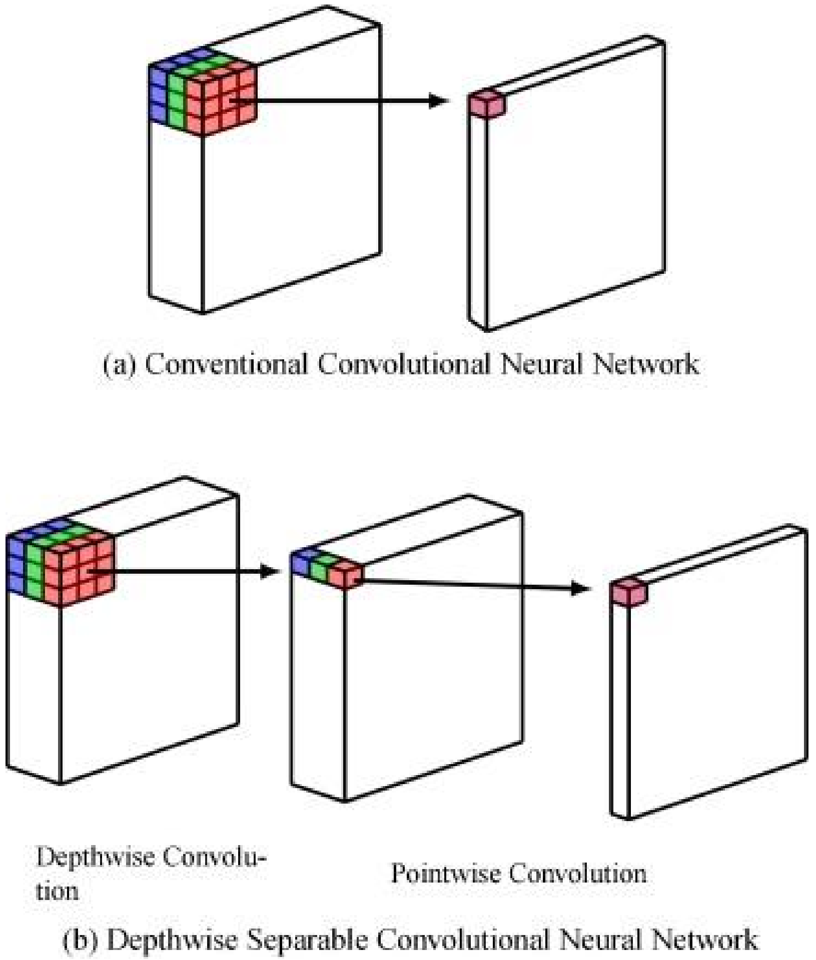
Figure 1. **Left**: A block of ResNet [14]. **Right**: A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

Xie, Saining, et al. "Aggregated residual transformations for deep neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

- Grouped convs have been shown to be very effective in image classification.
- Far fewer # of params.
- Wider nets can fit on GPU
- Implicit regularization through sparsity

ShuffleNets, MobileNets, MnasNets

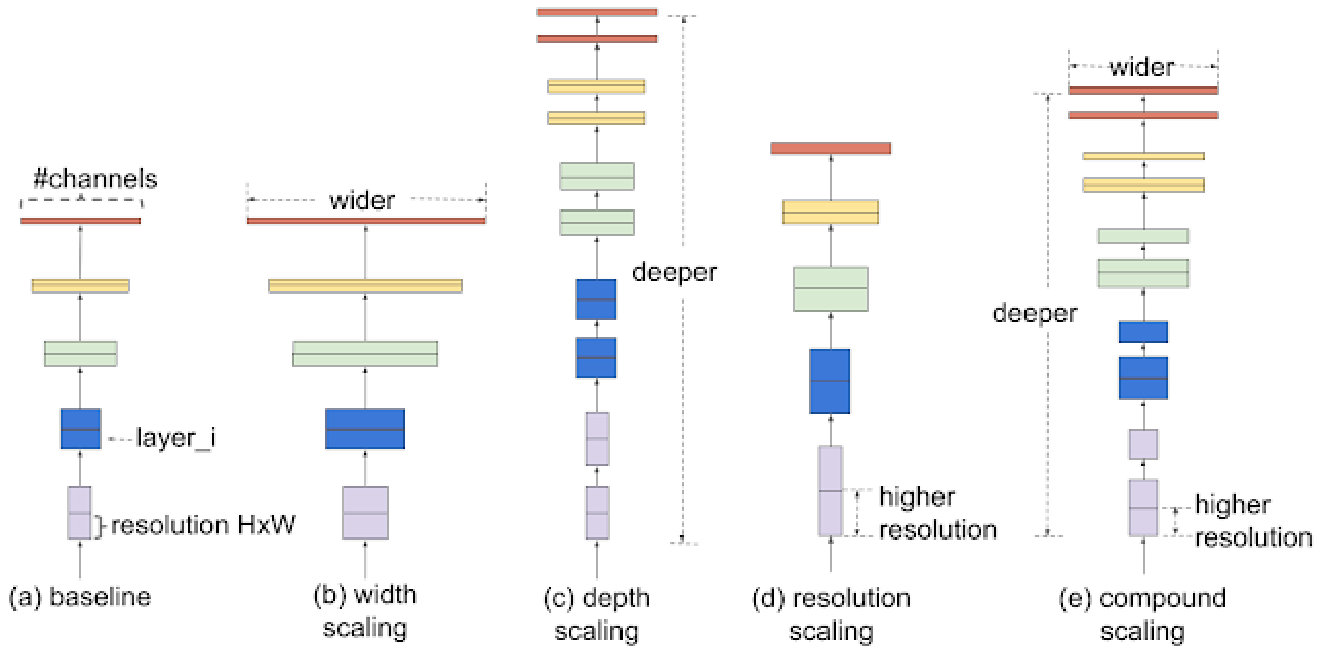
- Architectures optimized for smartphones and other less compute, less parallel platforms.
- Lots of tricks, but the major contribution is the **depthwise-separable convolution.**
- Reduces the number of operations required for a convolution by 9-25x.
- Decomposes a KxKxHxWxC convolution to a 1x1xHxWxC convolution and C 3x3xHxWx1 convolutions (i.e. a grouped convolution where the number of groups = number of input channels)



Kamal, K. C., et al. "Depthwise separable convolution architectures for plant disease classification." *Computers and Electronics in Agriculture* 165 (2019): 104948.

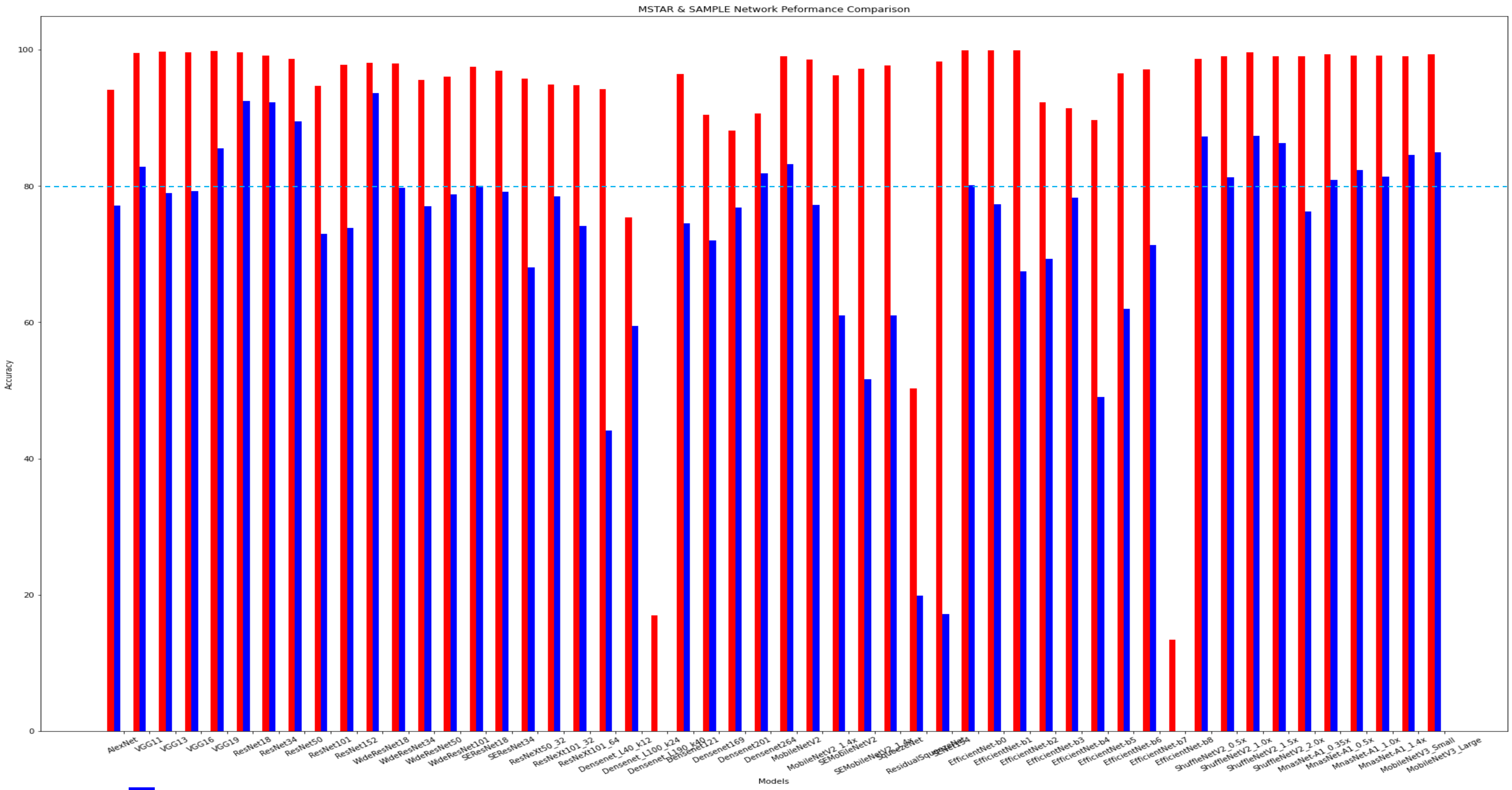
EfficientNet

- Roughly current state of the art in image classification and object detection
- Utilizes more intelligent model scaling
- A small baseline network was found through NAS (efficientnet-b0)
- The baseline network is then scaled up in width (number of channels), depth (number of layers), and resolution (input image size) in proportion



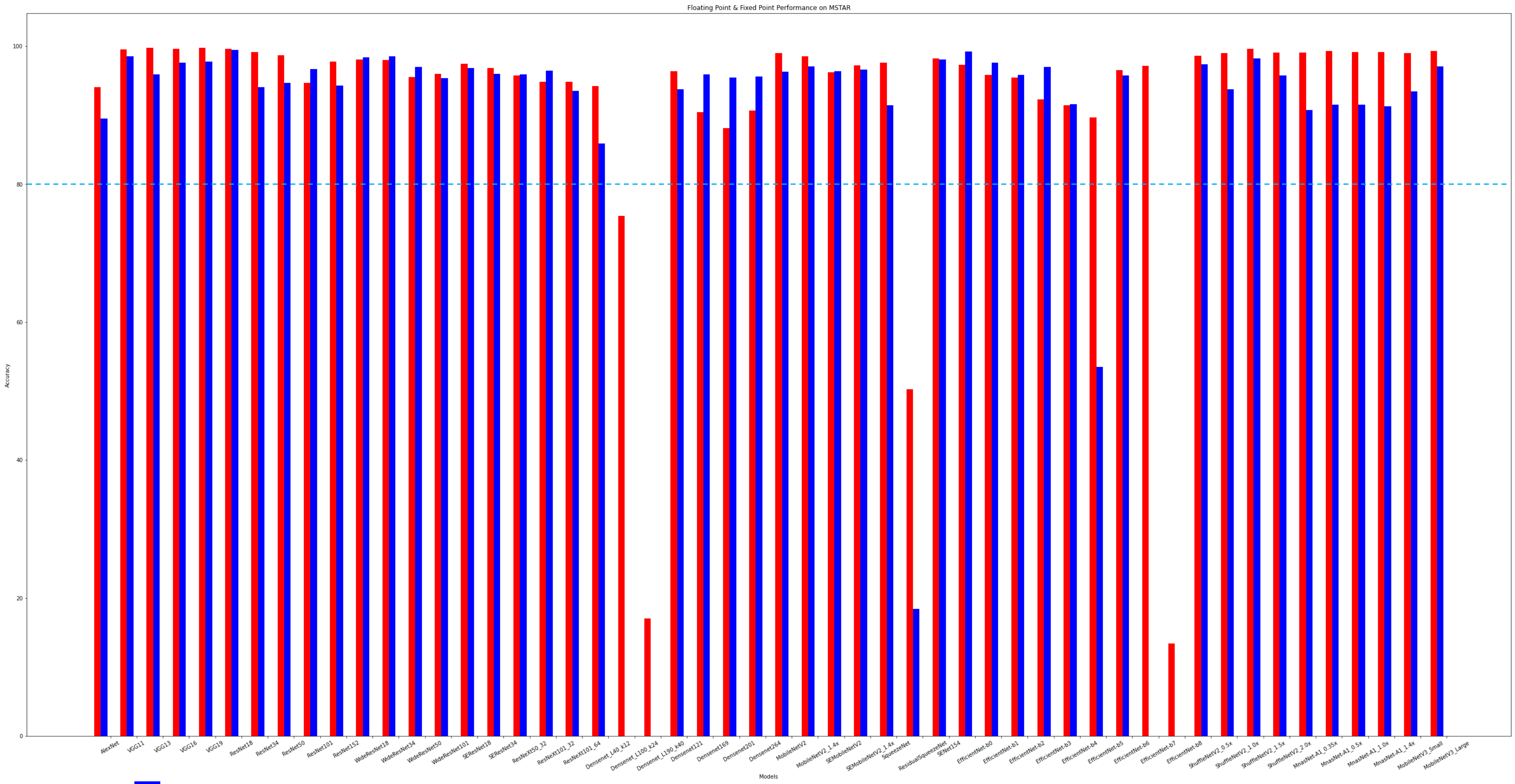
Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).

Results – Accuracy on MSTAR & SAMPLE



MSTAR SAMPLE

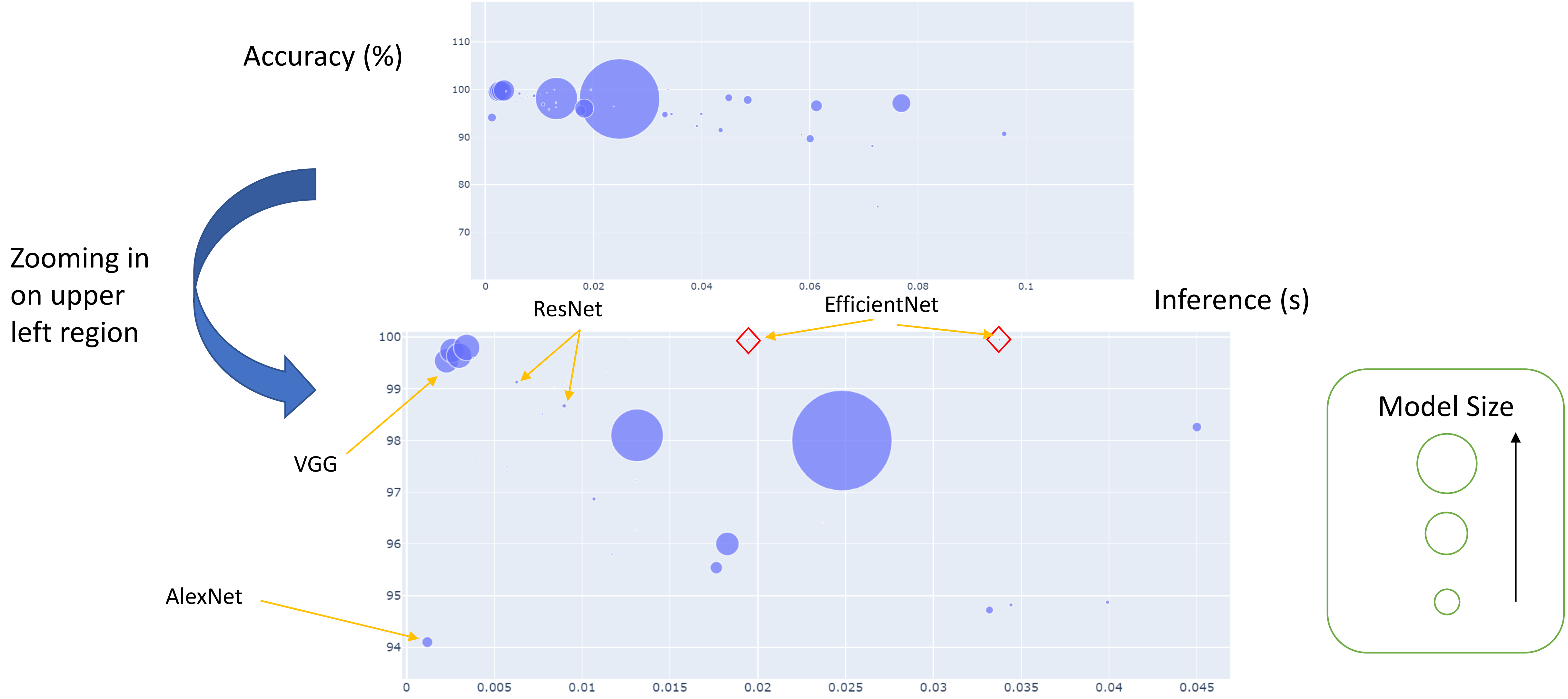
Results – FL vs FP on MSTAR



MSTAR SAMPLE

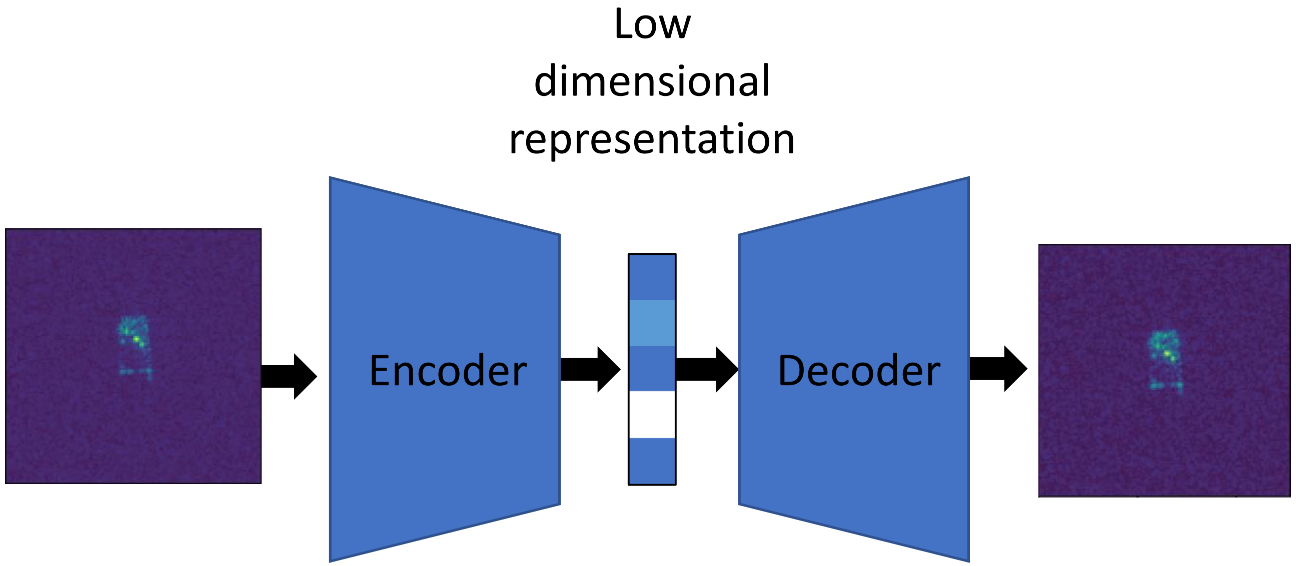
Performance Understanding (MSTAR)

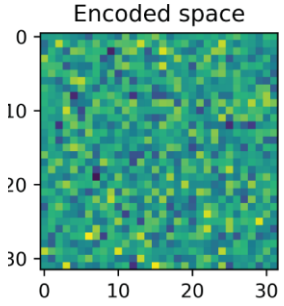
Performance understanding - considering latency (on GPU) & model size (memory impact)

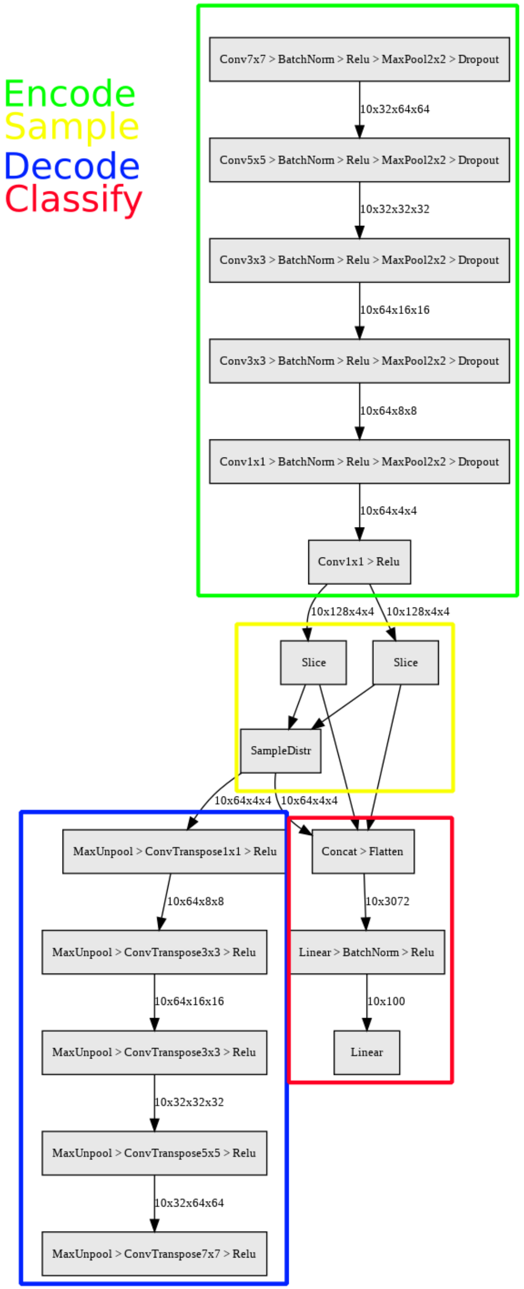


Variational Autoencoder (VAE)

Defined in 2013 by Kingma et al. and Rezende et al..

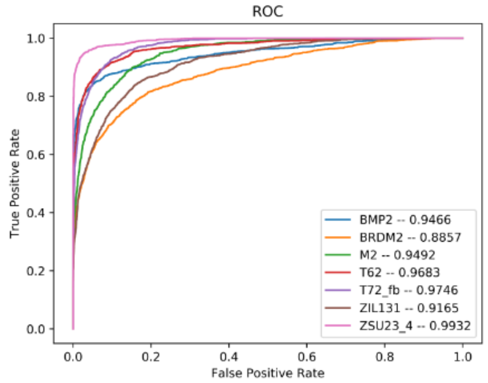


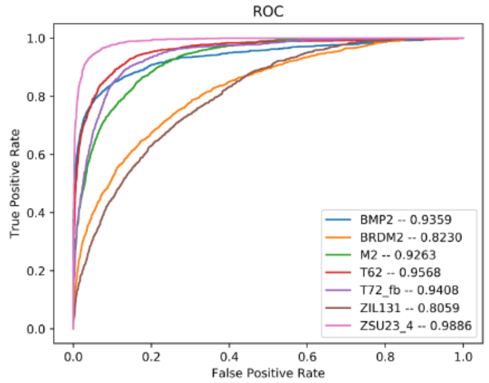




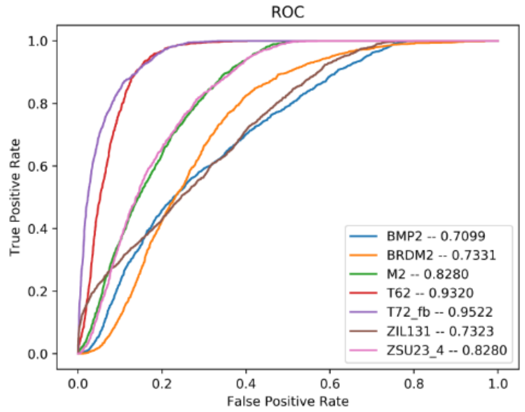
Variational Autoencoder (VAE) on MSTAR

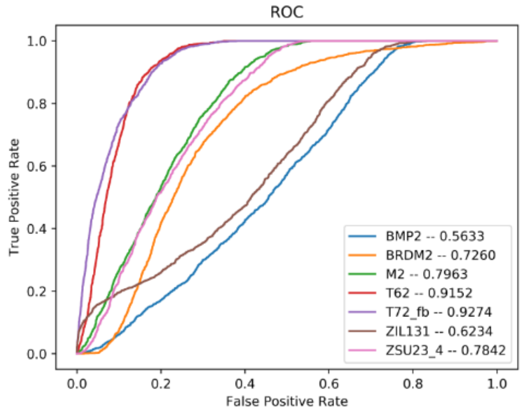
5 degrees **20 degrees**



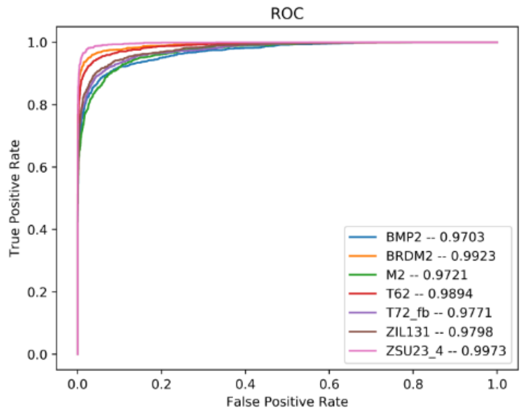


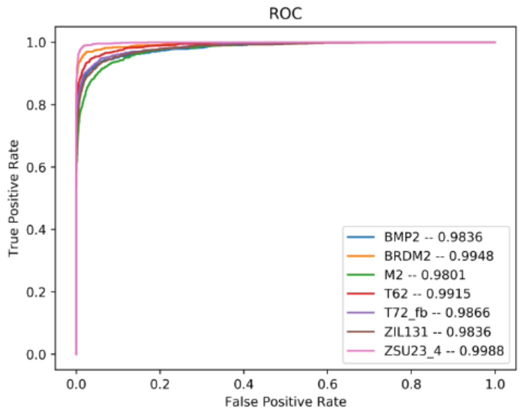
Baseline MPM – with masking



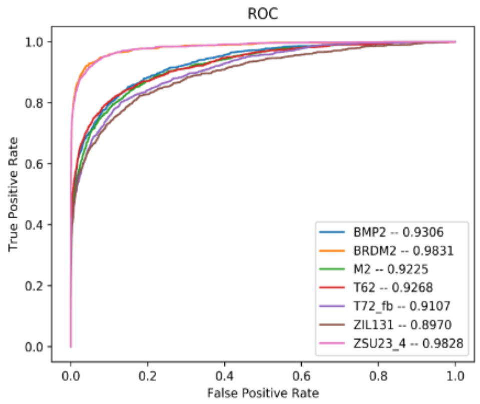


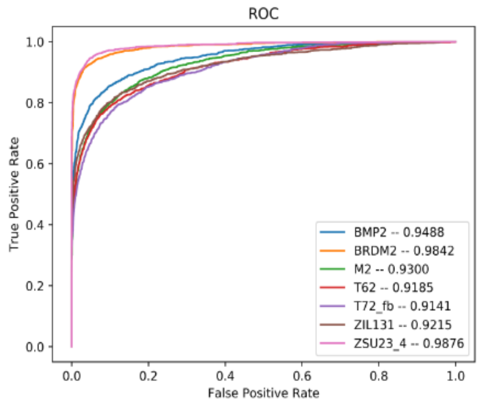
Baseline MPM – without masking





VAE





VAE – with shift

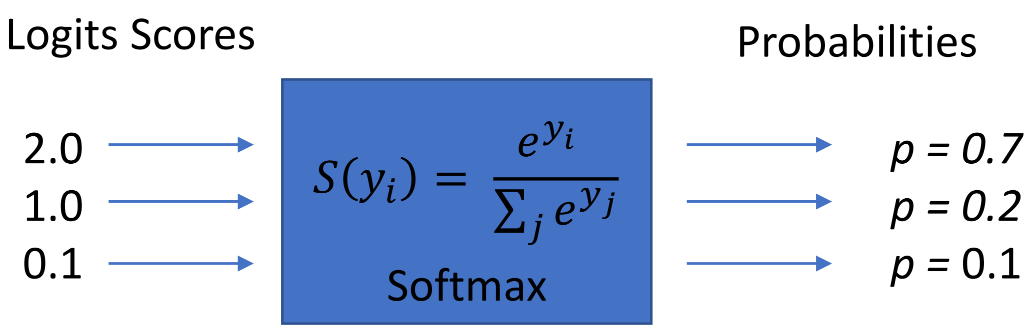
Additional Algorithmic Understanding

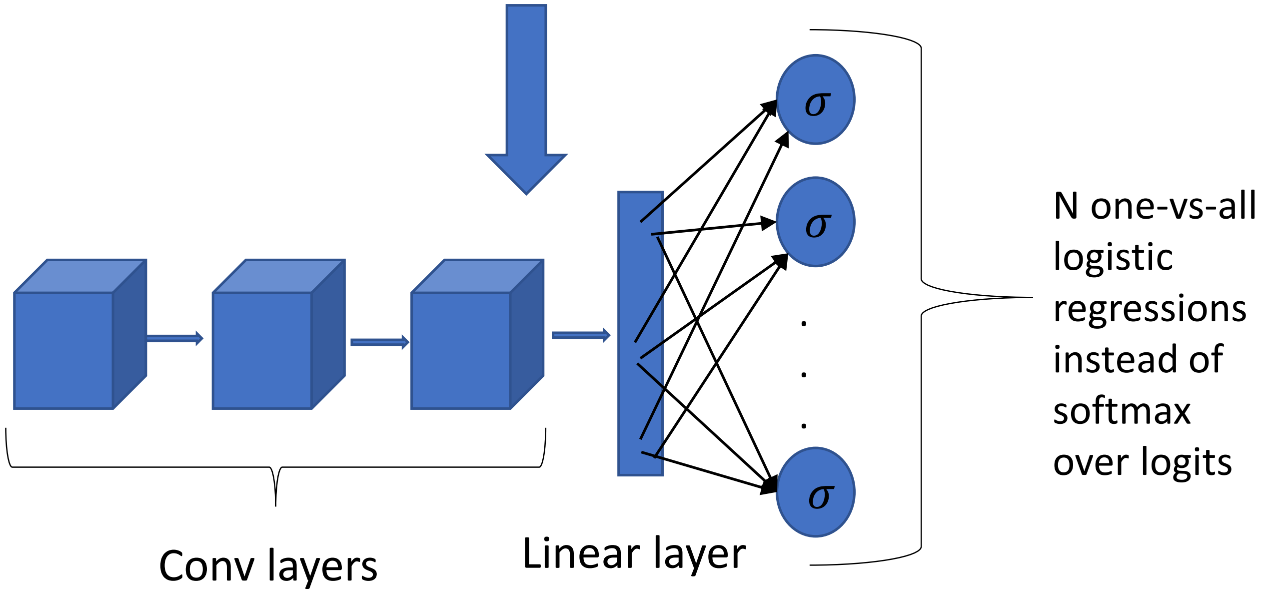
- Beyond accuracy and computational costs we’ve also begun exploring additional methods of insight into model behaviors

- Illustrative of techniques we’ve been exploring – but not intended to convey results or advocate for these methods

Logistic Regression Ensemble Models

- We are interested in per class accuracies and false positive/false negative rates
- This is unnatural to do with a standard softmax output in a DNN
- Idea: replace last layer of the DNN with N one-vs-all logistic regressions where the weights for each logistic regression are the weights for each output unit in the network.
- Convert each minibatch to N one-vs-all minibatches and train each logistic regression independently at every epoch with binary cross entropy.
- In inference, take max score. Works as good, if not better than softmax on MSTAR





Saliency Maps

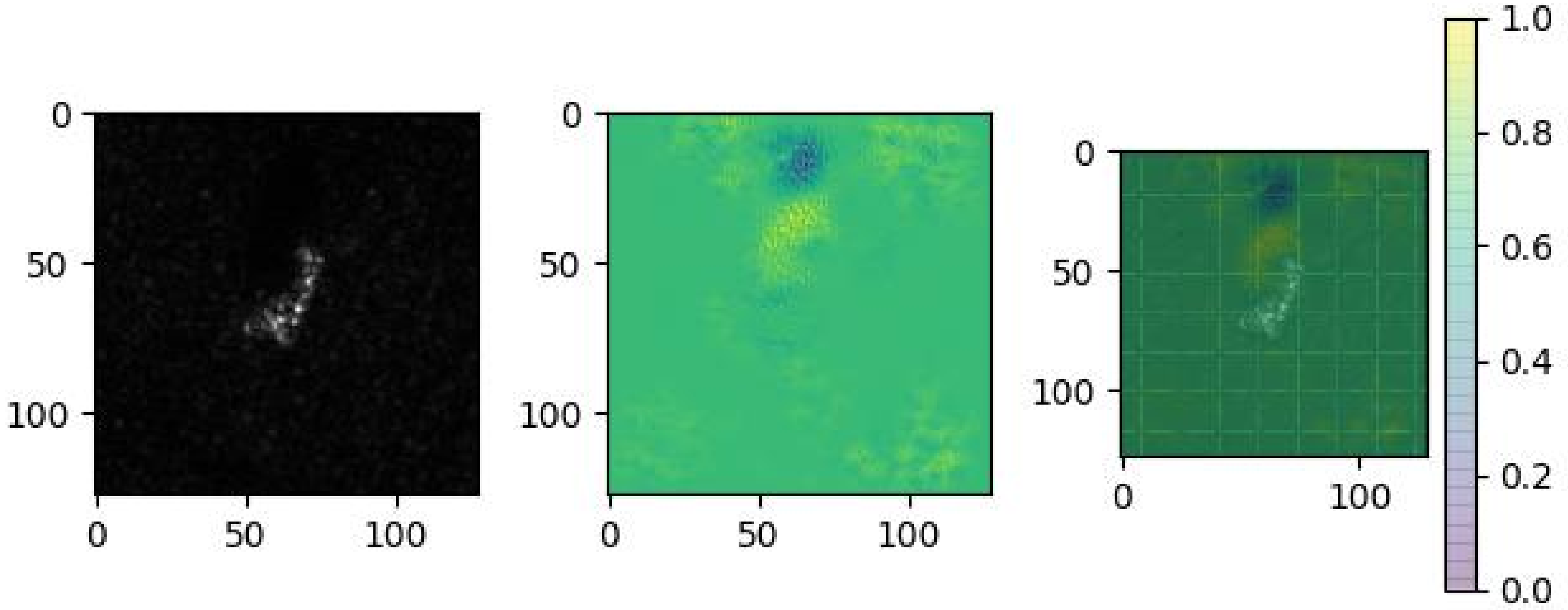
- Methods used for model explainability and interpretability

- Two techniques:
  - Vanilla gradient saliency maps (easy, applicable to any architecture)
  - Class activation maps (more complex, not applicable to any architecture, possibly more effective)

Vanilla Gradient

- Load up a pre-trained model, feed an example of interest through, backprop all the way back to the input pixels
- This gives you the gradient of the loss with respect to the input pixels
- Plot this as a normalized heatmap

Vanilla grad saliency map from VGG16



Class Activation Maps (CAM)

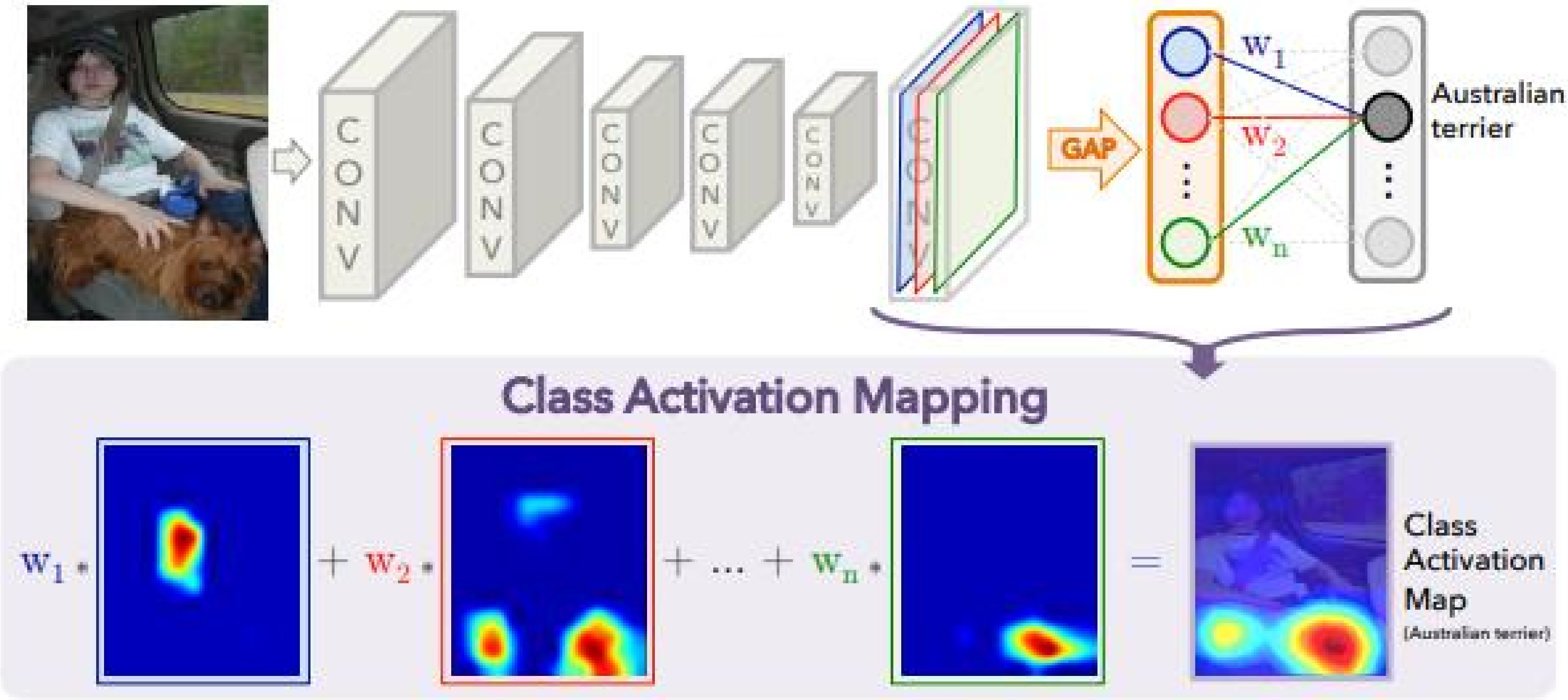
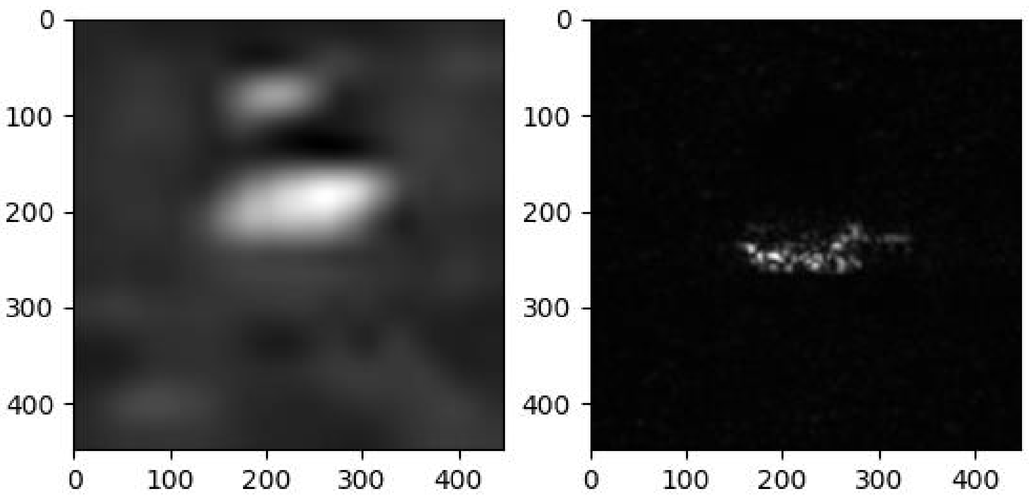


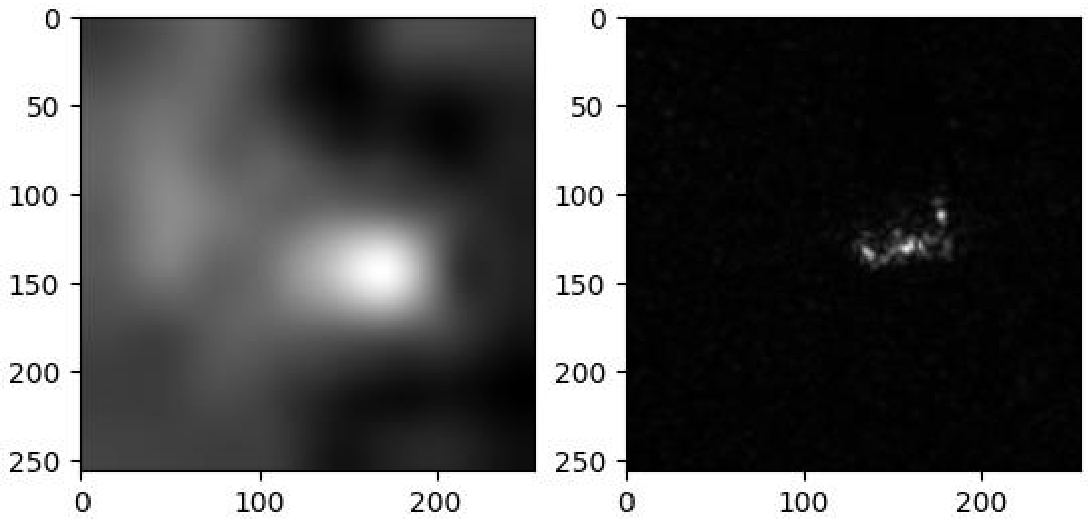
Figure 2. Class Activation Mapping: the predicted class score is mapped back to the previous convolutional layer to generate the class activation maps (CAMs). The CAM highlights the class-specific discriminative regions.

Zhou, Bolei, et al. "Learning deep features for discriminative localization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Class Activation Maps (CAM)



CAM from efficientnet-b0



CAM from resnet18

Conclusions

Beyond the algorithmic results shown here, we're also exploring computational costs of execution on neuromorphic hardware

- This includes a collaboration with Professor Naresh Shanbhag (UIUC) analyzing computational complexity costs balancing accuracy-representation-computation efficiency

Pursuing further understanding of the operation of neural network ATR approaches

- This includes exploring additional methods beyond saliency maps, ablation studies, analysis of activations, & analysis of representations/embeddings

Overall this research has enabled state-of-the-art performance in terms of accuracy as well as an understanding of associated computational costs while working towards understanding the neural computation

Thank you

Questions?