SAND2017-8366C

Convolutional Neural Networks for Automatic Threat Detection in Security X-Rays

Funding Source: DHS



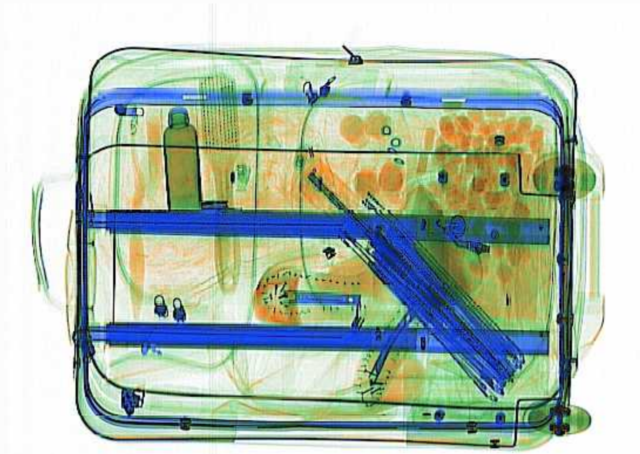
Trevor Morris/8754   Tiffany Chien/9365   Eric Goodman/9365

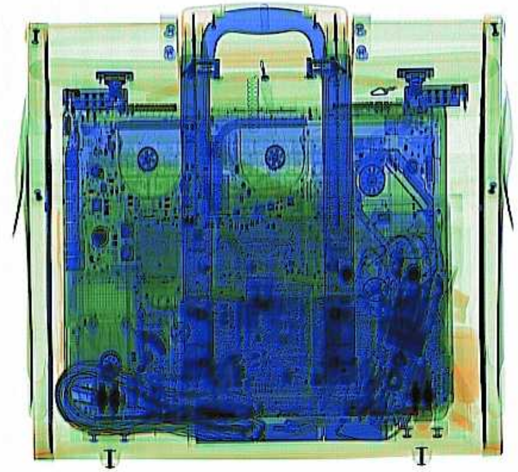
Abstract

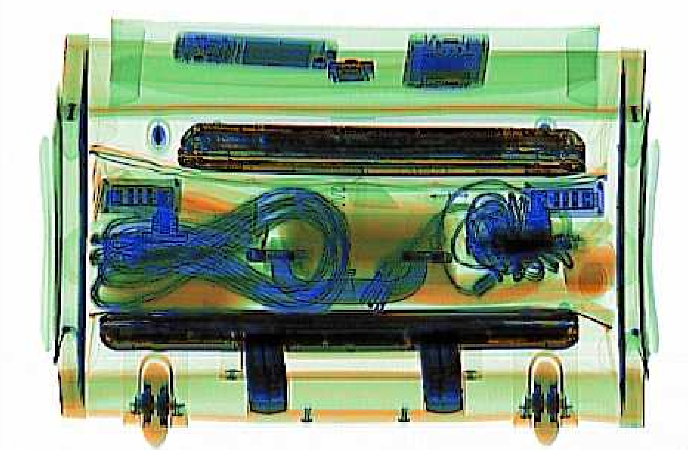
This project applies Convolutional Neural Networks (CNNs) to the task of automatic threat detection, specifically explosives, in security X-ray scans of passenger baggage. Our data preparation methods preserve important features of the images while making our data compatible with different models. Using three different prebuilt state-of-the-art models and taking advantage of the properties of the X-ray scanner, we achieve reliable detection of threats. We employ transfer learning and training set augmentation to overcome our relatively small dataset. We also use visualizations to help interpret and adjust the training of our models.

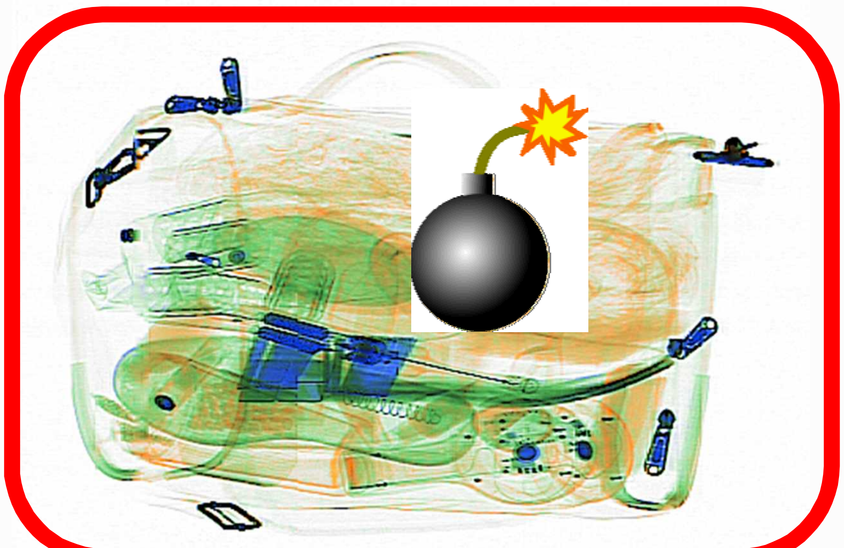
Problem you are trying to solve

Detect explosives in passenger baggage X-rays

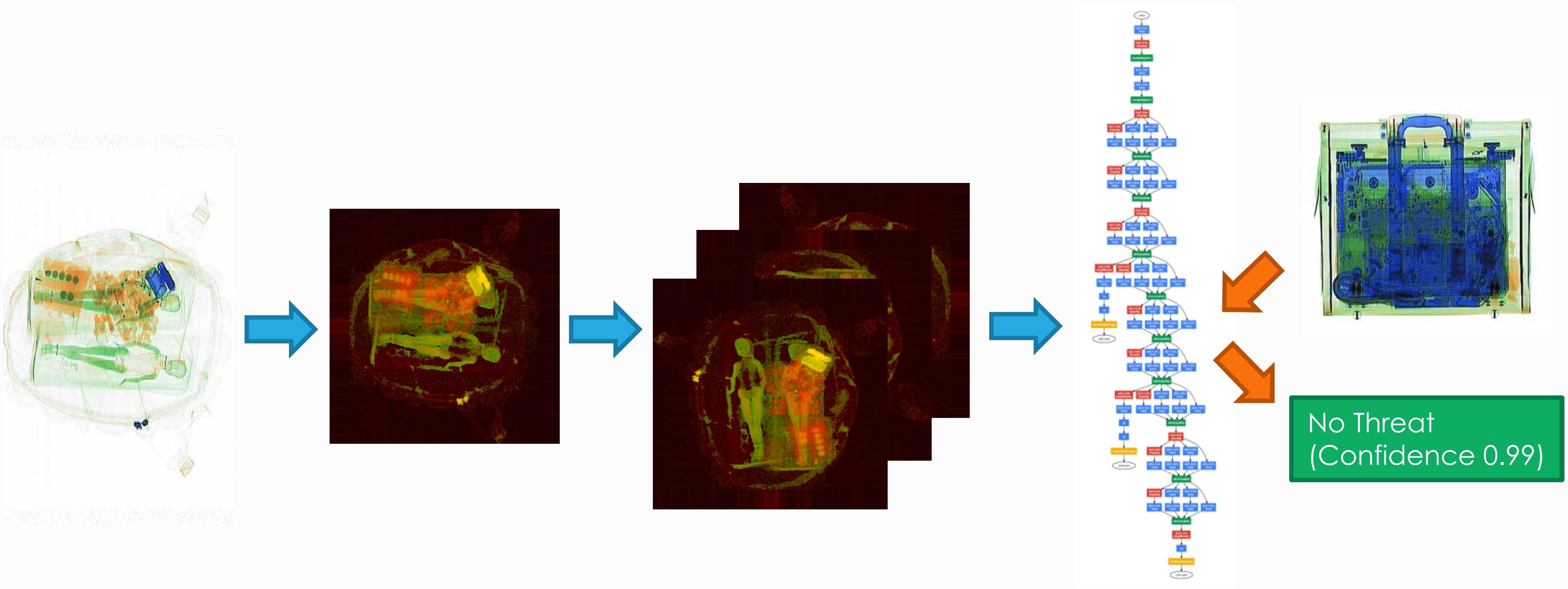








Algorithmic approach of your solution



Dataset of x-ray scans

Processing and normalization

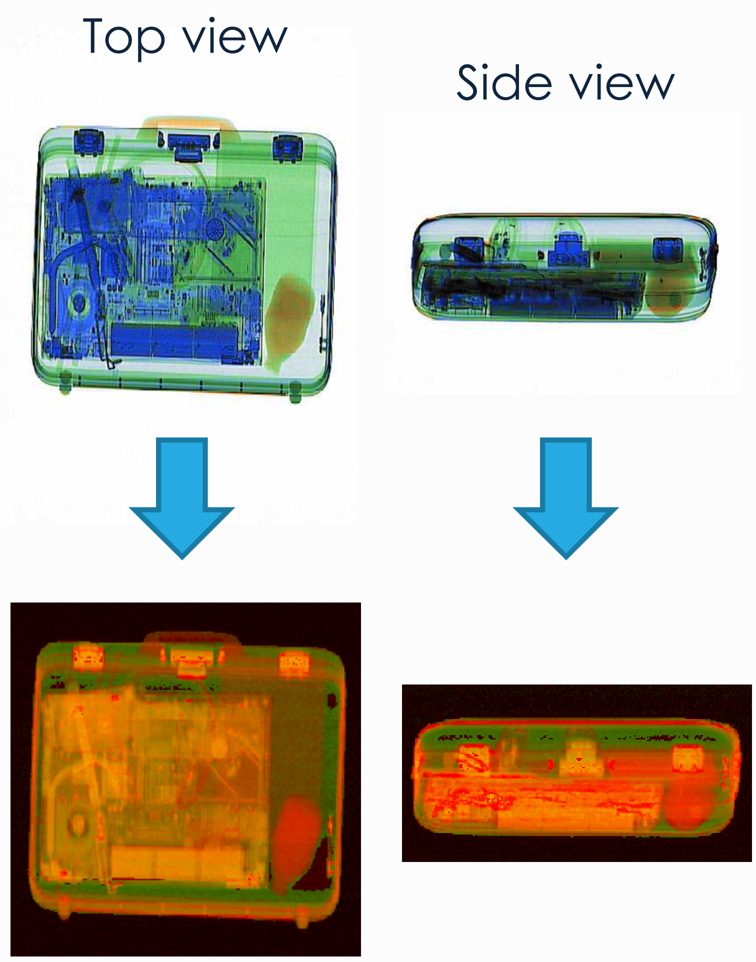
Data augmentation

State-of-the-art CNN pretrained on ImageNet data

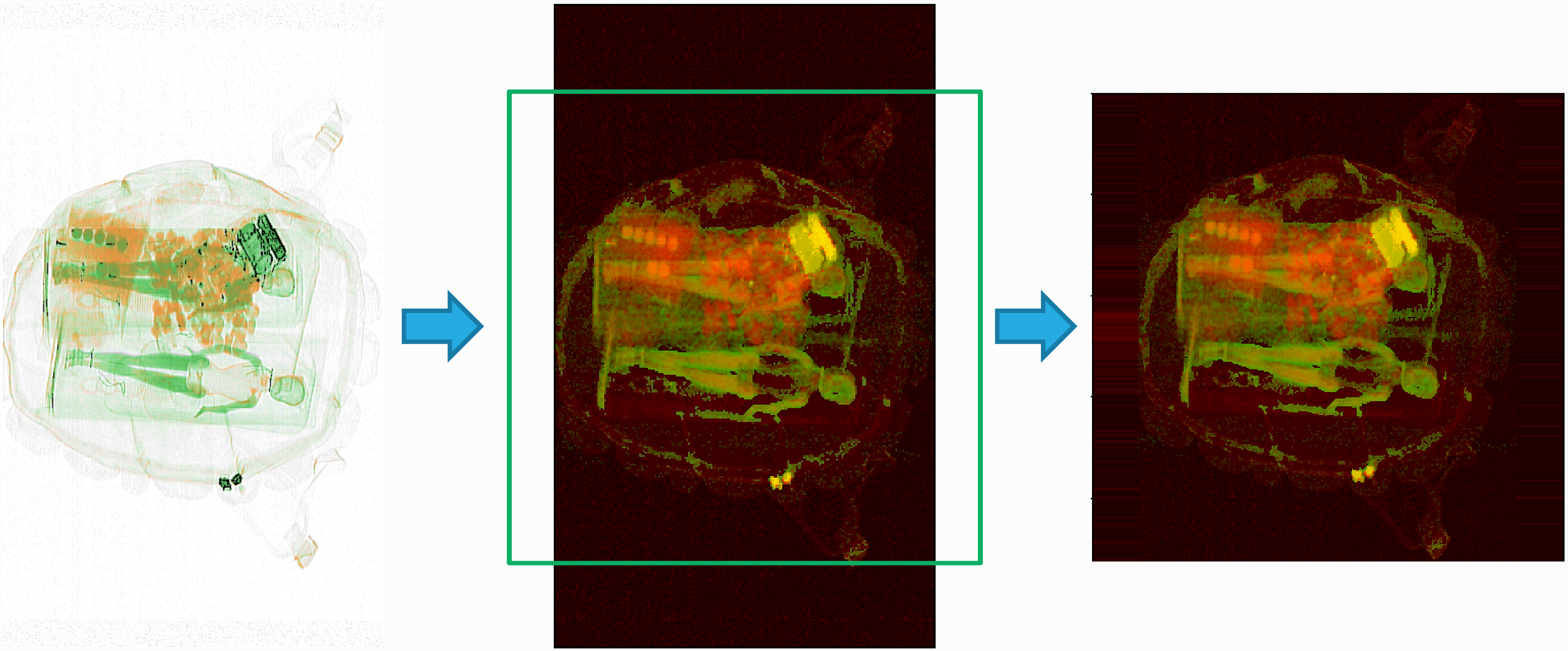
Predictions on novel data

Passenger Baggage Object Dataset (PBOD)

- Around 7000 scans in DICOS format
- Two channels:
  - Intensity (red)
  - Z-effective (green)
- Two views:
  - Top
  - Side
- Scans are repeated
  - Problematic



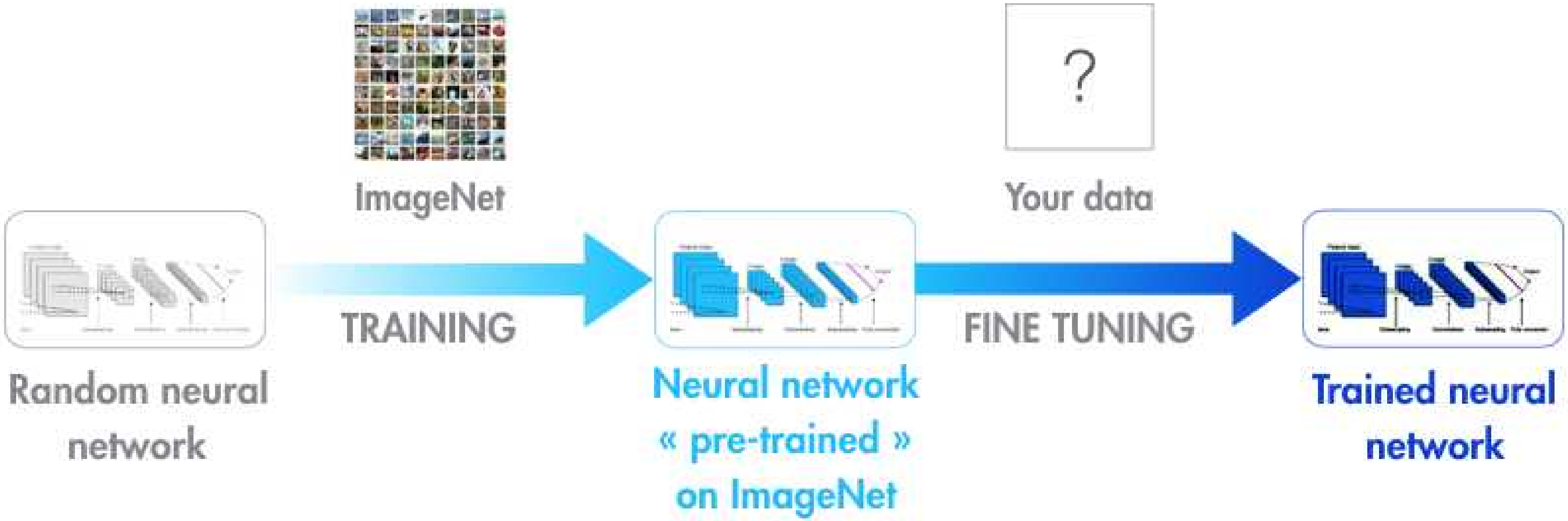
Data Processing



- Normalize channels to range from 0 to 1
- Crop to 299x299

Problem: Not a lot of data

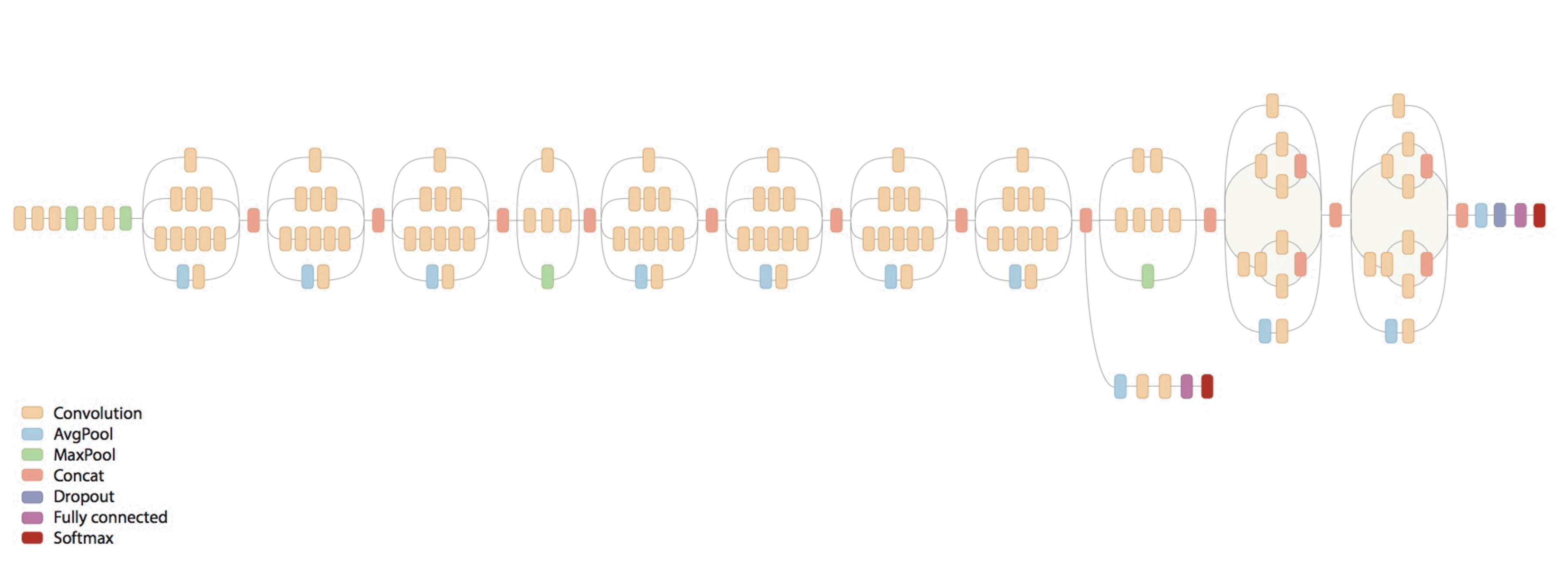
- Augmentation: Random flips, rotations, shifts, and zooms
- Transfer learning



Models

InceptionV3 (Google)

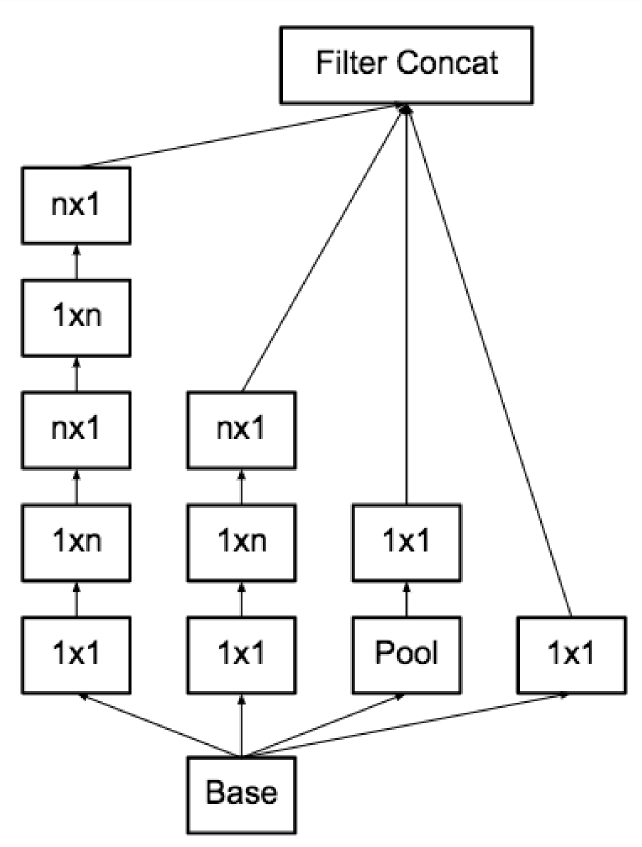
- 25 million parameters
- Batch normalization
- Inception modules



Models – InceptionV3 Inception Module

- "Multi-level feature extractor"

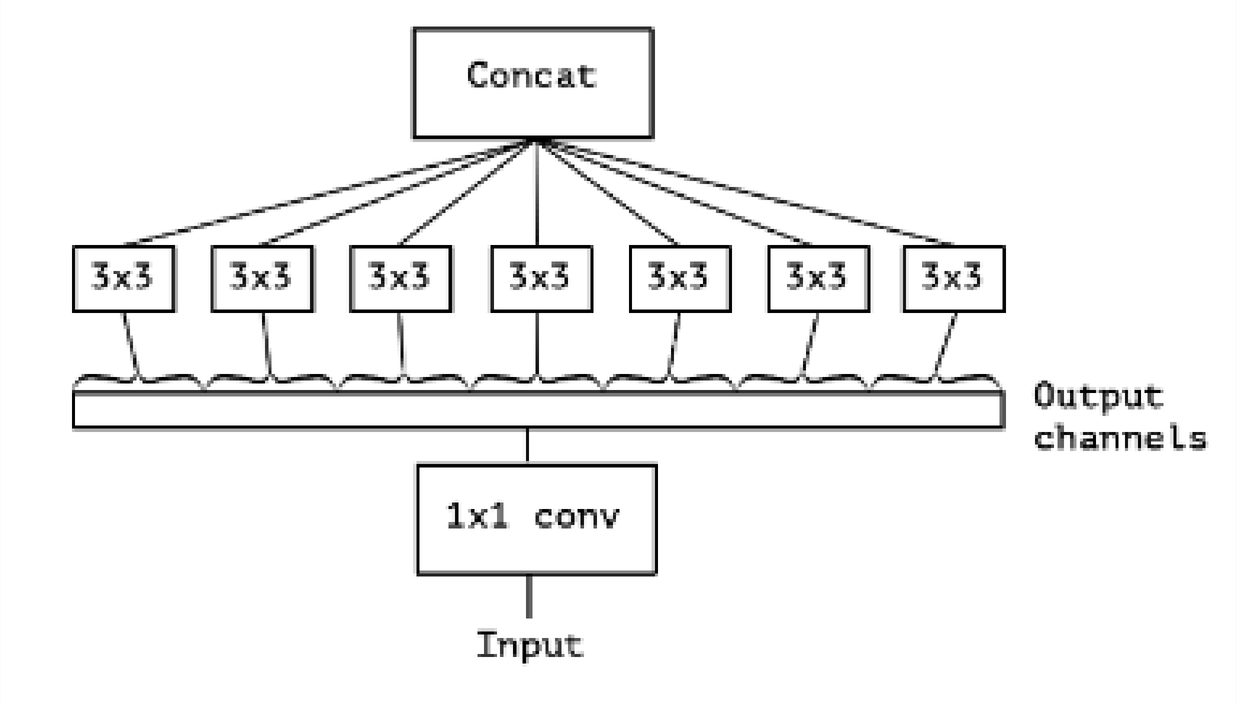
- Two performance increasing principles:
  - 5x5 convolution filters can be replaced by two consecutive 3x3 filters which achieves:
    - Same effective receptive input size and output depth
    - Less computationally expensive
  - 3x3 convolution filters can be replaced by consecutive asymmetric convolutions: 3x1 and then 1x3



Models

Xception (Google)

- Also 25 million parameters
- Replaced Inception modules with “Depthwise-separable convolutions”



Models

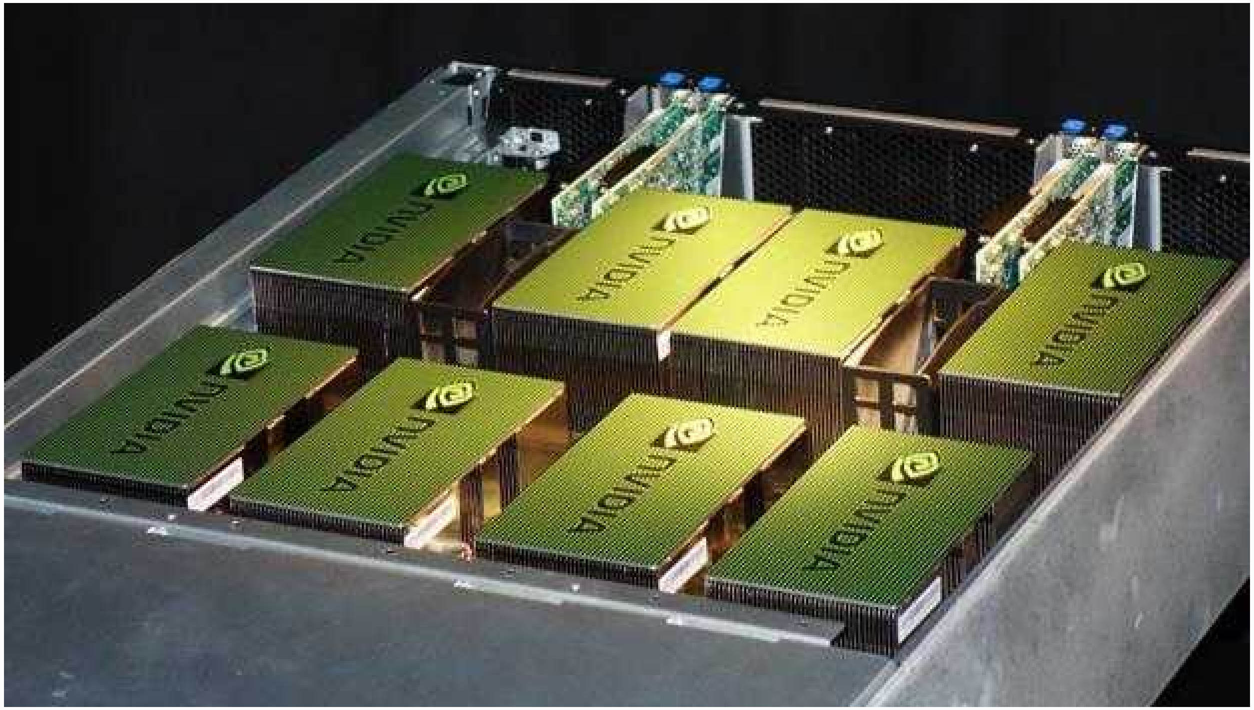
VGGNet 19 (Visual Geometry Group, Univ. of Oxford)

- 144 million parameters

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

NVIDIA DGX-1

- 8 Tesla P100s
- 28672 total CUDA cores, 16GB/GPU
- 2x 20-Core Intel Xeon CPU
- Thanks to Sandia HPC



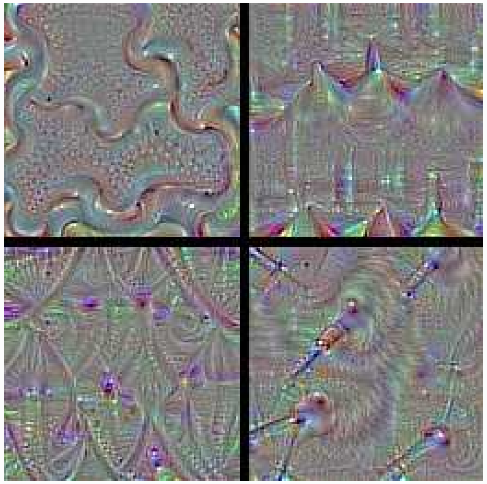
Results

Model	Single-view Accuracy (%)	Combined-view Accuracy (%)
InceptionV3	82.14	85.28
Xception	86.17	86.32
VGG19	**88.20**	88.69
InceptionV3 Pretrained	85.08	87.23
Xception Pretrained	87.00	87.63
VGG19 Pretrained	87.99	**90.60**

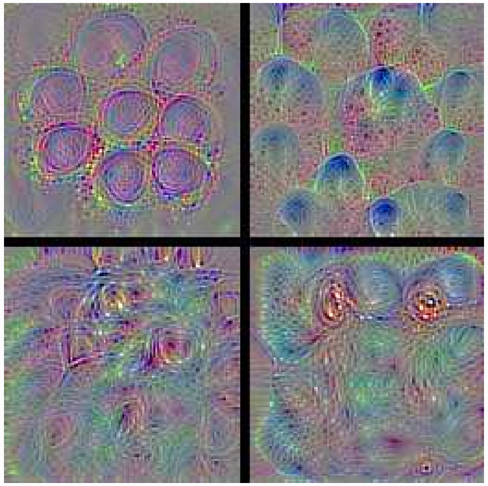
Visualizations – Maximal Activation

- Attempt to gain insight into how well our model is learning

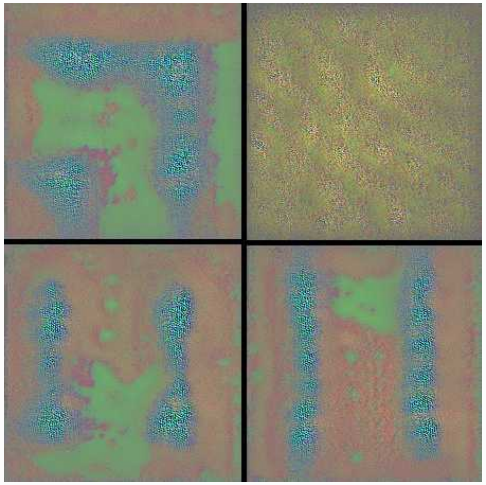
1. Take random noise and put it through the network
2. Measure how much a certain neuron activates, and using gradient ascent iteratively modify the noisy image to maximize activation



Pretrained on ImageNet



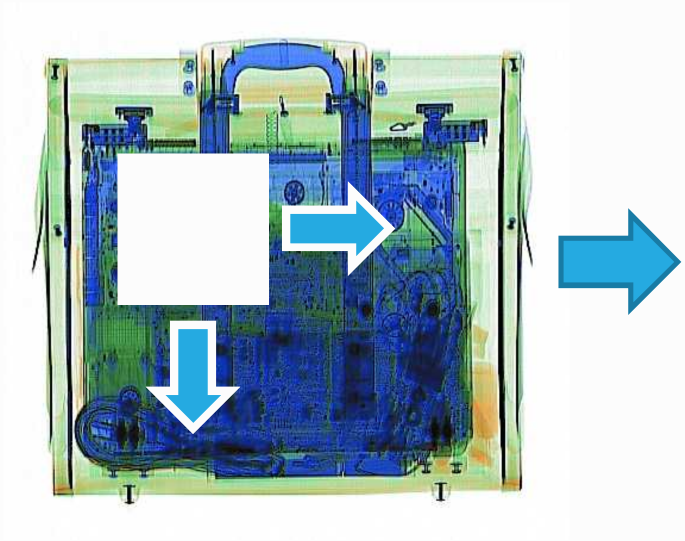
Retrained to our data

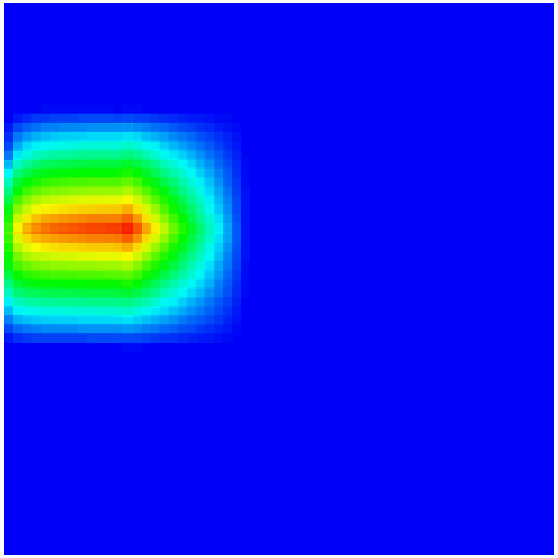


Trained from scratch

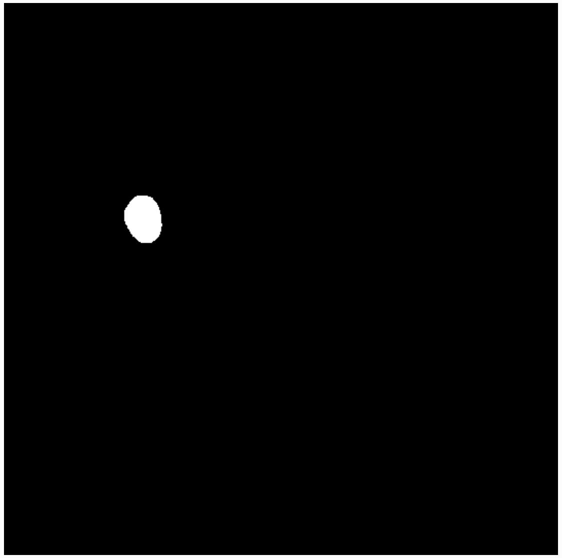
Visualization – Heatmaps

- Without any additional training or modification to the network, we applied a "sliding occluder" method to output a heat map which can show where detected threats are





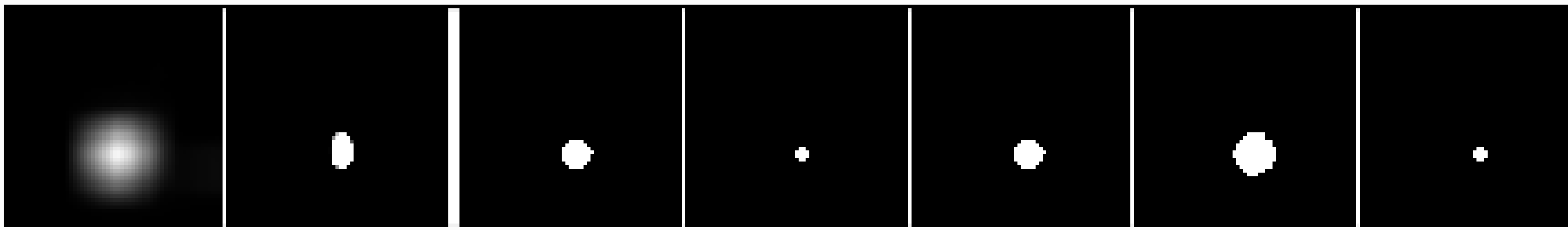
Heatmap of threat probabilities



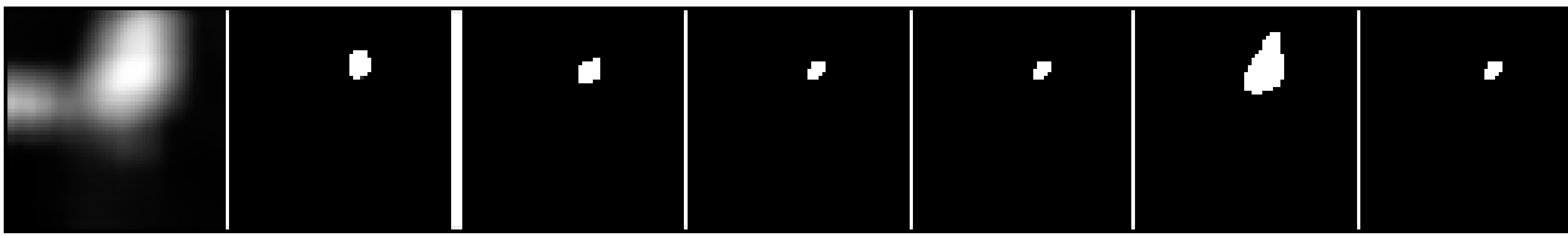
Ground truth location

Evaluation – Heatmaps

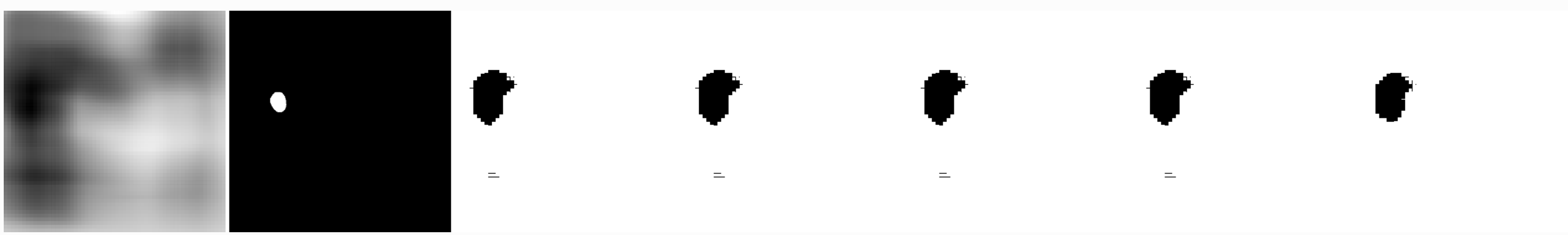
Heatmap	GT	F1	Precision	MCC*	BM*	MK*



		F1	Precision	MCC*	BM*	MK*
		0.8315	1.0	0.9152	0.9926	0.9953



		F1	Precision	MCC*	BM*	MK*
		0.6379	0.8519	0.8247	0.9872	0.9234



		F1	Precision	MCC*	BM*	MK*
		0.0099	0.0050	0.5059	0.5148	0.5025

* Modifed to scale from 0 to 1 instead of -1 to 1

Results - Heatmaps

Model	Avg F1	Avg PRE	Avg MCC	Avg BM	Avg MK	BM>0.7
INC	0.0881	0.0841	0.5530	0.6631	0.5393	0.3669
INC-P	0.1437	0.1581	0.5858	0.7361	0.5754	0.5556
XCP	0.1225	0.1301	0.5719	0.6964	0.5622	0.4393
XCP-P	0.2038	0.2381	0.6212	**0.8080**	0.6157	**0.7209**
VGG	0.0545	0.419	0.5331	0.6071	0.5183	0.2093
VGG-P	**0.2436**	**0.3027**	**0.6340**	0.7839	**0.6469**	0.6512

Conclusions

- Initial results indicate CNNs are effective in detecting explosives within x-rays images.
- Visualizations can provide insight into how to improve your models