Deep Neural Exposure:
You Can Run, But Not Hide Your Neural Network Architecture!

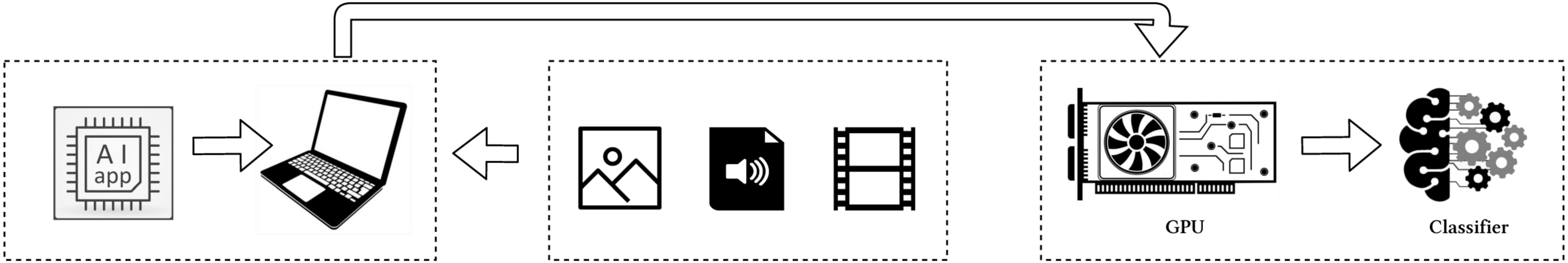
SAYED ERFAN AREFIN & DR ABDUL SERWADDA

TEXAS TECH UNIVERSITY, LUBBOCK, TEXAS

Neural Network Security

- Companies develop their own Neural Network
- Optimize networks
- Intellectual property
- Security of Neural Networks

Attack Scenario



1 Proprietary DNN app installed and running on personal computer

2 Malicious user providing inputs to repeatedly trigger DNN activity

3 GPU power measurement data from DNN activity feed to classifiers, which infers network architecture

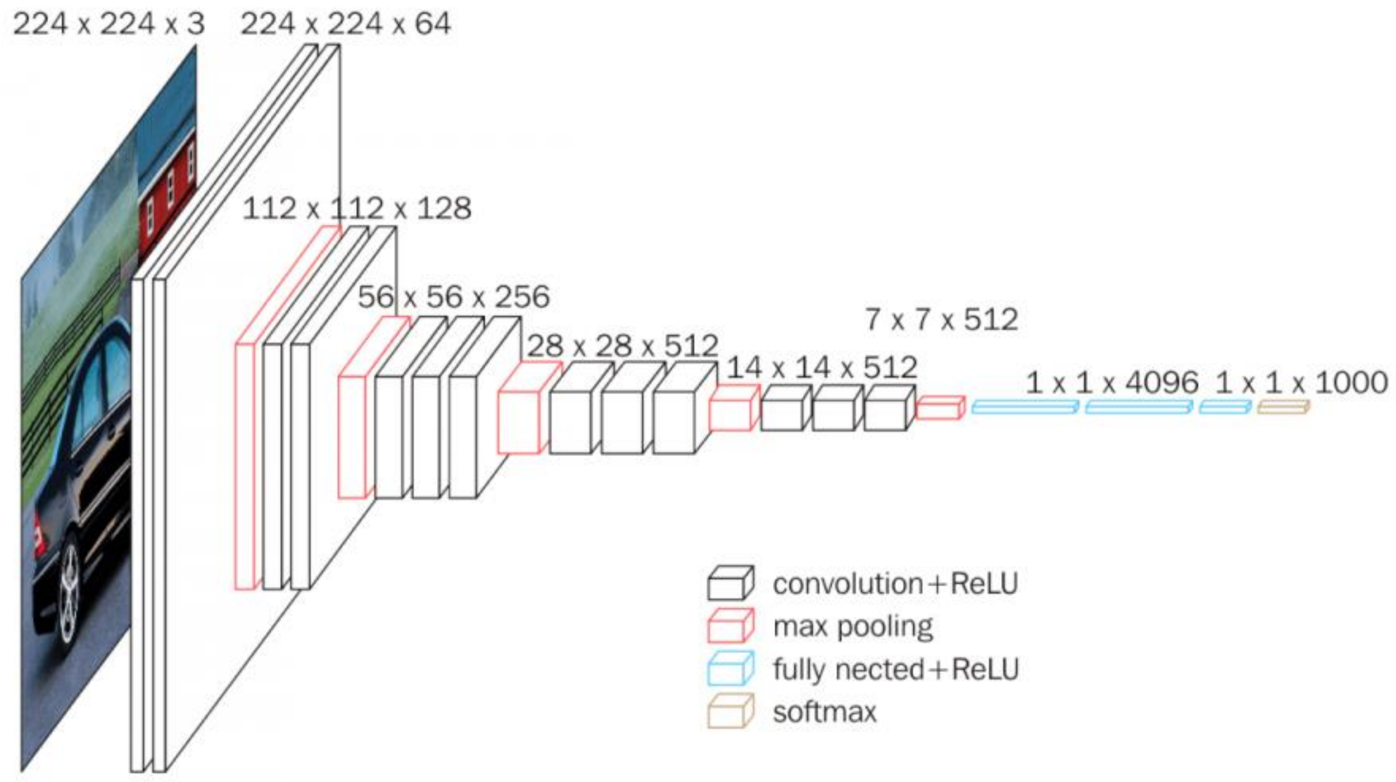
CNNs

IMAGENET

- The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale.
- Most of the CNNs used in this research are from ILSVRC.

CNN	Year
AlexNet	ILSVRC 2012
VGG Net	ILSVRC 2013
InceptionNet	ILSVRC 2014
ResNet	ILSVRC 2015
DenseNet	CVPR 2017
NASNet	Google 2018

VGG16 architecture

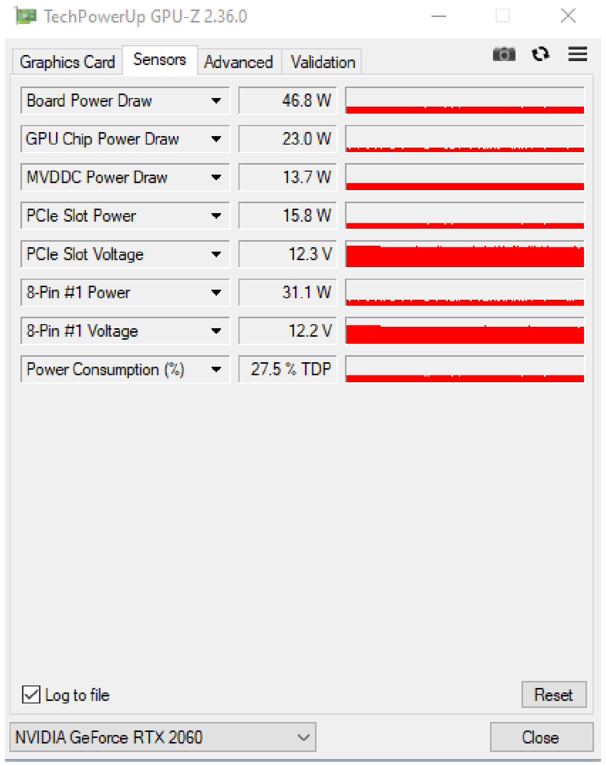


Comparison of CNN architectures

CNN Architecture	Architecture Layers			Network Properties				Accuracy	
	Conv	FC	PL	Input Size	Total Parameters	Dropout	Salient Feature	Top-5	Top-1
AlexNet	5	3	1	256x256x3	62,378,344	Used while training	Deeper	80.3%	57.2%
ResNet 18	17	1	2	224x224x3	11,511,784	None	Shortcut connections	90.58%	71.78%
SqueezeNet	14	0	4	256x256x3	1,248,424	50% after 9th fire module	Compressed	80.3%	57.5%
PolyNet	85	1	3	331x331x3	76.1 Million	0 to 25% Stochastic paths	Optimized deeper	95.75%	81.29%
NasNet Large	33	0	24	299x299x3	3.1 - 27.6 Million	50% on SoftMax.	Architectural search	96.2%	82.7%
VGG11	8	3	5	224x224x3	138,423,208	First 2 FC with 50%	Fixed-size kernels	90.4%	71.8%
Inception v4	38	0	5	299x299x3	43 Million	80% before SoftMax	Parallel kernels	95%	80%
Xception	36	1	5	299x299x3	22,855,952	50% before the Logistic regression layer	Extreme Inception	94.5%	79%
DenseNet 121	120	1	5	224x224x3	7.0 Million	20% after every Conv layer (except for the 1st)	Each layer connected to all	92.29%	74.98%
DPN 131	44	1	2	224x224x3	79.5 Million	None	New & reuse features	80.07%	94.88%

Power Measurement

- GPUz software by Techpowerup.
- Power sensors of GPU.
- Sampling rate: 100 samples/sec



TechPowerUp GPU-Z 2.36.0

Board Power Draw	46.8 W
GPU Chip Power Draw	23.0 W
MVDDC Power Draw	13.7 W
PCIe Slot Power	15.8 W
PCIe Slot Voltage	12.3 V
8-Pin #1 Power	31.1 W
8-Pin #1 Voltage	12.2 V
Power Consumption (%)	27.5 % TDP

☑ Log to file

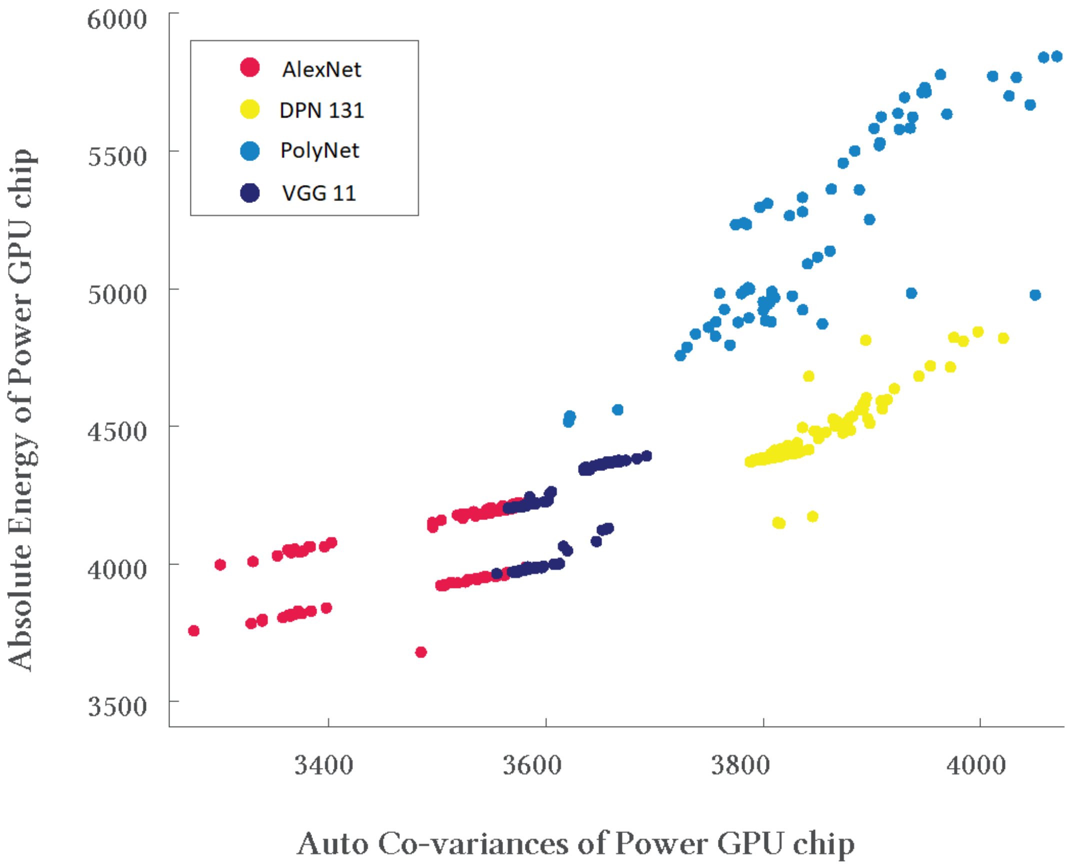
NVIDIA GeForce RTX 2060

Implementation

- Pre-trained CNNs.
- PyTorch provides pre-trained CNNs.
- Python to implement the experiment.
- Tested on Nvidia RTX 2060 GPU
- Data collected from GPU-z log file and cleaned.
- TSFresh Library.

Features	Sensors
Benford ABS Sum of Changes, Standard Deviation, Binned Entropy, Lempel Ziv Complexity, Variance, Abs Energy, Count Above Mean Minimum, Variation Co-efficient, Last Location of Minimum, C3, Last Location of Maximum, CidCe	All
Benford Correlation	All except MVDDC & PCIe Slot

A plot of 2 features for 4 of the architectures.



Datasets

Dataset-1

- Cats and dogs.
- Total 100 samples were used.

Dataset-2

- Random Images
- Total 100 samples were used.

Comparison of datasets

- 10 basic CNN architectures.
- Every CNN was tested with Dataset-1 & Dataset-2.
- Random Forest & Logistic Regression

	Logistic Regression	Random Forest
Dataset 1	62.96%	73.57%
Dataset 2	69.43%	82.81%

Attack Configuration

CNN Architecture	**Variants**
SqueezeNet	SqueezeNet 1, SqueezeNet 1.1
DPN	DPN68, DPN68B, DPN92, DPN98, DPN131
DenseNet	DenseNet121, DenseNet161, DenseNet169, DenseNet201
InceptionNet	BN InceptionNet, InceptionNet v3, InceptionNet v4
VGG	VGG11, VGG13, VGG16, VGG19, VGG11 BN, VGG13 BN, VGG16 BN, VGG19 BN
ResNet	ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, CaffeResNet101, FB ResNet152
NasNet	NASNet Mobile, NASNet-A Large, P-NASNet 5 Large

Attack Configuration

Attack Configuration	Description
1	Use the 10 core CNN architectures for training and testing the classifiers.
2	Use the 7 core CNN architectures (architectures that has variants) for training and all the 32 variants testing the classifiers.
3	Use the 32 variants of the CNN architectures for training and testing the classifiers.

Results

Attack Configuration	Average Accuracy
1	82.81%
2	42%
3	64.17%

Thank you!