A First Look at Deep Learning Apps on Smartphones [1]

Presented by James

Deep Learning

Apps

Who is using deep learning?

Types of deep learning apps

- **Image recognition:** Facebook, Google photos
- **Video recognition:** TikTok

- **Text analysis:** Google translate
- **Voice analysis:** Siri

- **Recommendation engine:**

  Google / YouTube Ads

Background

The Dawn - 2017

Most major venders launched their DL frameworks for smartphones.

- Google: TensorFlow Lite (Nov. 2017)
- Facebook: Caffe2 (Apr. 2017)
- Apple: Core ML (Jun. 2017)
- Tencent: Ncnn (Jul. 2017)
- Baidu: MDL (Sep. 2017) [1]

Same Goal:

Executing the DL inference task **solely** on smartphones.

Technical terms

DL models / frameworks

DL models:

Comprise neuron layers in different types.

Convolutional Neural Network (CNN)

Recurrent neural network (RNN)

DL frameworks:

DL frameworks produce DL models and execute the models over input data

Think it as a template for developing DL apps.

Research motivation

Deep learning on smartphones is relatively new, what is the progress now?

Current Situation of smartphone apps market

How to bridge the gap between research and development?

Need some tools to analyse the deep learning usage in mobile apps.

Research questions

Key Questions

1. The **characteristics** of apps that have adopted Deep learning
2. The **roles** of deep learning in those apps
3. Current **types** of DL frameworks adopted in apps

Research method

Scope:

Android apps (88% of market share in 2018)

Data:

Apps on Google play store (APK and reviews)

16,500 apps in total

Two datasets at June and September 2018

Tools:

Developed an analysing tool (DL Sniffer)

Manual inspection

Workflow

Searches for specific strings or class/method names

e.g. TensorFlow always have
"TF_AllocateTensor" in its rodata section.

Extracts DL models from those DL apps

Researchers filter false positive
through reverse engineering

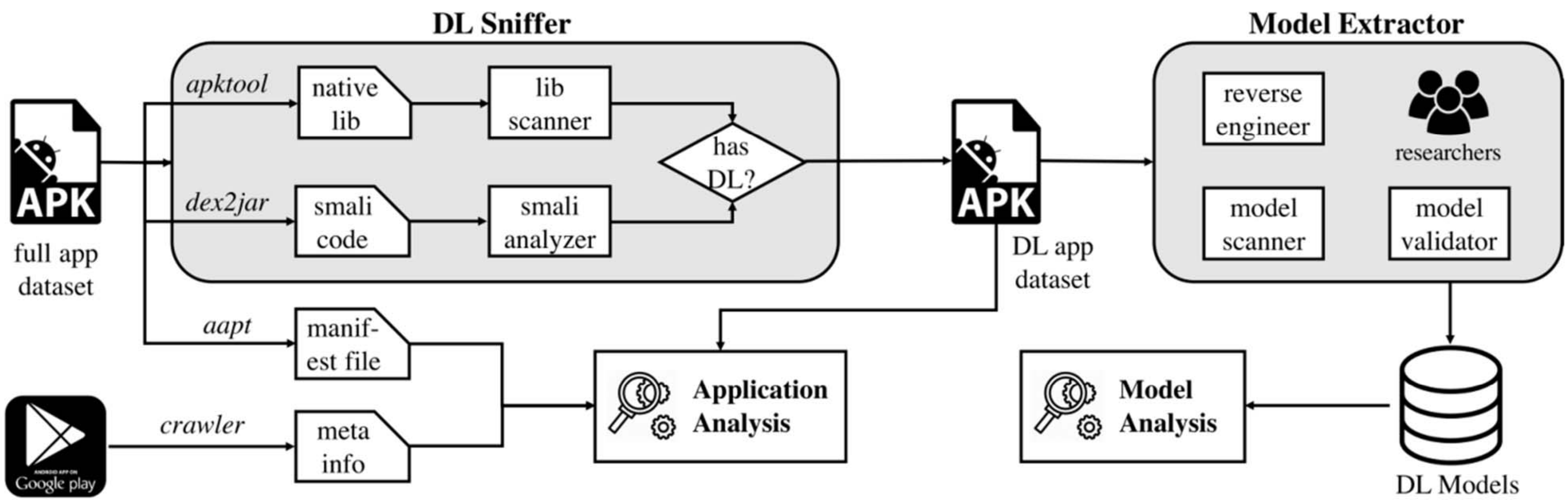
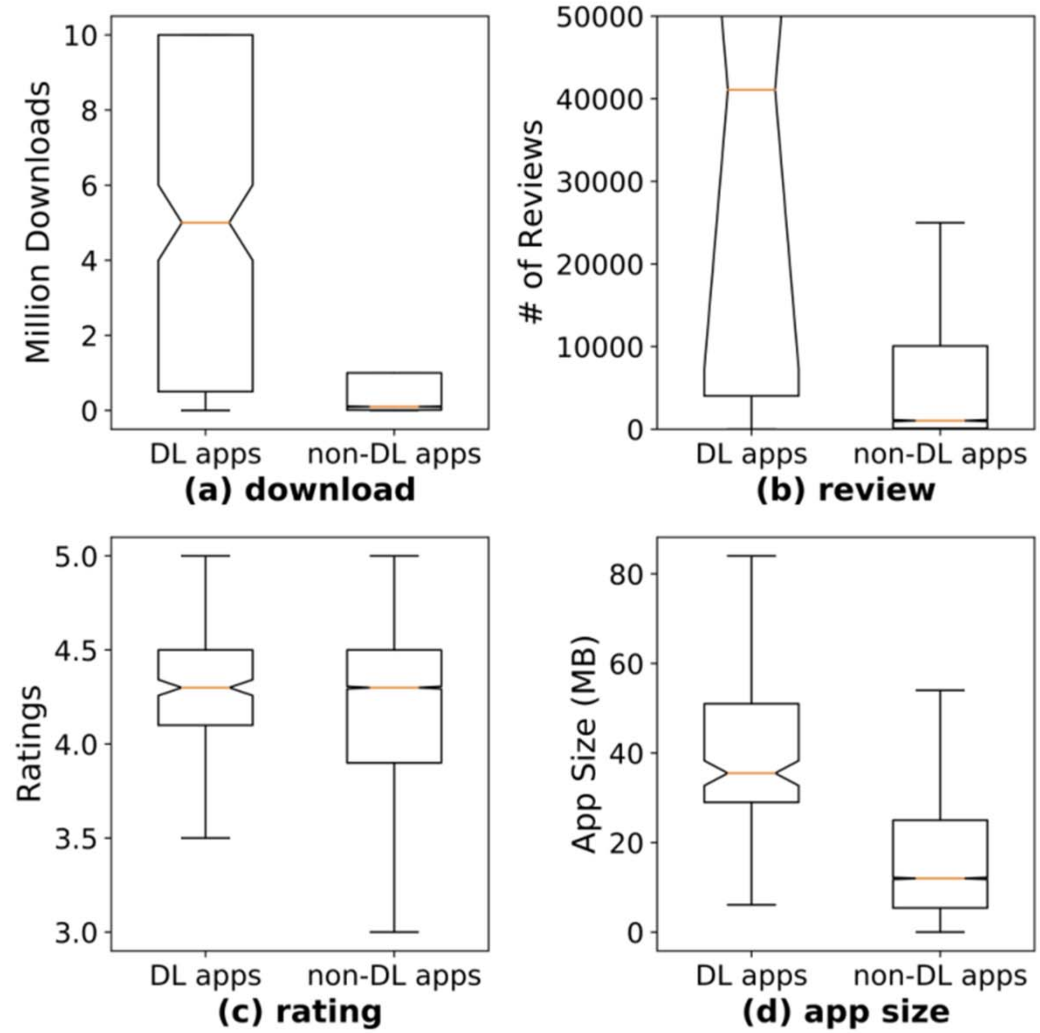


Figure 1: The overall workflow of our analyzing tool.

Characteristics of DL Apps

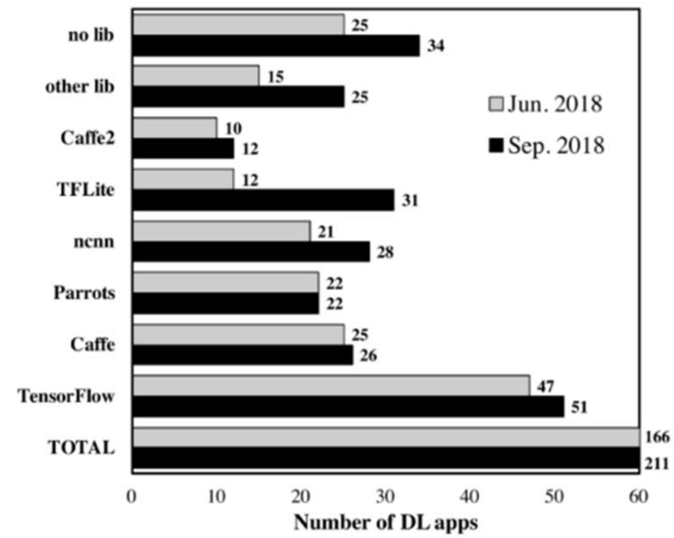
- More downloads
- More reviews
- More centred in rating
- Larger app size



Findings

DL is gaining popularity on smartphones

Big companies own more DL apps



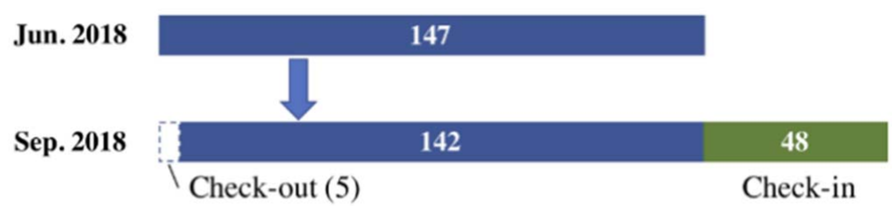


Figure 5: The number of check-in and check-out DL apps.

Roles of DL

- Image
- Text
- Audio
- Other

Becoming the core function

usage	detailed usage	as core feature
image: 149	photo beauty: 97	94 (96.9%)
	face detection: 52	44 (84.6%)
	augmented reality: 19	5 (26.3%)
	face identification: 8	7 (87.5%)
	image classification: 11	6 (54.5%)
	object recognition: 10	9 (90%)
	text recognition:11	4 (36.3%)
text:26	word&emoji prediction: 15	15 (100%)
	auto-correct: 10	10 (100%)
	translation: 7	3 (42.8%)
	text classification: 4	2 (50%)
	smart reply: 2	0 (0%)
audio: 24	speech recognition: 18	7 (38.9%)
	sound recognition: 8	8 (100%)
other: 19	recommendation: 11	2 (18.1%)
	movement tracking: 9	4 (44.4%)
	simulation: 4	4 (100%)
	abnormal detection: 4	4 (100%)
	video segment: 2	1 (50%)
	action detection: 2	0 (0%)
total: 211		171 (81.0%)

Popular frameworks

- TensorFlow (51)
- TFLite (31)
- Ncnn (28)
- Caffe (26)
- Parrots (22)
- Caffe2 (12)

DL framework is gaining traction

Mobile DL ecosystem is forming

Framework	Owner	Supported Mobile Platform	Mobile API	Is Open-source
TensorFlow [36]	Google	Android CPU, iOS CPU	Java, C++	✓
TF Lite [37]	Google	Android CPU, iOS CPU	Java, C++	✓
Caffe [65]	Berkeley	Android CPU, iOS CPU	C++	✓
Caffe2 [9]	Facebook	Android CPU, iOS CPU	C++	✓
MxNet [46]	Apache Incubator	Android CPU, iOS CPU	C++	✓
DeepLearning4J [13]	Skymind	Android CPU	Java	✓
ncnn [35]	Tencent	Android CPU, iOS CPU	C++	✓
OpenCV [26]	OpenCV Team	Android CPU, iOS CPU	C++	✓
FeatherCNN [16]	Tencent	Android CPU, iOS CPU	C++	✓
PaddlePaddle [24]	Baidu	Android CPU, iOS CPU & GPU	C++	✓
xNN [40]	Alibaba	Android CPU, iOS CPU	unknown	✗
superid [34]	SuperID	Android CPU, iOS CPU	unknown	✗
Parrots [30]	SenseTime	Android CPU, iOS CPU	unknown	✗
MACE [23]	XiaoMi	Android CPU, GPU, DSP	C++	✓
SNPE [31]	Qualcomm	Qualcomm CPU, GPU, DSP	Java, C++	✗
CNNDroid [70]	Oskouei et al.	Android CPU & GPU	Java	✓
CoreML [12]	Apple	iOS CPU, GPU	Swift, OC	✗
Chainer [10]	Preferred Networks	/	/	✓
CNTK [22]	Microsoft	/	/	✓
Torch [39]	Facebook	/	/	✓
PyTorch [29]	Facebook	/	/	✓

DL Model analysis

- Model/Layer types
- Resource Footprint
- Optimizations
- Security

CNN: 87.7%
RNN: 7.8%
Unknown: 4.5%

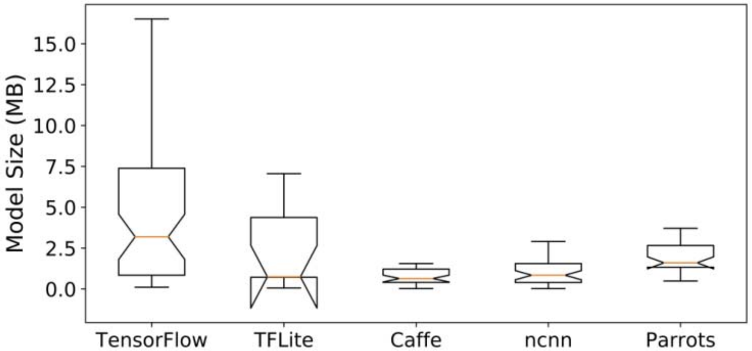
Layer type	% of models	# in each model	Layer type	% of models	# in each model
conv	87.7	5 / 14.8	relu	51.0	6 / 16.3
pooling	76.5	2 / 2.8	split	46.9	1 / 7.5
softmax	69.1	1 / 1.1	prelu	32.1	4 / 4.6
fc	60.5	3 / 5.6	reshape	28.4	2 / 24.1
add	56.8	9.5 / 23.8	dropout	21.0	1 / 1.0

Convolution (conv) is dominant, a core layer of CNN models

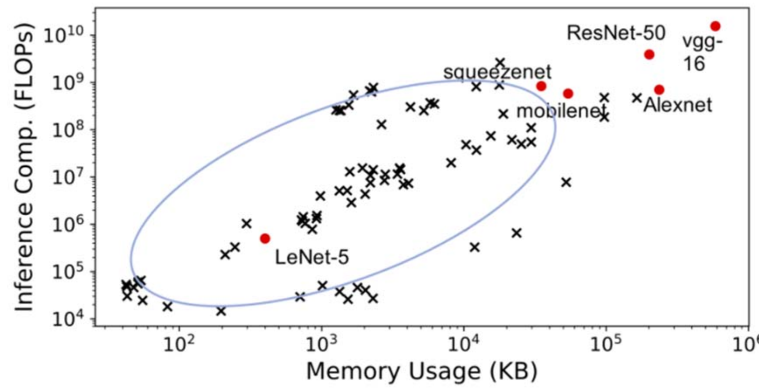
DL Model analysis

- Model/Layer types
- Resource Footprint
- Optimizations
- Security

Model Size



Memory Cost



DL models are very lightweight
running such models are inexpensive

DL Model analysis

- Model/Layer types
- Resource Footprint
- Optimizations
- Security

Quantization
Reduce the number of bits
Sparsity [2]
Reduce parameters

Table 4: Optimizations applied on DL models.

	1-bit Quan.	8-bit Quan.	16-bit Quan.	Sparsity
TF	unsupported	4.78%	0.00%	0.00%
TFLite	unsupported	66.67%	unsupported	unsupported
Caffe	unsupported	0.00%	unsupported	unsupported
Caffe2	unsupported	0.00%	unsupported	unsupported
ncnn	unsupported	0.00%	unsupported	unsupported
Total	*0.00%*	*6.32%*	*0.00%*	*0.00%*

Lack of DL optimizations

DL Model analysis

- Model/Layer types
- Resource Footprint
- Optimizations
- Security

Obfuscation 47/120(39.2%)
Shallow approach
Remove any meaningful text
Encryption 23/120(19.2%)
Better approach
Hide model structures. Decrypt at runtime

Findings:

Only **few frameworks** support obfuscation
Ncnn/Mace covert model to binaries or C++ codes

No framework provides supports in model encryption

Overall findings

- DL is gaining popularity on smartphones
- DL is becoming the core function of mobile apps
- DL framework is gaining traction
- DL models right now are very lightweight
- Well-known DL optimizations are missing in current apps
- Security issue is not handled in current apps, most DL framework doesn’t support Obfuscation Encryption

Limitations

Analysing tool only detects 16 DL frameworks

Small sample size

Only analysed on Android platform

Conclusions

- They have developed a new analysing tool to bridge the gap between research and practice.
- The paper revealed the current situation and the trend regarding DL apps in the real market.
- They pointed out that security and optimisations could be the major issue for current DL apps.

References

[1] M. Xu, J. Liu, Y. Liu, F.X. Lin, Y. Liu and X. Liu, “A First Look at Deep Learning Apps on Smartphones,” in Proc. of the ACM World Wide Web Conference, 2019, pp. 2125-2136, doi: 10.1145/3308558.3313591.

[2] Wei Wen, Chunpeng Wu, Yandan Wang, Yiran Chen, and Hai Li. 2016. Learning structured sparsity in deep neural networks. In Advances in Neural Information Processing Systems. 2074–2082.