

A First Look at Deep Learning Apps on Smartphones

Mengwei Xu¹, Jiawei Liu¹, Yuanqiang Liu¹

Felix Xiaozhu Lin², Yunxin Liu³, Xuanzhe Liu¹

¹Peking University, ²Purdue University, ³Microsoft Research



Deep learning is popular



Google is using deep learning to change the world!
- summary of 1st Keynote

- Not only in CV, NLP, Robotics, but also Economics, Biology...



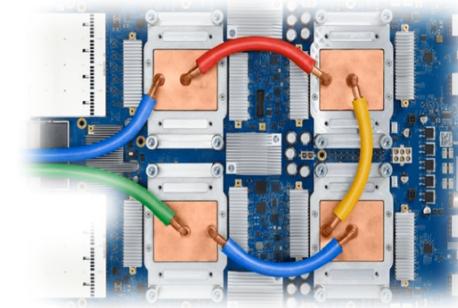
object detection



speech translation



robot control



NN accelerators

- In the Web Conference 2019: 16 (7%) papers are “deep”, 12 (5%) papers are connected to “neural” – only in title

DL inference on smartphones

DL on Cloud

- Centralized
- Powerful resources
- Well-studied
- Far from users



DL on smartphones

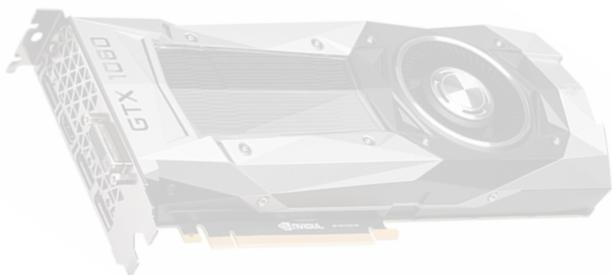
- Decentralized
- Low-end (challenging)
- Fewer prior efforts
- Close to users



DL inference on smartphones

DL on Cloud

- Centralized
- Powerful resources
- Well-studied
- Far from users



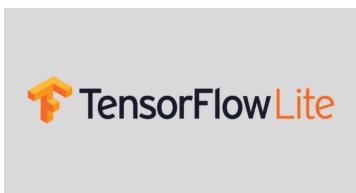
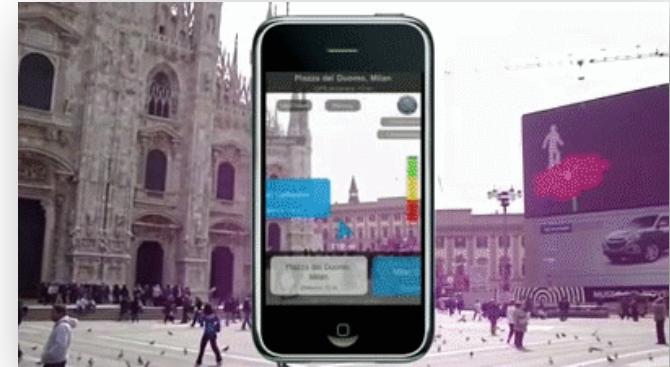
DL on smartphones

- Decentralized
- Low-end (challenging)
- Fewer prior efforts
- Close to users



DL inference on smartphones

- **From academia**
 - Cross-layer optimizations
 - Algorithm, System, Architecture
- **From industrial**
 - Android 8.1 officially introduces Neural Network APIs
 - There're emerging DL frameworks for mobile: Facebook – Caffe2, Google – TF Lite, Apple – Core ML, Tencent – ncnn, Baidu – MDL



DL inference on smartphones

- From academia

- C
- A

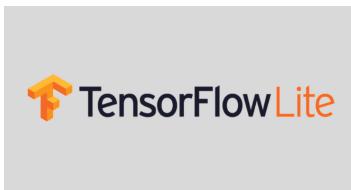
Early adopters have high influence or even decide the evolution of new technologies

- Diffusion of innovations theory, 1962

- From

- A

- There're emerging DL frameworks for mobile: Facebook – Caffe2, Google – TF Lite, Apple – Core ML, Tencent – ncnn, Baidu – MDL



Research goal and challenges

- **(The first)** to understand how DL technique is utilized in real-world mobile applications
 - What are the characteristics and usage of DL apps?
 - How do the DL models look like?
1. How to identify DL apps?
 2. How to extract DL models?



Android: the target platform

Identifying DL apps

Detecting DL functionality code? – Difficult 😰

- Heterogeneous: many ways to implement
- Binary code: written in C++ code and compiled in shared library



Identify
DL apps

Deep learning frameworks: the popular choice

mobile-oriented

Framework	Owner	Supported Mobile Platform	Mobile API	Is Open-source	Supported Model Format	Support Training
TensorFlow [37]	Google	Android CPU, iOS CPU	Java, C++	✓	ProtoBuf (.pb, .pbtxt)	✓
TF Lite [38]	Google	Android CPU, iOS CPU	Java, C++	✓	FlatBuffers (.tflite)	✗
Caffe [69]	Berkeley	Android CPU, iOS CPU	C++	✓	customized, json (.caffemodel, .prototxt)	✓
Caffe2 [9]	Facebook	Android CPU, iOS CPU	C++	✓	ProtoBuf (.pb)	✓
MxNet [49]	Apache Incubator	Android CPU, iOS CPU	C++	✓	customized, json (.json, .params)	✓
DeepLearning4J [13]	Skymind	Android CPU	Java	✓	customized (.zip)	✓
ncnn [36]	Tencent	Android CPU, iOS CPU	C++	✓	customized (.params, .bin)	✗
OpenCV [27]	OpenCV Team	Android CPU, iOS CPU	C++	✓	TesnorFlow, Caffe, etc	✗
FeatherCNN [17]	Tencent	Android CPU, iOS CPU	C++	✓	customized (.feathermodel)	✗
PaddlePaddle [25]	Baidu	Android CPU, iOS CPU & GPU	C++	✓	customized (.tar)	✓
xNN [41]	Alibaba	Android CPU, iOS CPU	unknown	✗	unknown	unknown
superid [35]	SuperID	Android CPU, iOS CPU	unknown	✗	unknown	unknown
Parrots [31]	SenseTime	Android CPU, iOS CPU	unknown	✗	unknown	unknown
MACE [24]	XiaoMi	Android CPU, GPU, DSP	C++	✓	customized (.pb, .yml, .a)	✗
SNPE [32]	Qualcomm	Qualcomm CPU, GPU, DSP	Java, C++	✗	customized (.dlc)	✗
CNNDroid [74]	Oskouei et al.	Android CPU & GPU	Java	✓	MessagePack (.model)	✗
CoreML [12]	Apple	iOS CPU, GPU	Swift, OC	✗	customized, ProtoBuf (.proto, .mlmodel)	✓
Chainer [10]	Preferred Networks	/	/	✓	customized (.chainermodel)	✓
CNTK [23]	Microsoft	/	/	✓	ProtoBuf (.proto)	✓
Torch [40]	Facebook	/	/	✓	customized (.dat)	✓
PyTorch [30]	Facebook	/	/	✓	customized, pickle (.pkl)	✓

support for

mobile platforms

Identifying DL apps

~~Detecting DL functionality code?~~ – ~~Difficult~~ 😢

- Heterogeneous: many ways to implement
- Binary code: written in C++ code and compiled in shared library

Detecting DL framework usage? – Life is easier 😎

- Most DL apps are developed by popular DL frameworks
- Those frameworks remain relatively stable across versions

```
$ readelf -p .rodata libMagicClean.so # from adobe app
```

```
[ba521c] ./tensorflow/core/framework/numeric_op.h
[ba5280] ./tensorflow/core/kernels/reduction_ops_common.h
[ba54c4] tensorflow/core/kernels/pooling_ops_common.cc
[ba56a4] tensorflow/core/kernels/pad_op.cc
[ba58d8] ./tensorflow/core/kernels/pooling_ops_common.h
[ba5a70] tensorflow/core/kernels/matmul_op.cc
[ba5b74] tensorflow/core/kernels/function_ops.cc
[ba5c08] ./tensorflow/core/kernels/winograd_transform.h
[ba5c44] tensorflow/core/kernels深深_conv2d.cc
[ba5c6c] tensorflow/core/kernels/cwise_ops_common.cc
[ba5de0] ./tensorflow/core/util/bcast.h
[ba5e38] ./tensorflow/core/kernels/cwise_ops_common.h
```



Identify
DL apps

The .rodata section of
libMagicClean.so in
adobe app

- *apktool + readelf*
- *TensorFlow detected*

Extracting DL models

Scanning apk & check file format

- Tflite models use protobuf format
- Ncnn models start with a magic number: 7767517



Extract
DL Models

Complementary-1: install apk and check local files

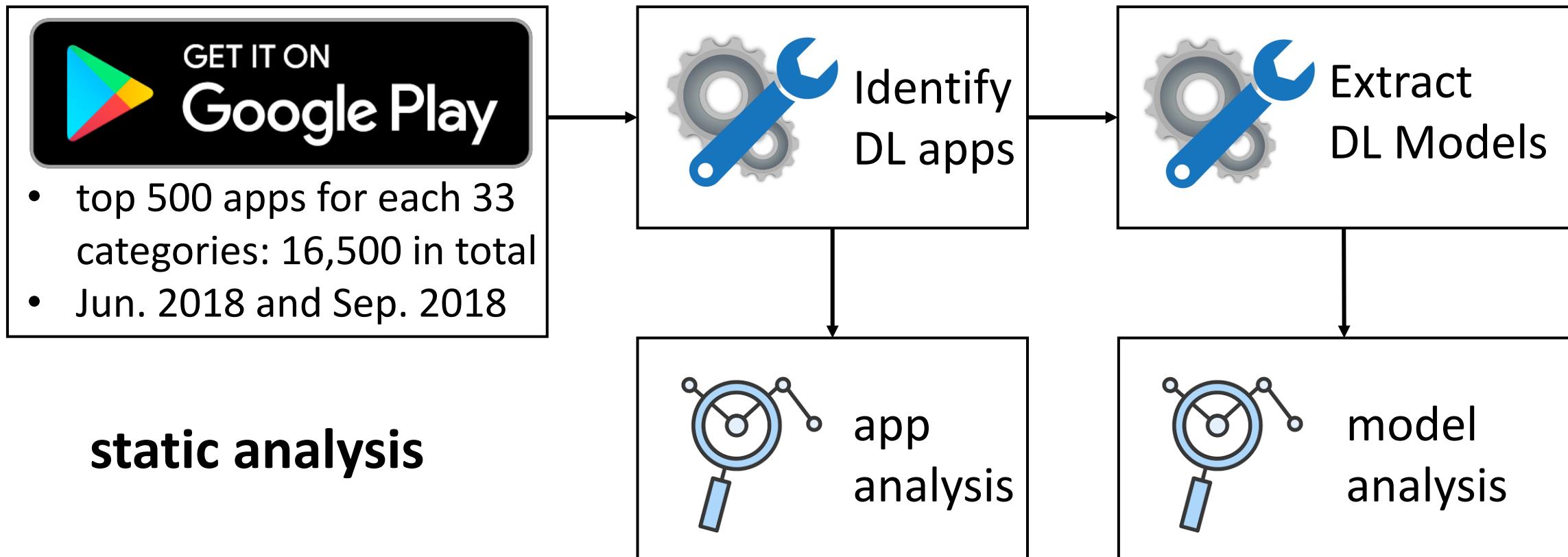
- Some models are fetched from network

Complementary-2: decrypt files

- Some models are encrypted

Analysis workflow

<https://github.com/xumengwei/MobileDL>



How many DL apps?

- There're ???/16,500 DL apps found in Sep. 2018.



How many DL apps?

- There're **211**/16,500 DL apps found in Sep. 2018.
 - only 1.3%

How many DL apps?

- There're **211**/16,500 DL apps found in Sep. 2018.
 - only 1.3%
 - DL apps are downloaded by **13M (11.9%)**
 - DL apps receive **9.6M (10.5%) reviews**
- Jun. 2018 –
Sep. 2018

★★★★★ 5,616,407 

4.2



 5,616,407 total

How many DL apps?

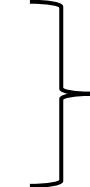
- There're **211/16,500** DL apps found in Sep. 2018.
 - only 1.3%

- DL apps are downloaded by **13M (11.9%)**
 - DL apps receive **9.6M (10.5%) reviews**
- } Jun. 2018 –
Sep. 2018

	Downloads (median)	Reviews (median)
DL apps	~5,000,000	~100,000
Non-DL apps	~40,000	~1,000

DL apps vs. non-DL apps

How many DL apps?

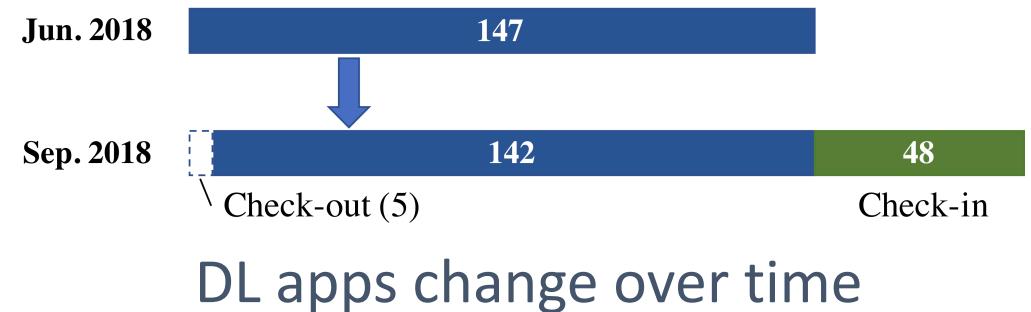
- There're **211**/16,500 DL apps found in Sep. 2018.
 - only 1.3%
 - DL apps are downloaded by **13M (11.9%)**
 - DL apps receive **9.6M (10.5%) reviews**
 - Early adopters of DL are top apps
- 
- Jun. 2018 –
Sep. 2018

How many DL apps?

- There're **211**/16,500 DL apps found in Sep. 2018.
 - only 1.3%

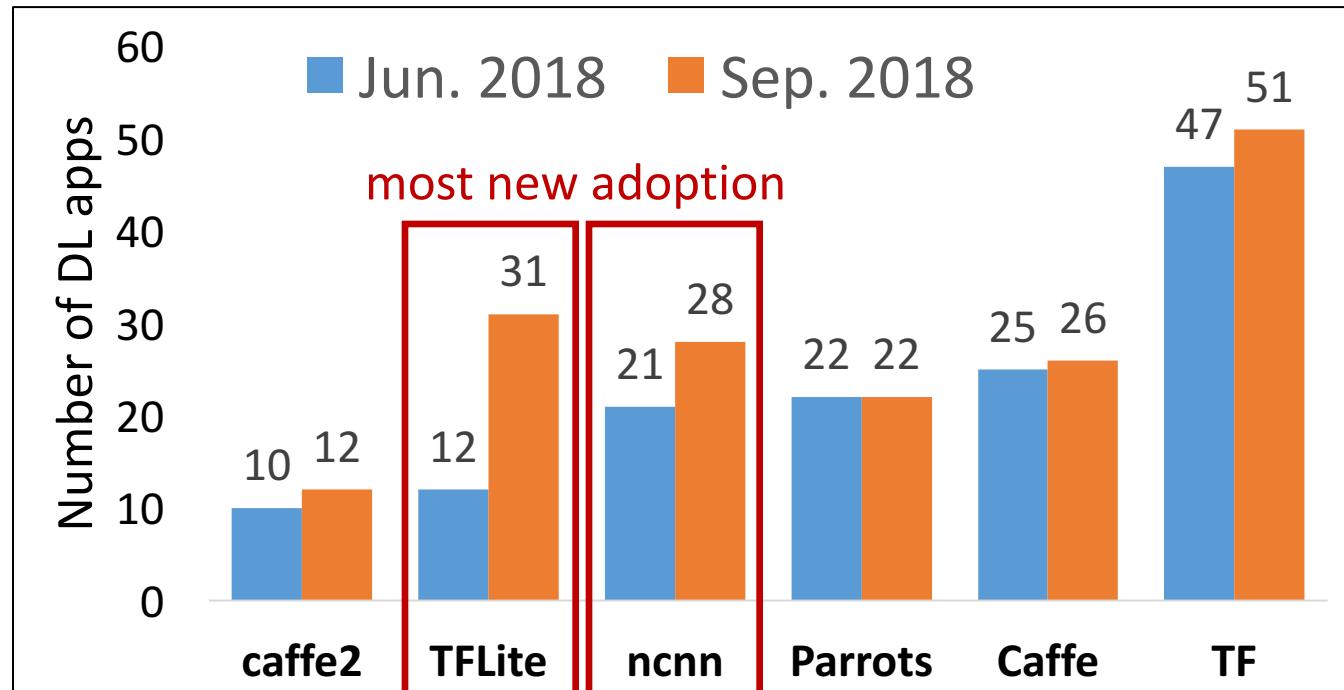
- DL apps are downloaded by **13M (11.9%)**
- DL apps receive **9.6M (10.5%) reviews**
- Early adopters of DL are top apps

- DL apps are increasing rapidly
 - ↑27% within 3 months



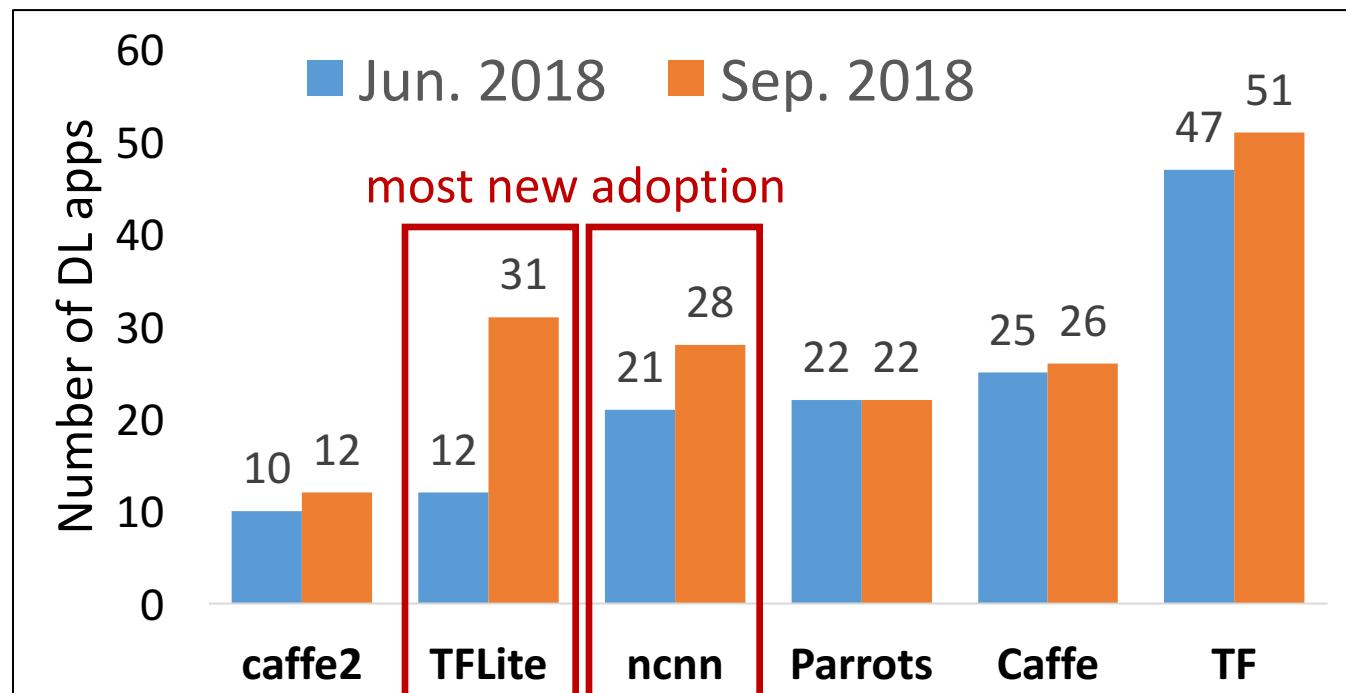
Which DL frameworks are most popular?

- Mobile-oriented DL frameworks are gaining traction
 - e.g., TFLite, ncnn, Caffe2, etc



Which DL frameworks are most popular?

- Mobile-oriented DL frameworks are gaining traction
 - e.g., TFLite, ncnn, Caffe2, etc



framework	library size (MB)
TF	15.3
Caffe	10.1
ncnn	2.5
TFLite	2.1
Caffe2	4.1

generic

mobile-oriented

What's the usage of DL in smartphone?

- The usage is diverse – of course

Image photo beauty face detection AR face identification Text recognition	Audio Speech recognition Sound recognition	Text word prediction auto-correct Translation text classification Smart reply
Other recommendation movement tracking action detection		

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?



Image photo beauty face detection AR face identification Text recognition	Audio Speech recognition Sound recognition	Text word prediction auto-correct Translation text classification Smart reply
	Other recommendation movement tracking action detection	

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?

Image: 149 photo beauty face detection AR face identification Text recognition	Audio: 24 Speech recognition Sound recognition	Text: 26 word prediction auto-correct Translation text classification Smart reply
	Other: 19 recommendation movement tracking action detection	

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?

Image: 149

photo beauty: 94

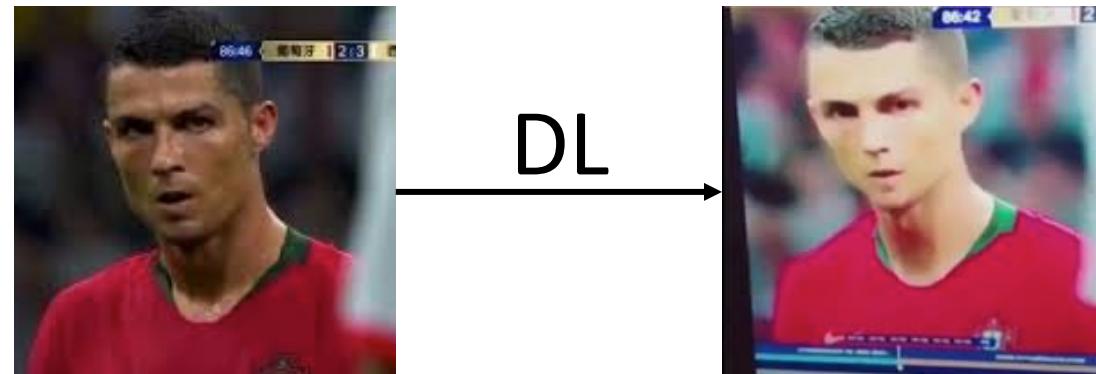
face detection

AR

face identification

Text recognition

- 44.5% DL apps use DL for photo beauty



What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?

Image: 149

photo beauty: 94

face detection

AR

face identification

Text recognition

- 29% DL apps in photography category



BeautyCam

Meitu, Inc. **Photography**

Everyone

This app is compatible with your device.

★★★★★ 350,883

Installed

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?
- What's the role played by DL? Important or dispensable?

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?
- What's the role played by DL? Important or dispensable?
- ??% apps use DL as core feature

What makes a core feature?

- Hot: invoked frequently
- Essential: needed badly

Text recognition in different apps

- ✓ Scanner app (Adobe) ✓
- ✗ Payment app (Alipay) ✗

What's the usage of DL in smartphone?

- The usage is diverse – of course
- What's the most popular use case?
- What's the role played by DL? Important or dispensable?
- 81% apps use DL as core feature: a strong motivation for future research to improve DL experience

Developers and bad practice

- **Most DL apps come from big companies**
 - Google *, Facebook, Adobe, Meitu, etc



Google



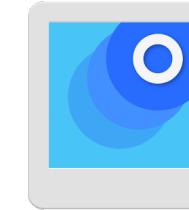
Gboard



Google
text-to-speed



Messages



PhotoScan



Google
Street View

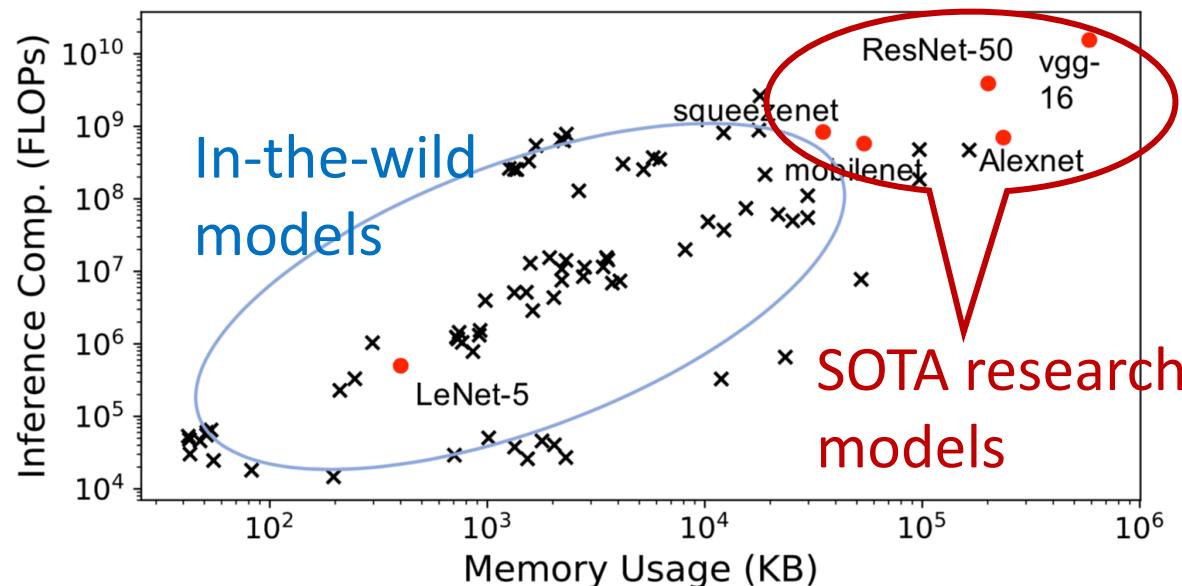


Google
Photos

- **Some DL apps use more than one DL frameworks (bad practice?)**
 - 24 apps; 5.4MB (13.6%) storage “overhead”
 - Can be avoided as DL frameworks can be used interchangeably

DL model resource footprint

- DL models are very lightweight
 - vs. state-of-the-art research models
 - The execution time of 80% models are less than 15ms on typical smartphone



	memory usage (MB)	Inference complexity (FLOPs)
In the wild	2.5	10M
ResNet-50	200	4G
MobileNet	54	500M

DL model optimizations

- **DL models are rarely optimized**
 - Only 6% models have been compressed

	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
<i>Total</i>	<i>0.00%</i>	<i>6.32%</i>	<i>0.00%</i>	<i>0.00%</i>

DL model protection

- For intellectual property and security



stop sign ->
speed limit sign

DL model protection

- For intellectual property and security



stop sign ->
speed limit sign

- However, only small number of apps protect their models
 - 19.2% models are encrypted
 - 39.2% models are obfuscated
- Very few frameworks (2) provide model protection
 - strip text information, convert to C++ code

Implications

- **To app developers**

- ❖ DL is affordable on smartphones: feel free to try it
- ❖ Prefer mobile DL frameworks: they are often faster and more lightweight

- **To DL framework developers**

- ❖ Make optimizations for smartphones (speed, memory, etc)
- ❖ Build model protection as a functionality

- **To DL researchers**

- ❖ Validate your brilliant ideas on lightweight models as well

Summary: DL is becoming increasingly popular and important on smartphones, but its power is still underexploited – *we are at dawn now*



Conclusion

- The first empirical study to understand how deep learning is utilized in real-world mobile applications.
- We find that deep learning is becoming increasingly popular and important on smartphones, but its power is still underexploited.

Thank you for attention!



Limitations and Future Work

- Our tool has false positive and false negative
- More platforms, longer time ranges