

# IESTI01 - TinyML

## Visual Wake Words (VWW) Introduction

Prof. Marcelo Rovai

July 21<sup>st</sup>, 2021



What is VWW,  
Visual Wake Words?

## Person Detection

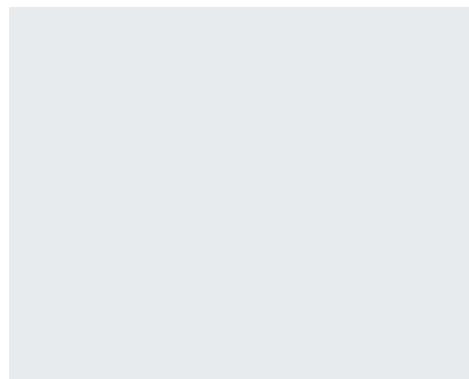


✓

## Mask Detection



✓



✗



✗



# Next big thing!

## Augmented Reality

- Smart Shopping



# Next big thing!

## Augmented Reality

- Smart Shopping
- Navigation



# Opportunities



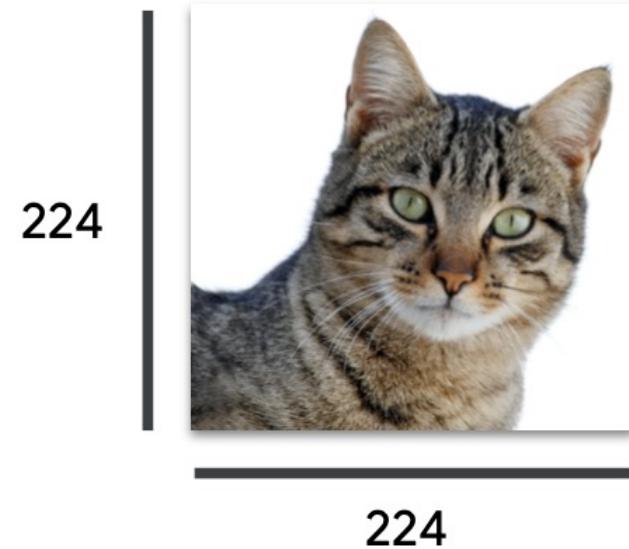
Run visual wake words  
**on-device**

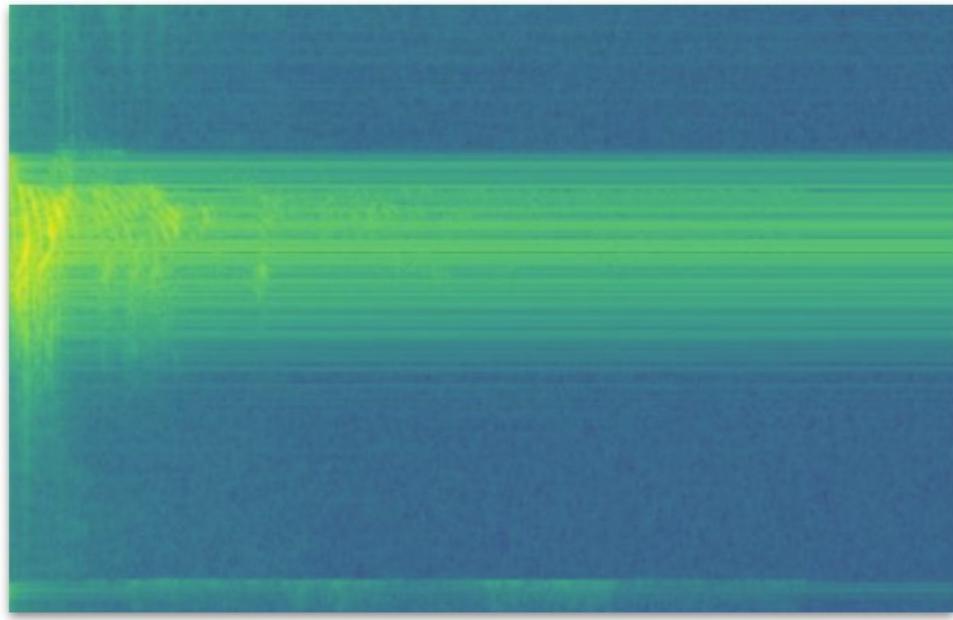
# Visual Wake Words, Challenges

# Simple Experiment

$$224 \times 224 \times 3 \times 4 = 602,112 \text{ Bytes}$$

Pixels      RGB  
              (# channels)      Bytes/Pixel





$$49 \times 40 \times 1 \times 4 = 7,840 \text{ Bytes}$$

Pixels

RGB  
(# channels)

Bytes/Pixel

49

$$224 \times 224 \times 3 \times 4 = 602,112 \text{ Bytes}$$

Pixels

RGB  
(# channels)



224



224

# Simple Experiment

**Always-on** (Visual Wake Words)?

- Much more data (than KWS)
  - Higher **latency**
  - Higher **power consumption**  
(drains battery)
- Lower **user satisfaction**

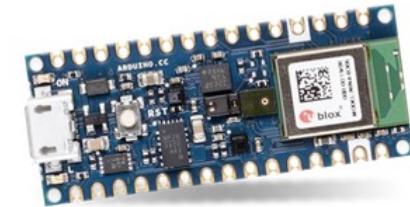
224



224

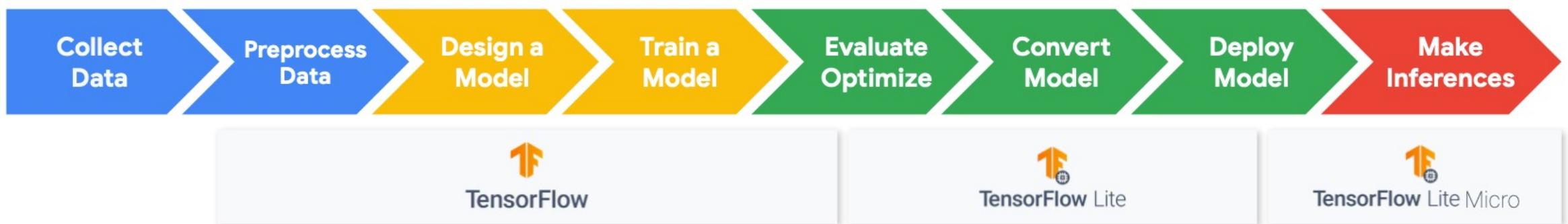
# Memory (CNN Models)

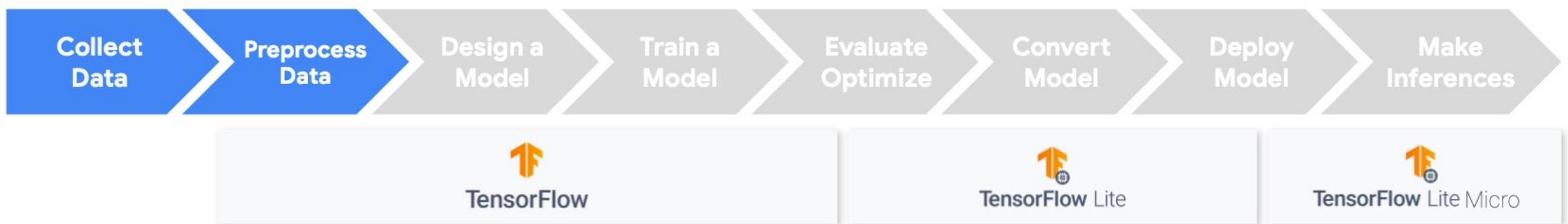
Model	Size	Top-1 Accuracy
Xception	<b>88 MB</b>	0.790
VGG16	<b>528 MB</b>	0.713
ResNet50	<b>98 MB</b>	0.749
Inception v3	<b>92 MB</b>	0.779
MobileNet v1	<b>16 MB</b>	0.713
DenseNet 201	<b>80 MB</b>	0.773
NASNetMobile	<b>23 MB</b>	0.825



Our board  
has **256 KB** of RAM (memory)

# Visual Wake Words, Data Collection and Processing





# Visual Wake Words Dataset

## Visual Wake Words Dataset

Aakanksha Chowdhery, Pete Warden, Jonathon Shlens,  
Andrew Howard, Rocky Rhodes

Google Research

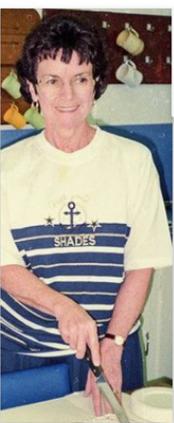
{chowdhery, petewarden, shlens, howarda, rocky}@google.com

<https://arxiv.org/pdf/1906.05721.pdf>

# Visual Wake Words Dataset



**Label: "person"**



**Label: "person"**



**Label: "not-person"**

(Labeled from COCO dataset)

# Visual Wake Words Dataset

Data collection is **DIFFICULT**

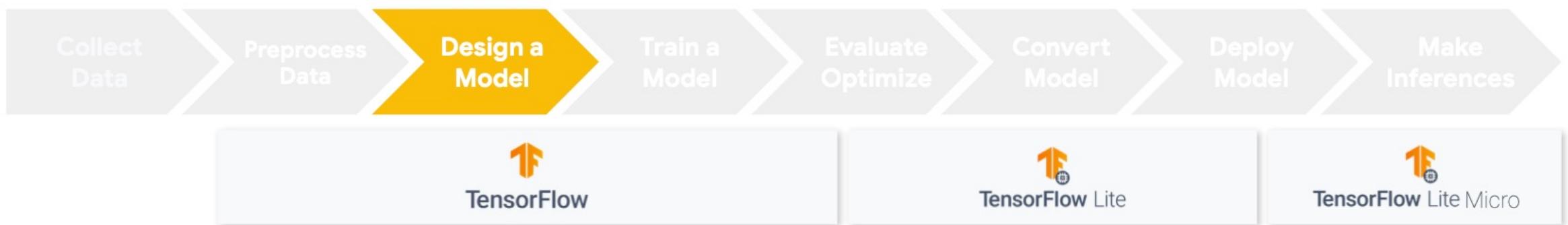
- This dataset and collection process is ***limited*** and has bias
- Small number of relevant images
- Large quantity of irrelevant images

## Visual Wake Words Dataset

Aakanksha Chowdhery, Pete Warden, Jonathon Shlens,  
Andrew Howard, Rocky Rhodes  
Google Research

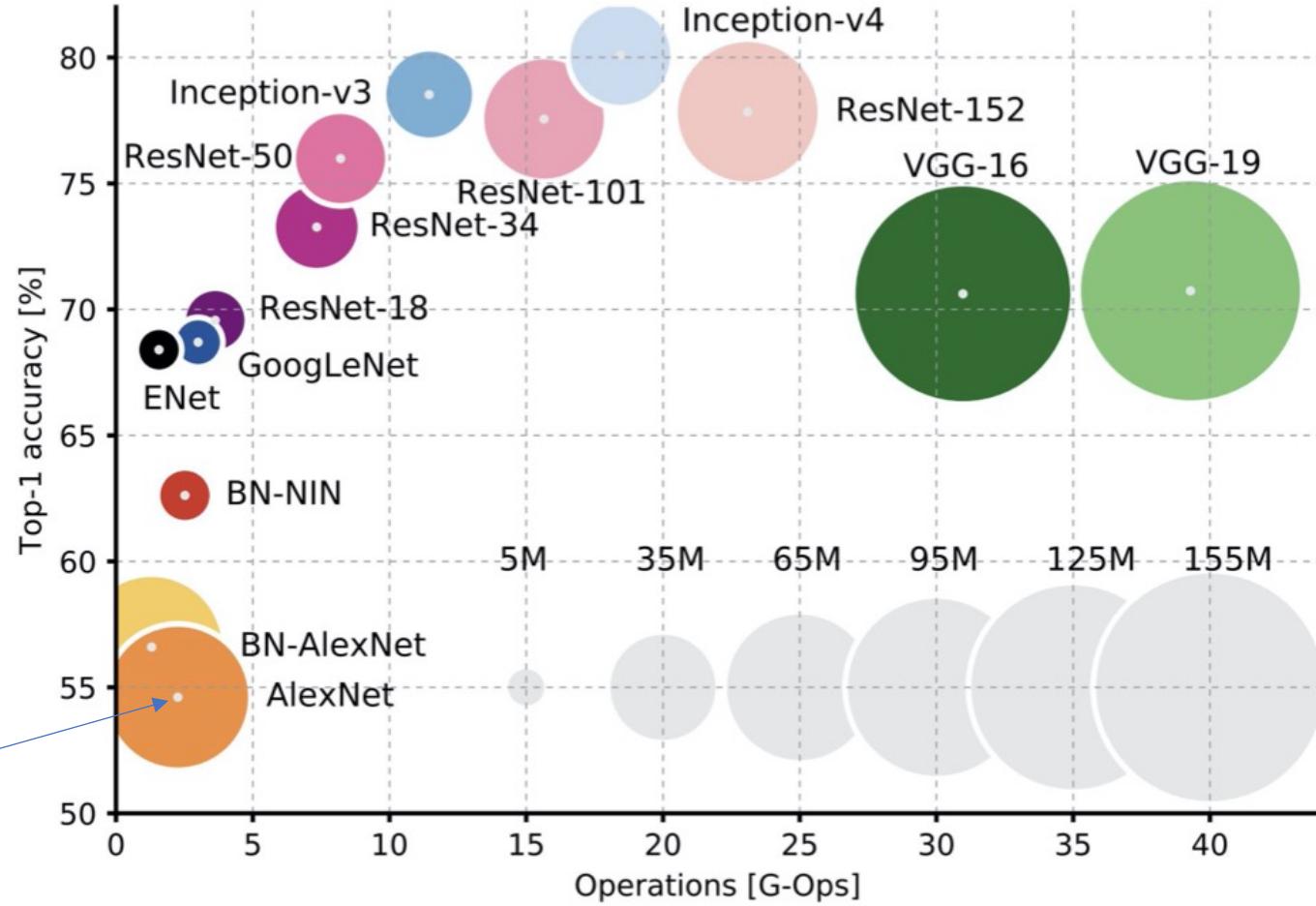
{chowdhery, petewarden, shlens, howarda, rocky}@google.com

# Visual Wake Words, Model



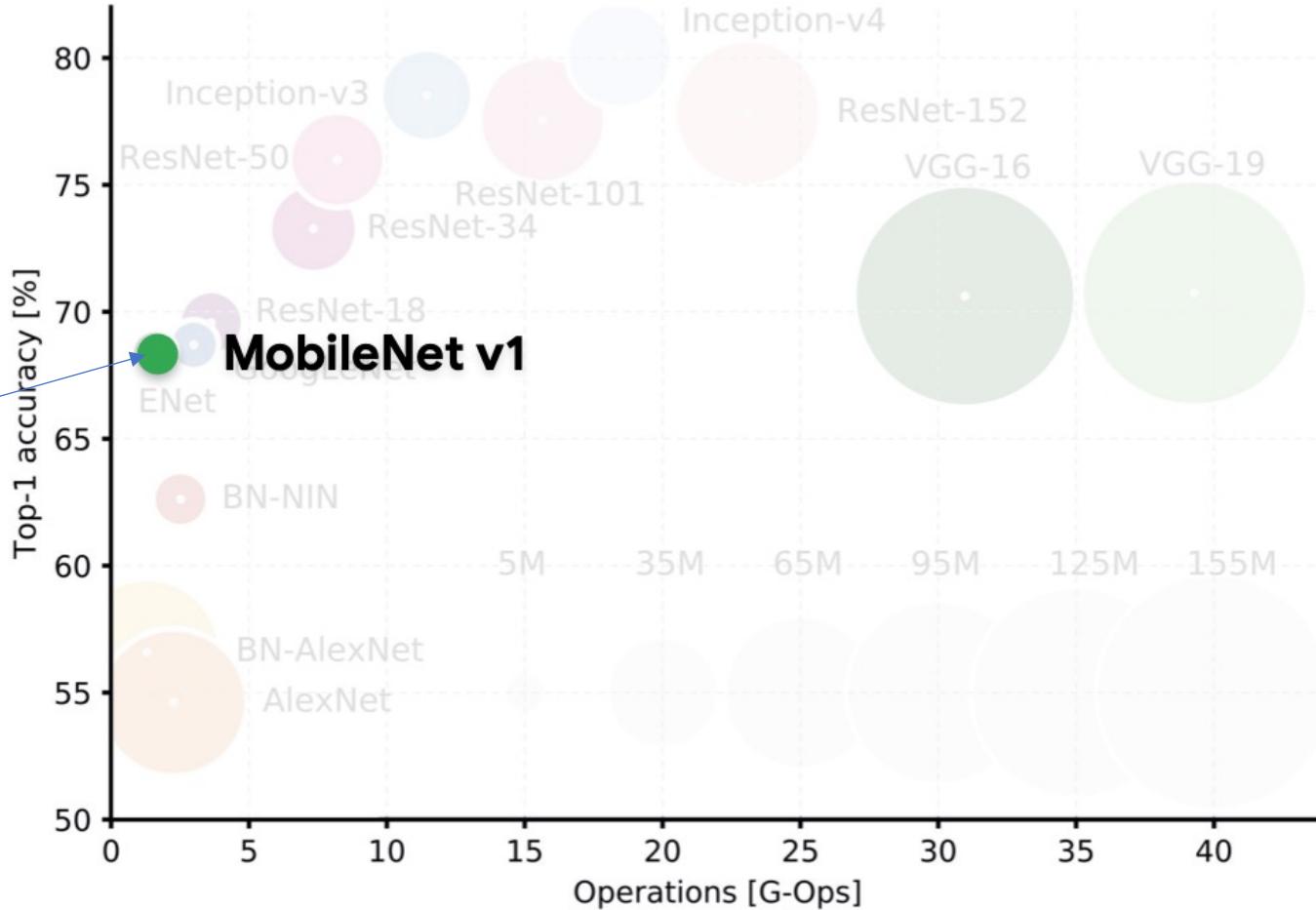
# Model Evolution

(2012)



# Model Evolution

(2017)



# MobileNet v1

## **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**

Andrew G. Howard

Weijun Wang

Menglong Zhu

Tobias Weyand

Bo Chen

Marco Andreetto

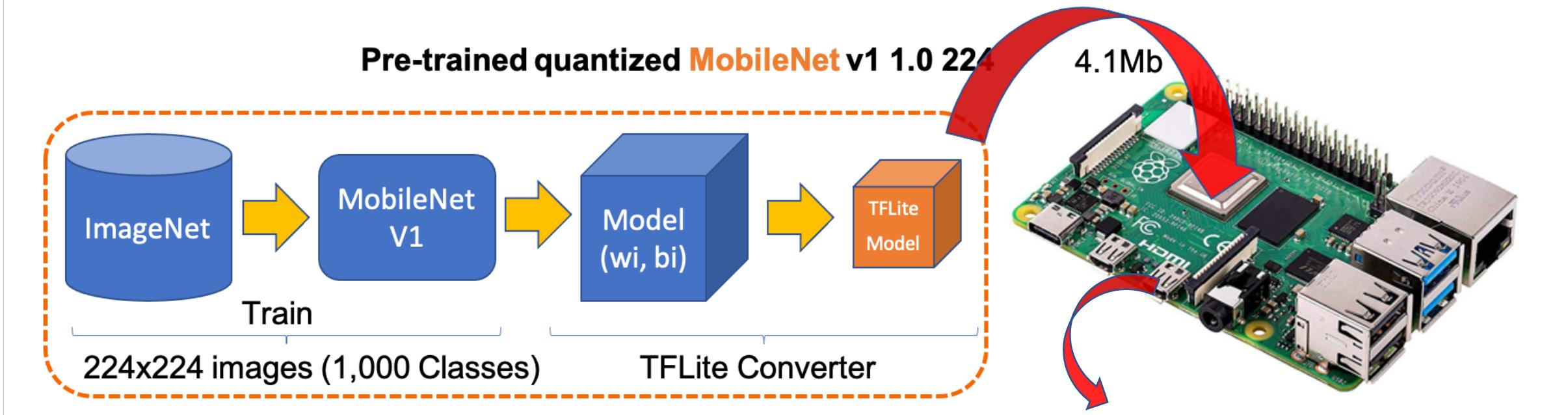
Dmitry Kalenichenko

Hartwig Adam

Google Inc.

{howarda,menglong,bochen,dkalenichenko,weijunw,weyand,anm,hadam}@google.com

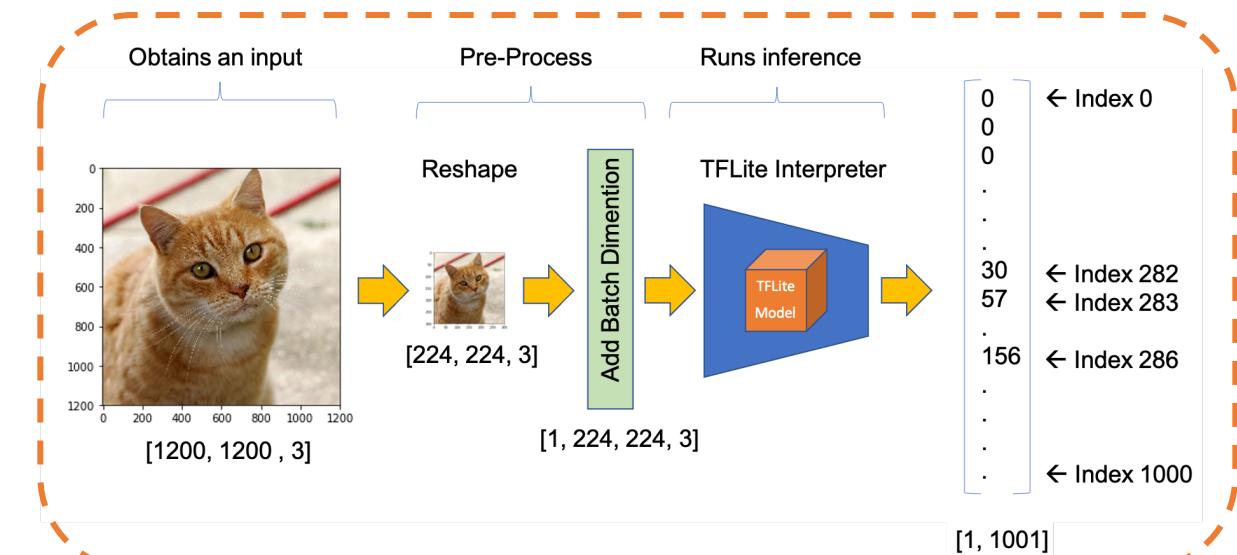
<https://arxiv.org/pdf/1704.04861.pdf>



Exploring IA at the Edge!



UNIFEI TELC03 30-Oct-20

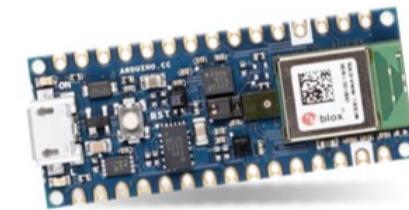


# MobileNet v1

Model	Size	Top-1 Accuracy
MobileNet v1	<b>16 MB</b>	0.713

\* Not Quantized

Fine for mobile phones  
with GB of RAM, but 64X  
microcontroller RAM



Our board [Course 3 Kit] only  
has **256KB** of RAM (memory)

# Further Optimizations

Multiply-Accumulates

$a$	Image Size	MACs (millions)	Params (millions)	Top-1 Accuracy
1	224	569	4.24	70.7
1	128	186	4.14	64.1
0.75	224	317	2.59	68.4
0.75	128	104	2.59	61.8
0.5	224	150	1.34	64.0
0.5	128	49	1.34	56.2
0.25	224	41	0.47	50.6
0.25	128	14	0.47	41.2

## Model

MobileNetV1 96x96 0.25

A pre-trained multi-layer convolutional network designed to efficiently classify images. Uses around 105.9K RAM and 301.6K ROM with default settings and optimizations. Works best with 96x96 input size. Supports both RGB and grayscale.

## Image Size

MobileNetV1 96x96 0.2

Uses around 83.1K RAM and 218.3K ROM with default settings and optimizations. Works best with 96x96 input size. Supports both RGB and grayscale.

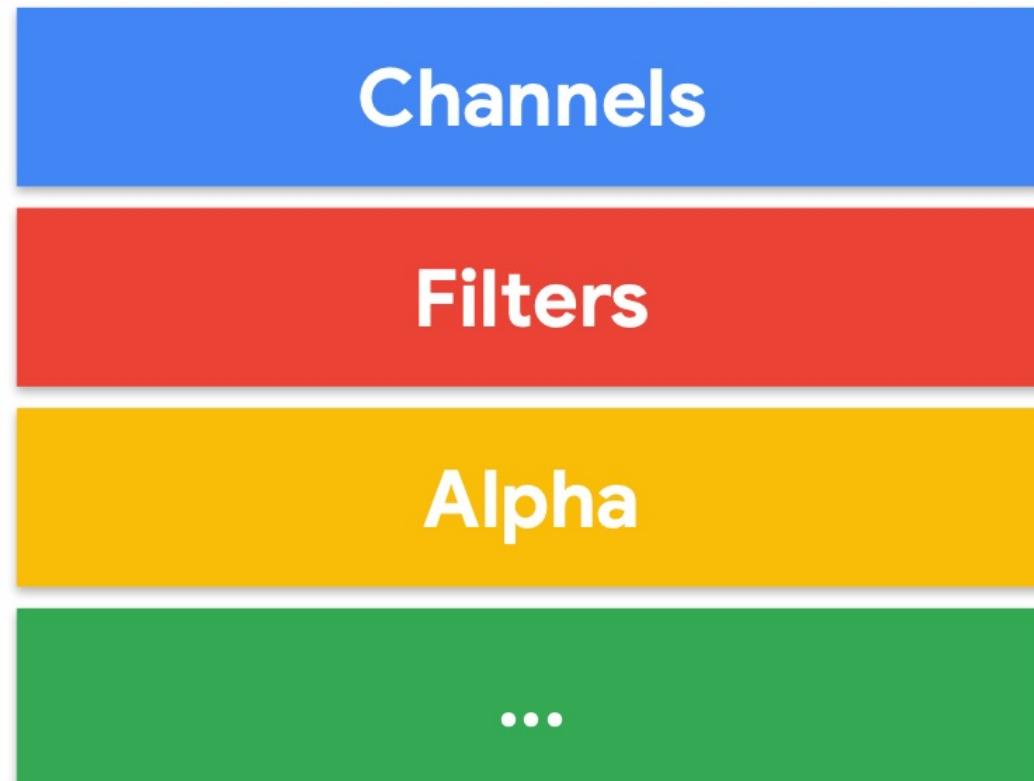
## Alpha

MobileNetV1 96x96 0.1

Uses around 53.2K RAM and 101K ROM with default settings and optimizations. Works best with 96x96 input size. Supports both RGB and grayscale.



# Neural Architecture Search (NAS)



# Visual Wake Words, Training a Model



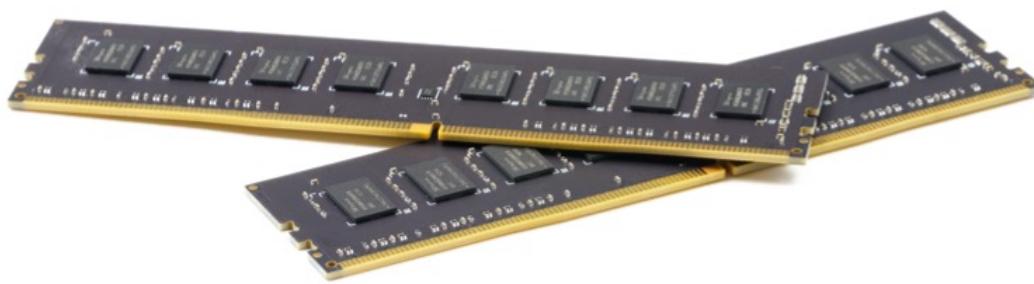
# Training Pipeline: Need Lots of Data



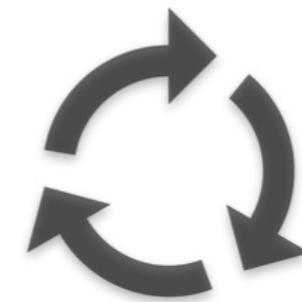
1000 Classes

1000 Images / Class

# Training Pipeline: Need Compute Resources



Memory



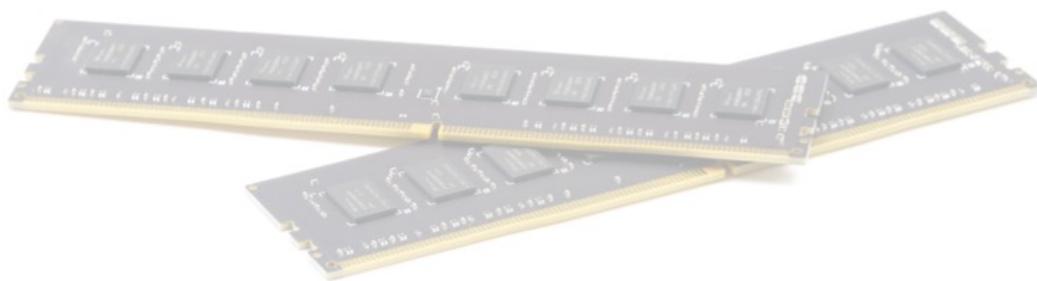
Compute



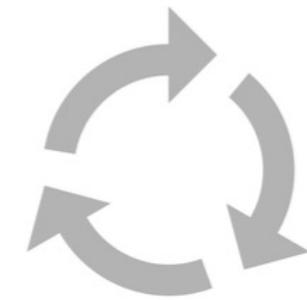
GPU and  
Accelerators

# Training Pipeline: Need Compute Resources

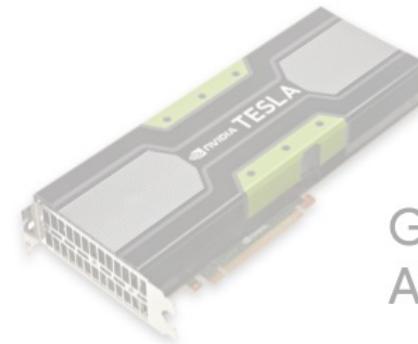
***Computationally Intensive  
Repeated Many Times (Epochs)***



Memory

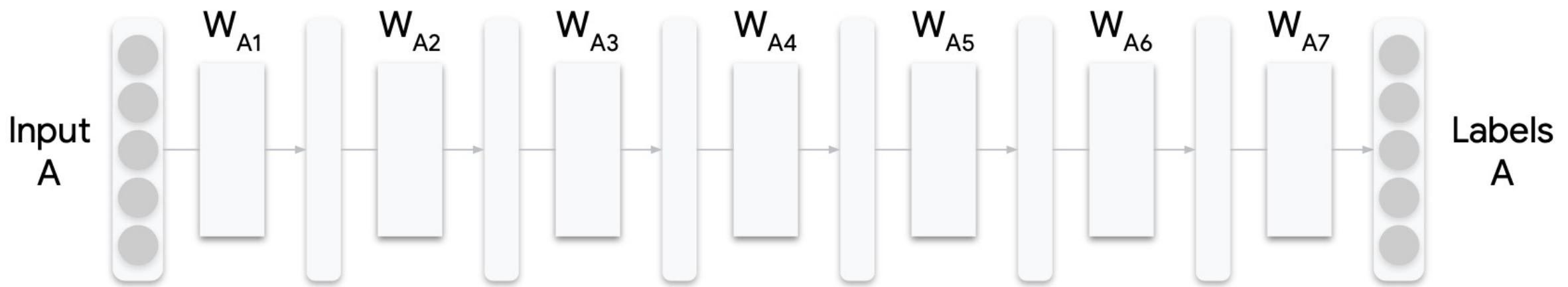


Compute

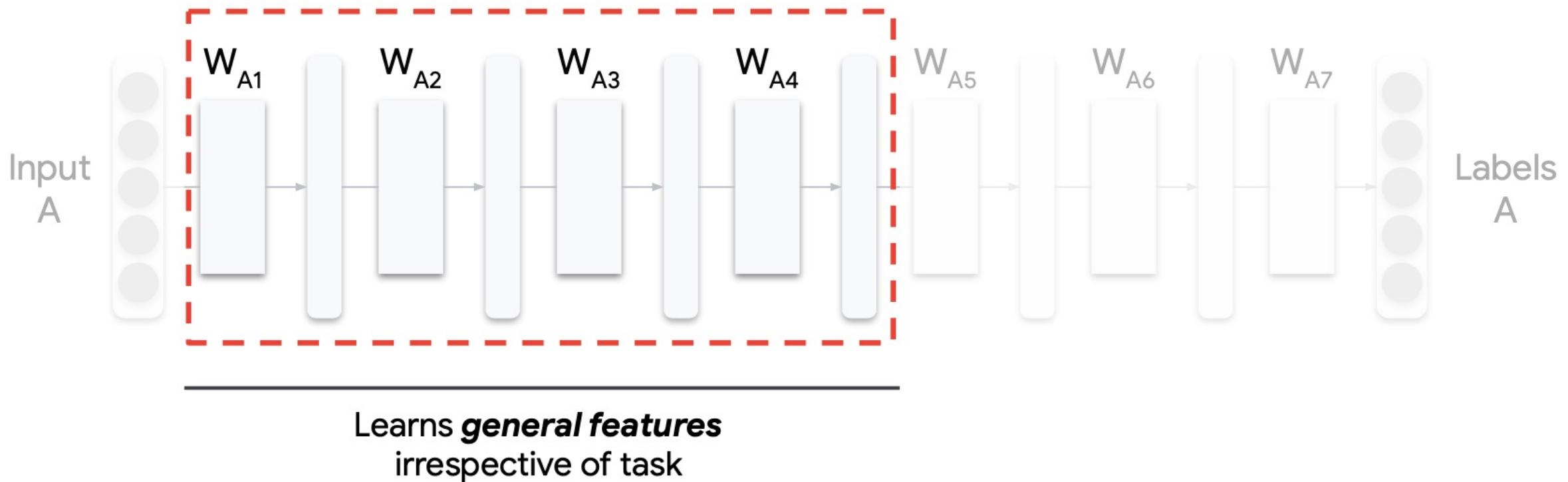


GPU and  
Accelerators

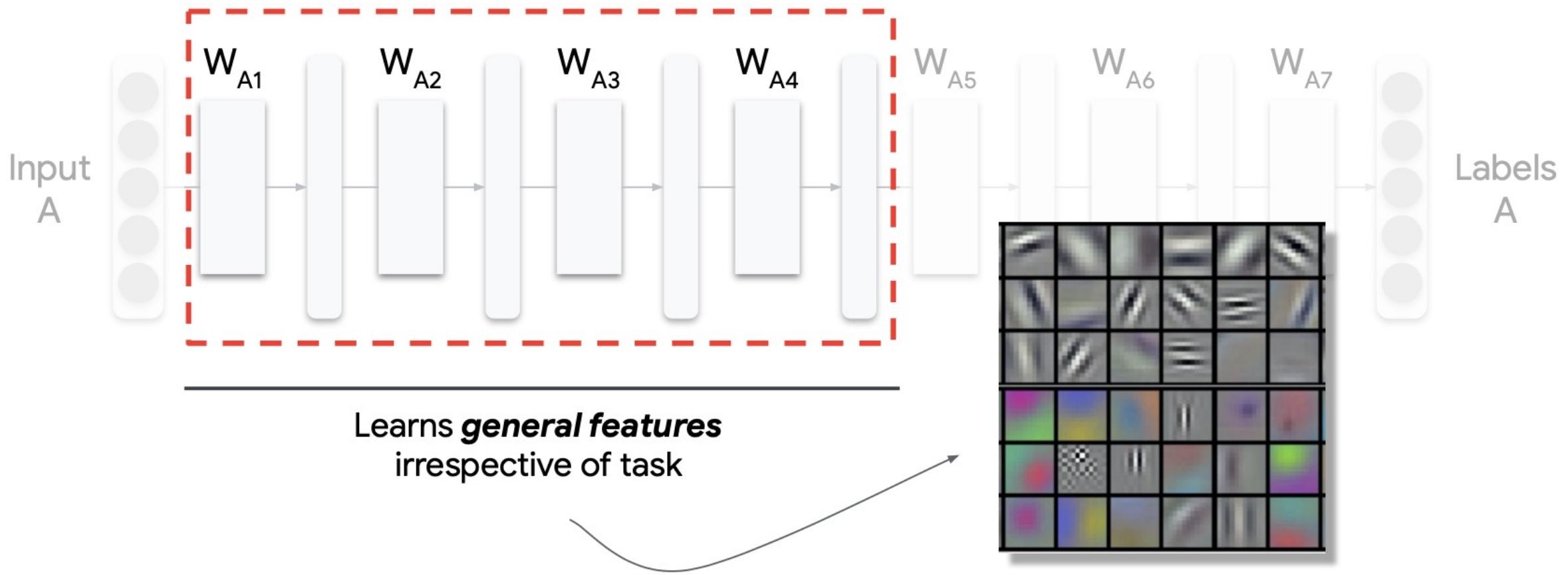
# End Result of Training



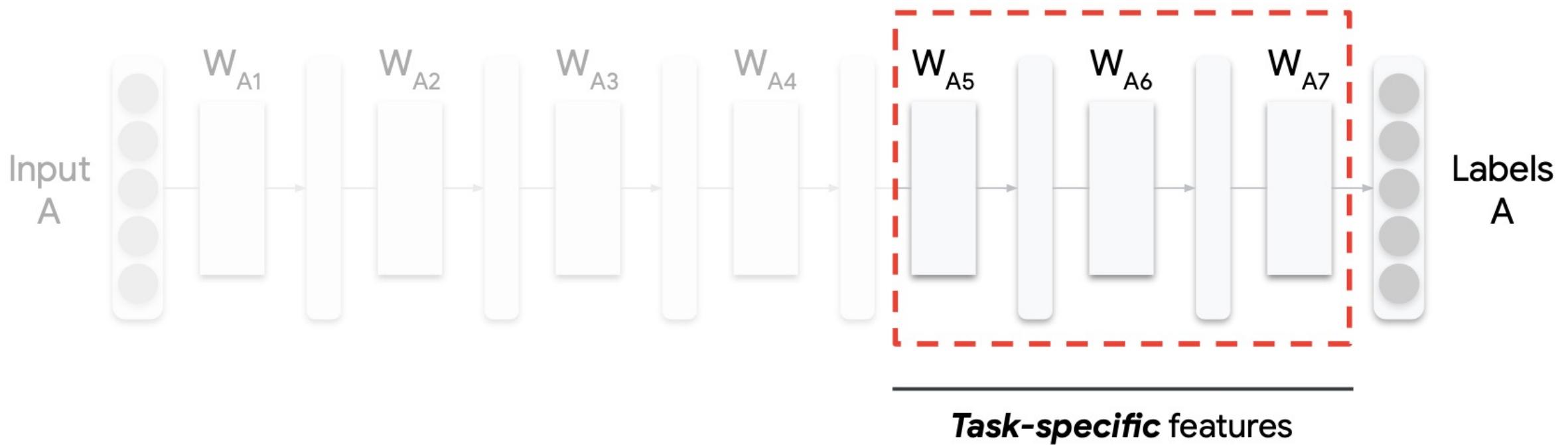
# End Result of Training



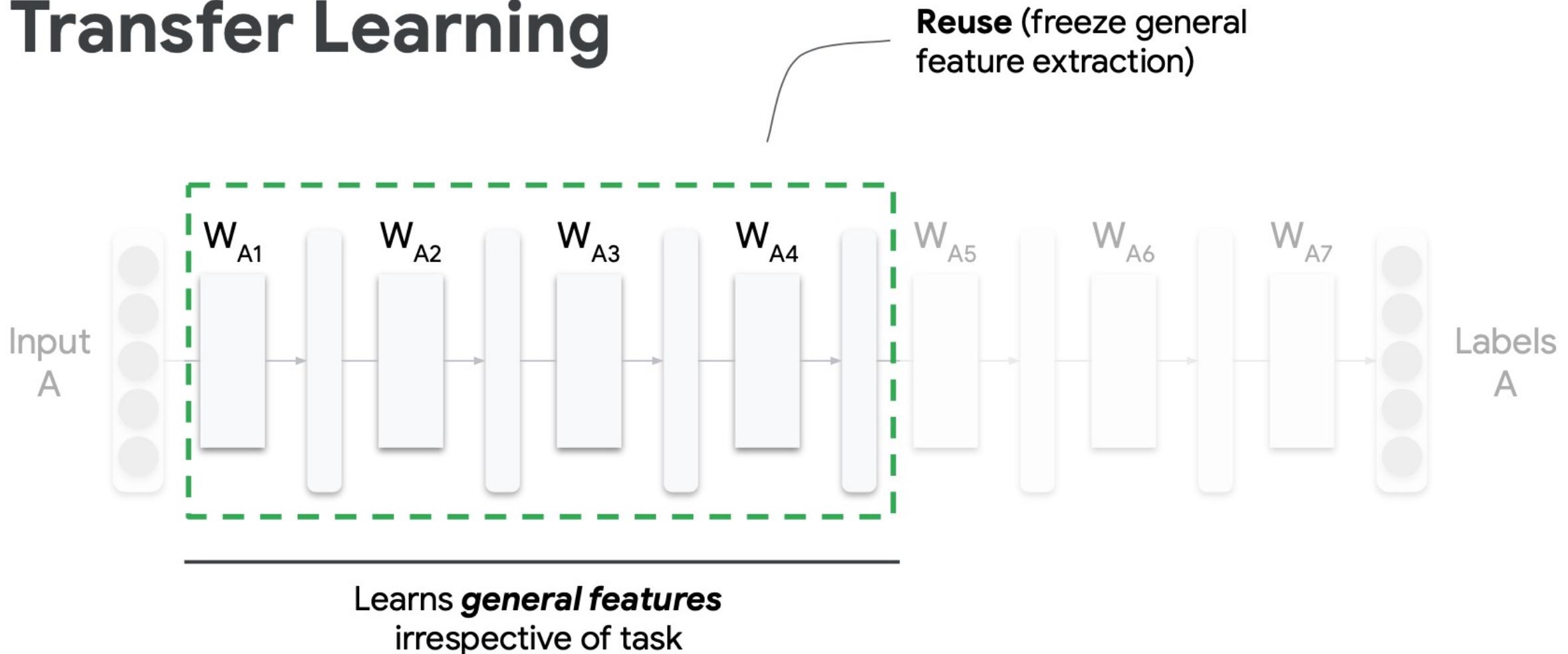
# End Result of Training



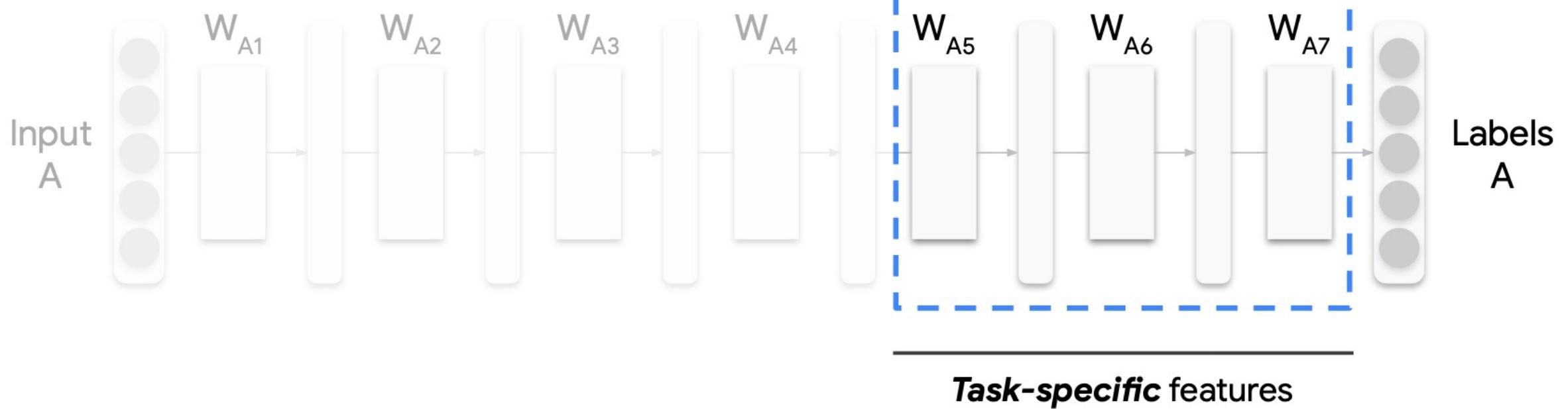
# End Result of Training



# Transfer Learning



# Transfer Learning



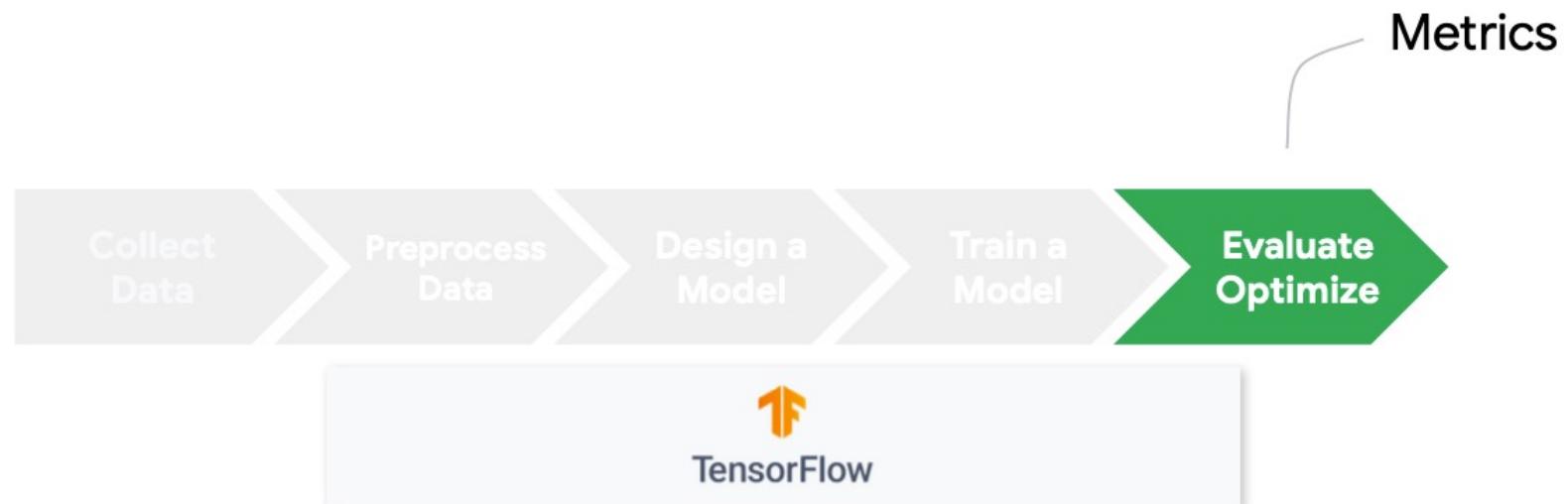
# Mask Detection using Transfer Learning

## Code Time!

Mask\_Detection\_using\_Transfer\_Learning.ipynb



# Visual Wake Words, Metrics



# Common Metrics



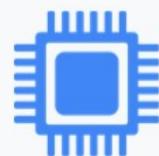
Accuracy

**Quantitative**



Efficiency

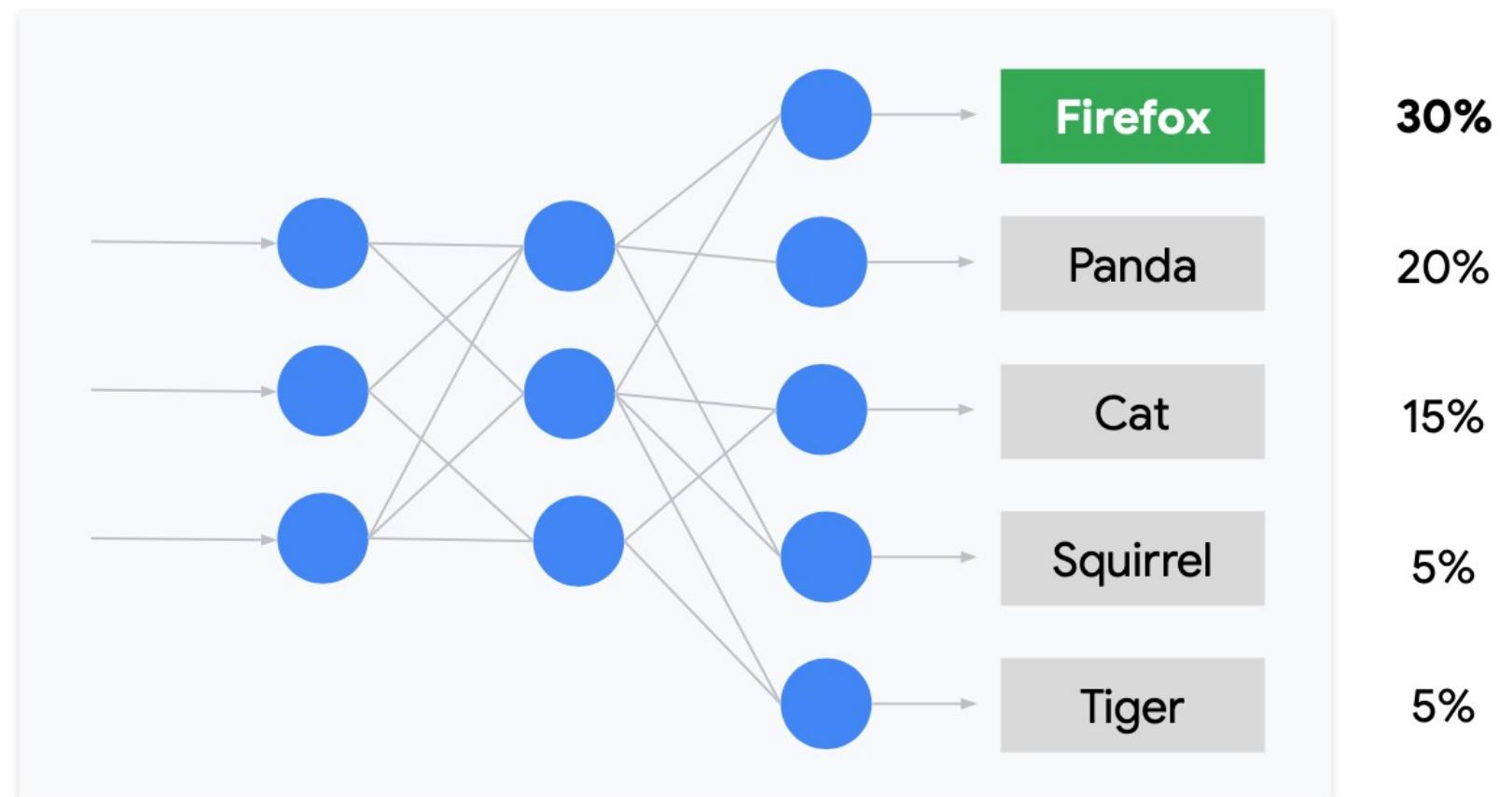
**Quantitative**



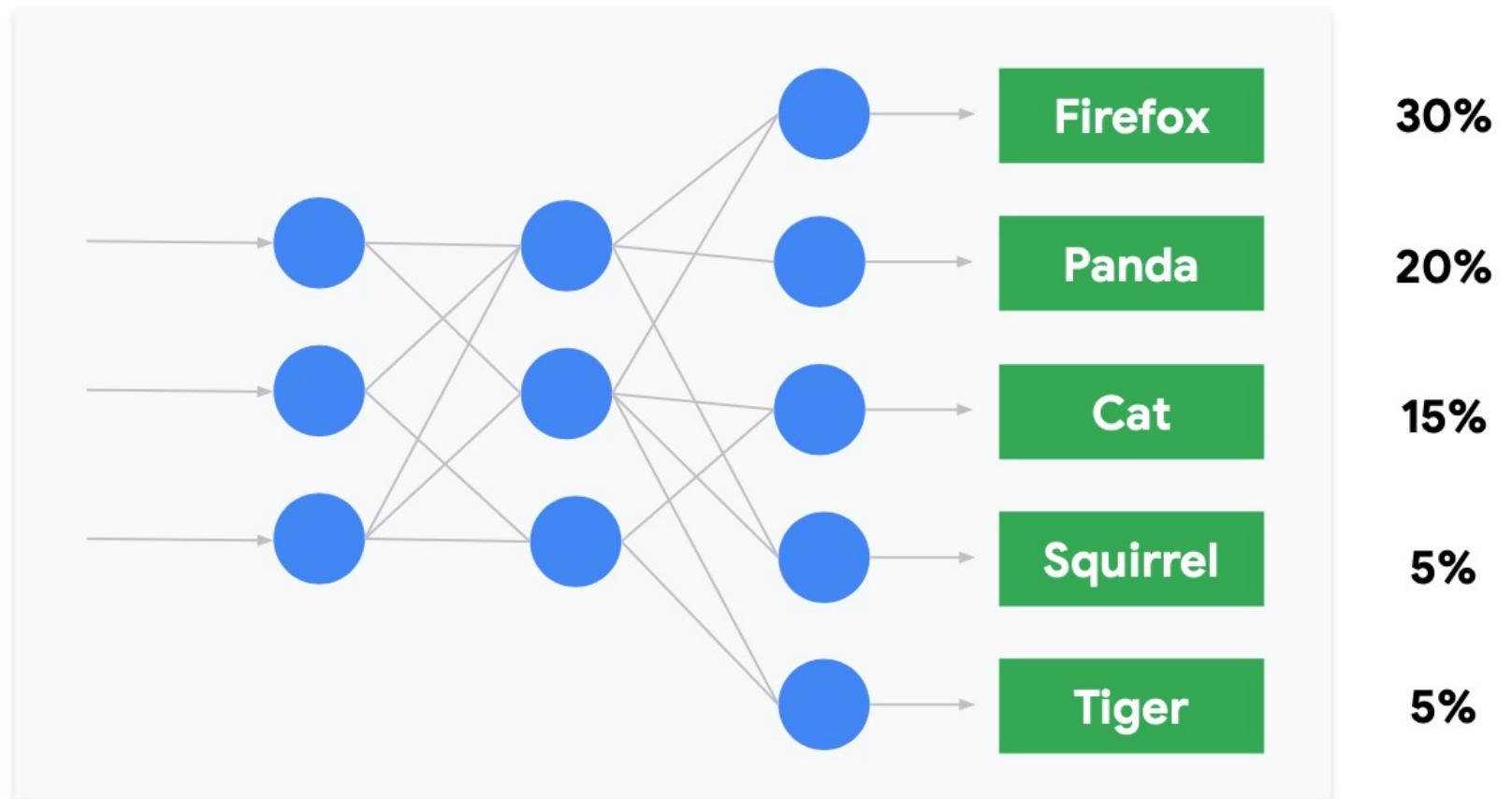
User Experience

**Qualitative**

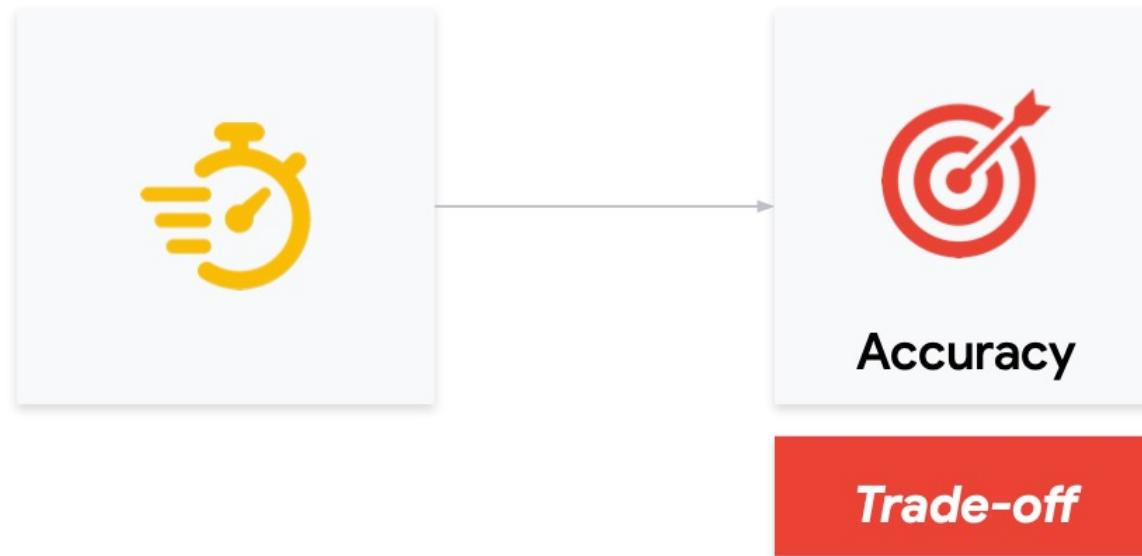
# Top-1 Accuracy



# Top-5 Accuracy

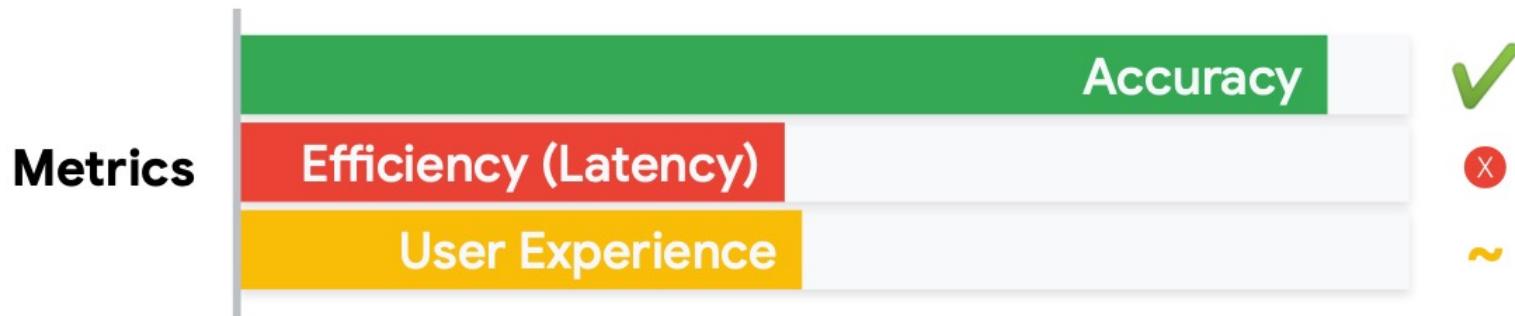


# Latency



# Latency

Accurate but *SLOW* model?



# Latency

Lower quality, but *faster* model?



Metrics	Accuracy	~
	Efficiency (Latency)	✓
	User Experience	✓

# Fairness

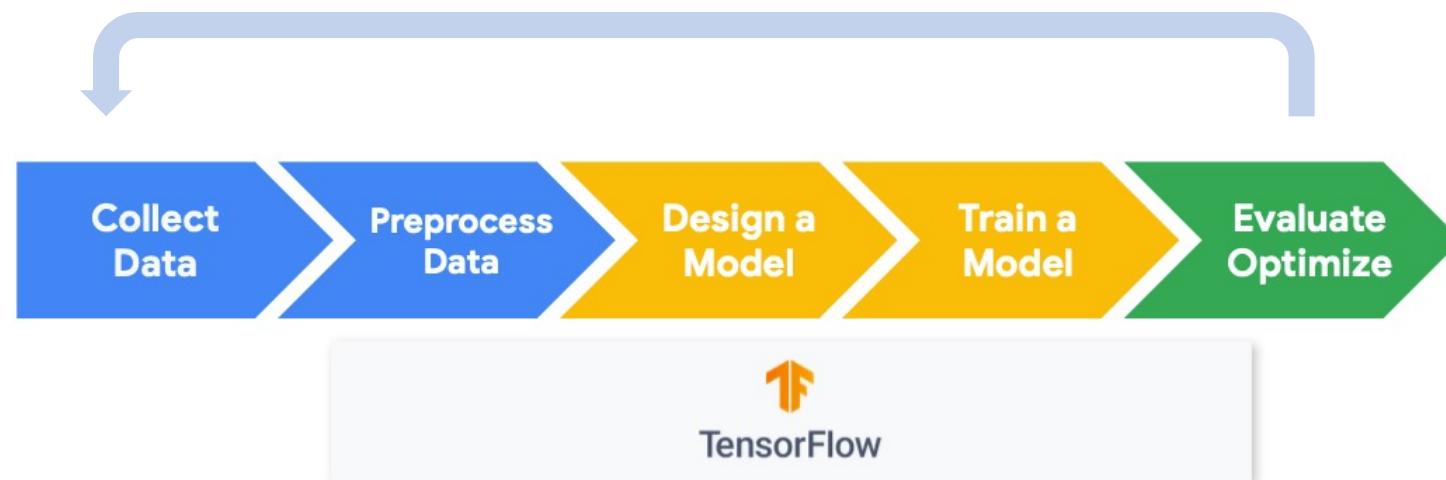
User in *majority* group of training data?



Metrics	Accuracy	✓
	Efficiency (Latency)	✓
	User Experience	✓

**Diverse, representative data** is important because it enables fair use (equal performance) across populations

# Achieving Ideal Metrics: Revisit pipeline



# Reading Material

# Main references

- [Harvard School of Engineering and Applied Sciences - CS249r: Tiny Machine Learning](#)
- [Professional Certificate in Tiny Machine Learning \(TinyML\) – edX/Harvard](#)
- [Introduction to Embedded Machine Learning \(Coursera\)](#)
- [Text Book: "TinyML" by Pete Warden, Daniel Situnayake](#)

I want to thank [Shawn Hymel](#) and [Edge Impulse](#), [Pete Warden](#) and [Laurence Moroney](#) from Google and specially Harvard professor [Vijay Janapa Reddi, Ph.D.](#) student [Brian Plancher](#) and their staff for preparing the excellent material on TinyML that is the basis of this course at UNIFEI.

The IESTI01 course is part of the [TinyML4D](#), an initiative to make TinyML education available to everyone globally.

**Thanks**  
And stay safe!

