

TF DEV SUMMIT 2020

KEYNOTE

Jinho Seo (synabreu@gmail.com)



TensorFlow

DEV SUMMIT 2020



76,000,000

downloads

80,000+

commits

13,000+

pull requests

2,400+

contributors



TF Dev Summit '19
TF 2.0 Alpha

TF Federated
TF Privacy

TF Graphics

TF 2.0 Beta

TF.text

SIG Addons

Roadshows
Shenzhen, Beijing,
Shanghai

TF at NeurIPS

MLIR

Roadshows
Nairobi, Lagos,
Johannesburg

Roadshow
Paris

Mar

Apr

May

Jun

Jul

Aug

Sept

Oct

Nov

Dec

tensorflow.org

Deeplearning.ai
and Udacity
courses

TF at
Google I/O

Google
Summer of
Code

Neural
Structured
Learning

TF 2.0 Final

Roadshow
São Paulo

TF World

Roadshow
Bengaluru

TF
Enterprise

Google
Code-In

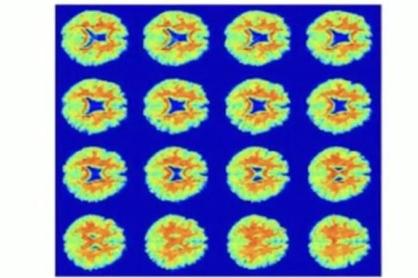
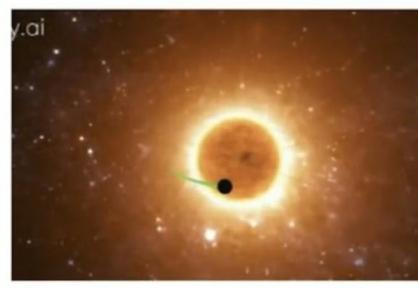


Simplified APIs

Easy model building with Keras and
eager execution

Robust model deployment anywhere

Powerful experimentation for research



1441

Bad Images

281 (2/006, 1/014)	280 (2/007, 1/021)	280 (2/007, 1/007)	280 (2/002, 1/016)	281 (2/004, 1/019)
282 (1/014, 1/004)	279 (2/012, 1/021)	282 (2/011, 1/004)	279 (2/012, 1/007)	280 (2/001, 1/005)

And here's a sample of what was labelled as 'good'.

1471 If this is just an example, you need the raw data to cut this script.

Mrs Bennet was really in a most pitiable state. The very mention of anything concerning the match threw her into an agony of ill-humour, and wherever she went she was sure of hearing it talked of. The sight of Miss Lucas was odious to her. As her successor in that house, she regarded her with jealous abhorrence. Whenever Charlotte came to see them, she concluded her to be anticipating the hour of possession; and whenever she spoke in a low voice to Mr Collins, was convinced that they were talking of the Longbourn estate, and resolving to turn herself and her daughters out of the house, as soon as Mr Bennet were dead. She complained bitterly of all this to her husband.

'Indeed, Mr Bennet,' [said she](#),
"it is very hard to think that Charlotte Lucas should ever be mistress of this house, that I should be forced to make way for her, and live to see her take her place in it!"

'My dear, do not give way to such gloomy thoughts. Let us hope for better things. Let us flatter ourselves that I may be the survivor.'



Neurascale

CLARA

Convo-Miner

Amazing Signs



Feedback on TF 2.0

Easy to use with better performance

Want migration to be easy



Emphasis on performance

Compatibility with the rest of the
TensorFlow ecosystem

Stability in the core library



TensorFlow

Ecosystem



TensorFlow Ecosystem

TensorFlow Core	tf.keras	TensorFlow Probability	Nucleus
TensorFlow.js	tf.data	Tensor2Tensor	TensorFlow Federated
TensorFlow Lite	TF Runtime	TensorFlow Agents	TensorFlow Privacy
TensorFlow Lite Micro	CoLab	Dopamine	Fairness Indicators
TensorBoard	TensorFlow Research Cloud	TRFL	Sonnet
TensorBoard.dev	MLIR	Mesh TensorFlow	Neural Structured Learning
TensorFlow Hub	TensorFlow Lattice	Ragged Tensors	JAX
TensorFlow Extended	Model Optimization Toolkit	TensorFlow Ranking	TensorFlow Quantum
Swift for TensorFlow	TensorFlow Graphics	Magenta	I/O and Addons



TensorFlow

Ecosystem

From research

To production

Deployed everywhere

Empowering Responsible AI

Powered by the community



TensorFlow

Ecosystem

From research

To production

Deployed everywhere

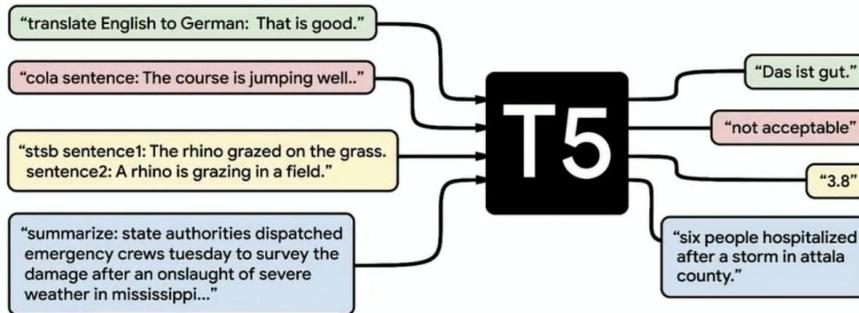
Empowering Responsible AI

Powered by the community



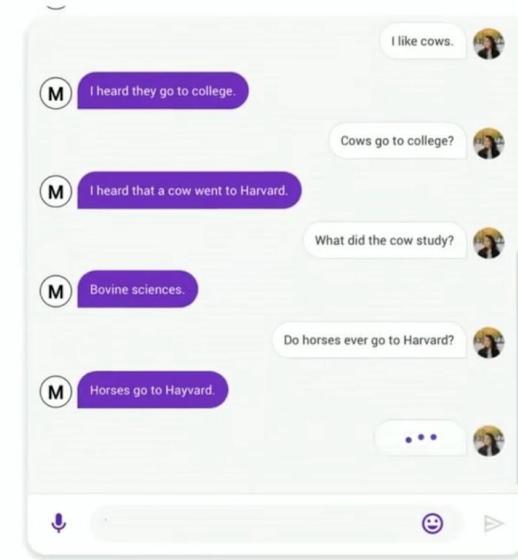
Cutting edge ML research with TensorFlow

T5: Talk-to-Text Transfer Transformer



[Arxiv.org: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer](https://arxiv.org/abs/1910.10683)

Meena: open-domain chat bot



[Arxiv.org: Towards a Human-like Open-Domain Chat Bot](https://arxiv.org/abs/1806.04375)



...and many more

Unsupervised Learning of Object Structure and Dynamics from Videos

Matthias Minderer*, Chen Sun, Ruben Villegas, Forrester Cole
 Kevin Murphy, Honglak Lee
 Google Research
 (mjm, chensun, rubenville, fcole, kpmurphy, honglak)@google.com

Trust Region-Guided Proximal Policy Optimization

Learning Conditional Deformable Templates with Convolutional Networks

Afshan V. Dalca*, Marianne Rakic, John Guttag, Mert R. Sabuncu
 CSAIL, MIT D-ITET, ETH CSAIL, MIT ECE and BME, Cornell

Abstract

We develop a learning framework for building deformable templates, which play a fundamental role in many image analysis and computational anatomy tasks. Conventional methods for template creation and image alignment to the template have undergone decades of rich methodical development. In these frameworks, template creation is often interleaved with steps of template estimation and alignment, which is often computationally very expensive. Due in part to this shortcoming, most methods compute a single template for the entire population of images, or a few templates for specific sub-groups of the data. In this work, we propose a learning framework for building deformable templates, either universal or *conditional* templates, jointly with a neural network that provides efficient alignment of the images to these templates. We demonstrate the usefulness of this method on a variety of domains, with a special focus on nonrigid registration. This is a methodological breakthrough, as creating a new template from scratch does not exist, or creating a new one with traditional methods can be prohibitively expensive. Our code and atlases are available online as part of the VoxelMorph library at <http://voxelmorph.csail.mit.edu>.

Keywords: deformable templates, conditional atlases, diffeomorphic image registration, probabilistic models, neuroimaging

8:0764v3 [cs.CV] 7 Dec 2019

Saccader: Improving Accuracy of Hard Attention Models for Vision

Gamaeldin F. Elsayed, Simon Kornblith, Quoc V. Le
 Google Research, Brain Team Google Research, Brain Team Google Research, Brain Team
 gamaeldin@google.com

Abstract

Although deep convolutional neural networks achieve state-of-the-art performance across nearly all image classification tasks, their decisions are difficult to interpret. One approach that offers some level of interpretability by design is *hard attention*, which uses only relevant portions of the image. However, training hard attention models with only class label supervision is challenging, and hard attention has proved difficult to scale to complex datasets. Here, we propose a novel hard attention model, which we term Saccader. Key to Saccader is a pretraining step that trains a hard attention model on a dataset of images and their ground truth attention optimization. Our best models narrow the gap to common ImageNet baselines, achieving 75% top-1 and 91% top-5 while attending to less than one-third of the image.

1 Introduction

Despite the success of convolutional neural networks (CNNs) across many computer vision tasks, their predictions are difficult to interpret. Because CNNs compute complex nonlinear functions of their inputs, it is often unclear what aspects of the input contributed to the prediction. Although many researchers have attempted to design methods to interpret predictions of off-the-shelf CNNs

DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction

in² Radomir Mech² Ulrich Neumann¹
²Adobe
 San Jose, California
 (ceylan, rmech}@adobe.com

Importance Weighted Hierarchical Variational Inference

Artem Sobolev, Dmitry Vetrov
 ng AI Center Moscow, Russia NRU HSE; Moscow, Russia
 obolev@bayesgroup.ru

Abstract

Inferential Inference is a powerful tool in the Bayesian modeling toolkit, however, effectiveness is determined by the expressivity of the utilized variational distributions in terms of their ability to match the true posterior distribution. In turn, expressivity of the variational family is largely limited by the requirement of a tractable density function. To overcome this roadblock, we introduce a family of variational upper bounds on a marginal log density in the case of graphical models (also known as latent variable models). We then give an upper bound on the Kullback-Leibler divergence and derive a family of increasingly *r* variational lower bounds on the otherwise intractable standard evidence bound for hierarchical variational distributions, enabling the use of more expressive posteriors. We show that previously known methods, such as Hierarchical Variational Models, Semi-Implicit Variational Inference and Doubly Implicit Variational Inference can be seen as special cases of the proposed approach, and empirically demonstrate superior performance of the proposed method on experiments.

Conclusion



Better experimentation with TensorFlow

TensorBoard.dev

Upload and share your ML experiments with anyone

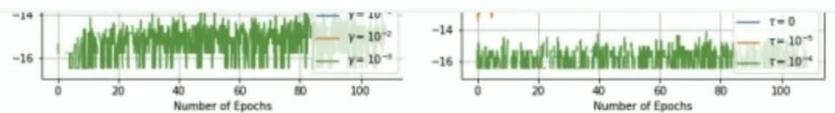


Figure 2: On the left τ is fixed at 10^{-5} and we compare three values for γ . On the right γ is fixed at 10^{-2} and we compare three values for τ . The tensorboard graphs can be found at <https://tensorboard.dev/experiment/up4UbhojT6uZKjnKPyZarQ>



Performance Profiler

Available in TensorBoard, Profiler provides overview of model performance and better debugging guidance





Eager at the core and simple, performant data input pipelines in 2.x

eager execution,
supporting numpy arrays

tf.data

TensorFlow Datasets

The screenshot shows the TensorFlow website's 'Datasets' section. The page title is 'TensorFlow Datasets: a collection of ready-to-use datasets.' Below the title, there is a brief description of what TensorFlow Datasets are and how they can be used. To the right of the description is a code editor containing Python code for working with the MNIST dataset using tf.data.Dataset.

```
import tensorflow.compat.v2 as tf
import tensorflow_datasets as tfds

# tfds works in both Eager and Graph modes
tf.enable_v2_behavior()

# Construct a tf.data.Dataset
ds = tfds.load('mnist', split='train', shuffle_files=True)

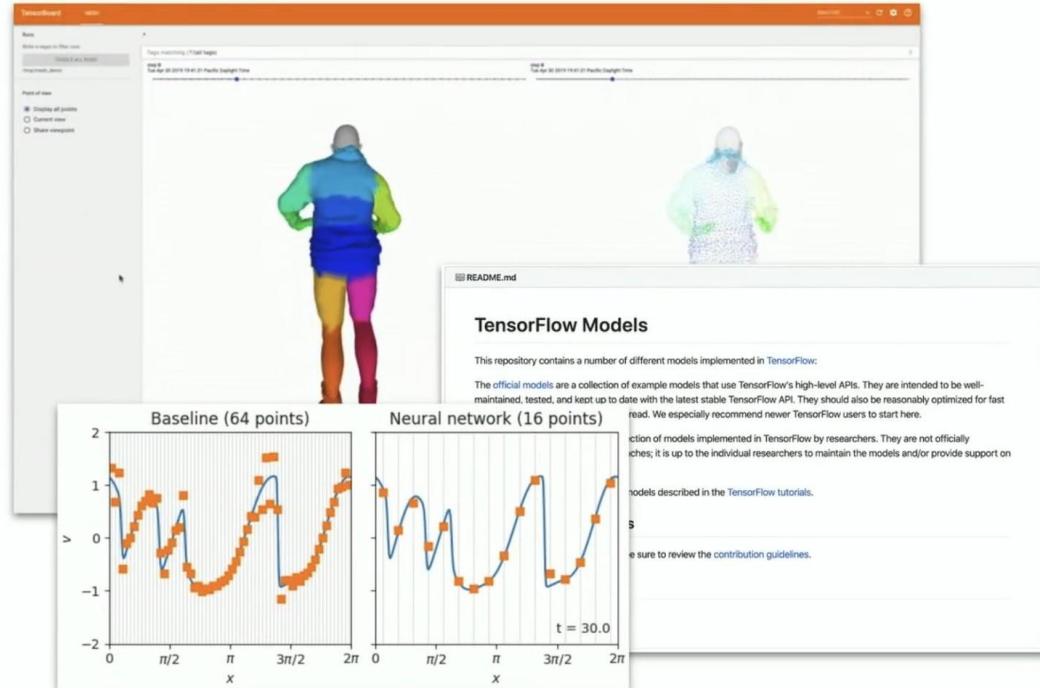
# Build your input pipeline
ds = ds.shuffle(1024).batch(32).prefetch(tf.data.experimental.AUTOTUNE)
for example in ds.take(1):
    image, label = example["image"], example["label"]
```

Below the code editor, there are two buttons: 'RUN IN COLAB' and 'NOTEBOOK'.



Add-ons and extensions to the TensorFlow ecosystem

- TF Probability
- TF Graphics
- Mesh TensorFlow
- TF Model Garden
- TF Agents
- TF Text
- Swift for TensorFlow
- Sonnet
- JAX
- Neural Structured Learning
- TF Quantum
- ...and more on tensorflow.org!





TensorFlow

Ecosystem

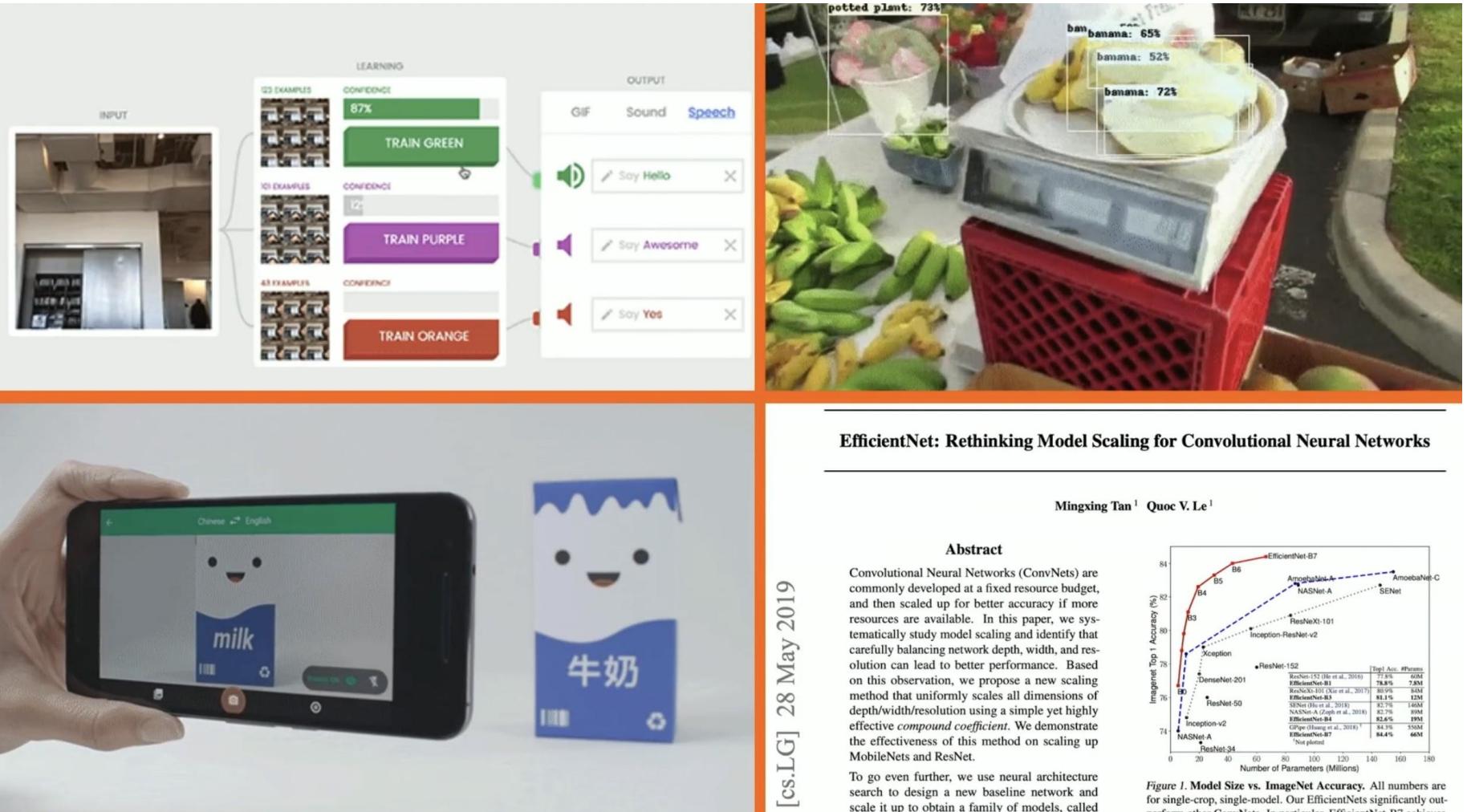
From research

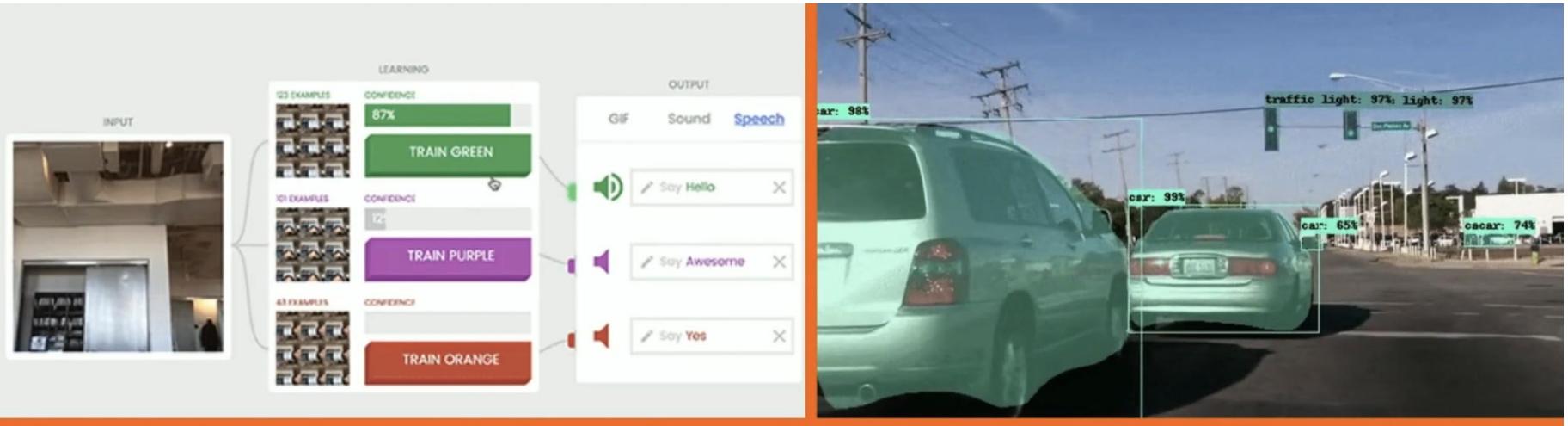
To production

Deployed everywhere

Empowering Responsible AI

Powered by the community





EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Mingxing Tan¹ Quoc V. Le¹

Abstract

Convolutional Neural Networks (ConvNets) are commonly developed at a fixed resource budget, and then scaled up for better accuracy if more resources are available. In this paper, we systematically study model scaling and identify that carefully balancing network depth, width, and resolution can lead to better performance. Based on this observation, we propose a new scaling method that uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective *compound coefficient*. We demonstrate the effectiveness of this method on scaling up MobileNets and ResNet.

To go even further, we use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called

[cs.LG] 28 May 2019

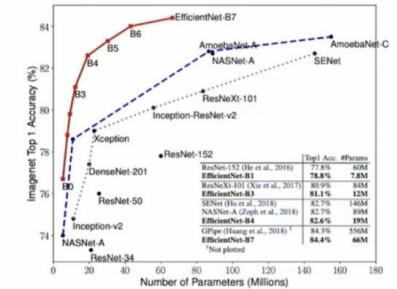
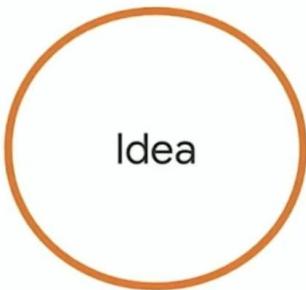
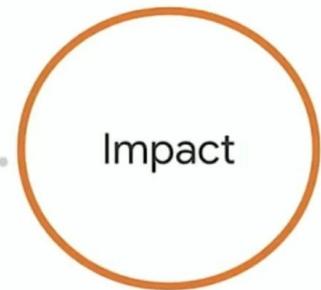


Figure 1. Model Size vs. ImageNet Accuracy. All numbers are for single-crop, single-model. Our EfficientNets significantly outperform other ConvNets. In particular, EfficientNet-B7 achieves



TensorFlow
Ecosystem





+





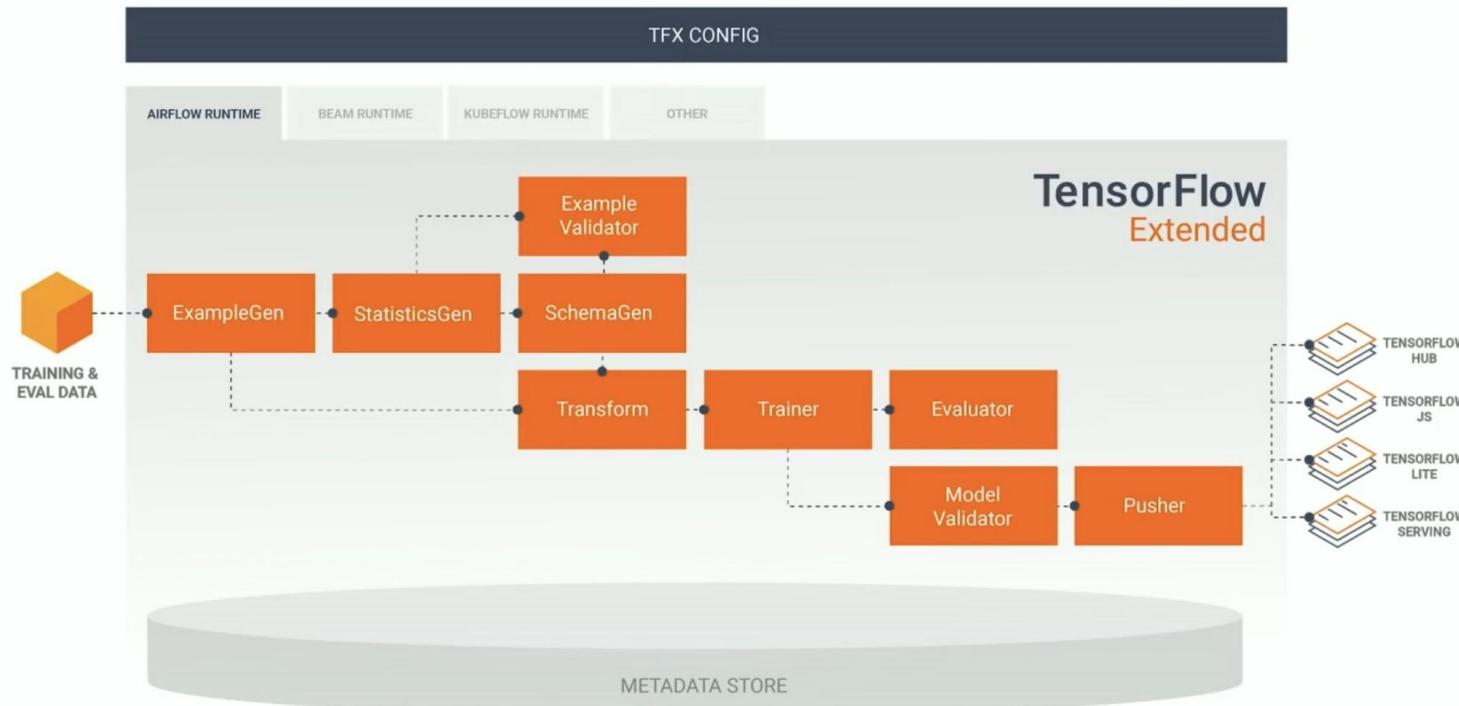
TensorFlow Hub

- Comprehensive collection of models across the TF ecosystem
- Pre-trained and ready to use
- Powered by the community, including:
 - DeepMind
 - Google
 - Microsoft AI for Earth
 - NVIDIA
 - The Metropolitan Museum of Art
 - Global Biodiversity Information Facility
 - And more...

The screenshot shows the TensorFlow Hub interface. At the top, there's a search bar with the text "text embedding spanish". Below it, a breadcrumb navigation shows "TensorFlow Hub" and "universal-sentence-encoder-multilingual". On the right, there's a "Jump to Model Formats" dropdown. The main content area has several sections: "Network" (16 languages), "Publisher" (Google), "Module type" (text-embedding), and "Format" (Hub Module). To the right, under "Model Details", there's a bulleted list: "Developed by researchers at Google, 2019, v1 [1].", "Convolutional Neural Net.", "Covers 16 languages, showing strong performance on cross-lingual retrieval.", and "A larger model with more complicated encoder architecture is available." Below this, a note says "To learn more about text embeddings, refer to the [TensorFlow Embeddings](#) documentation." Under "Intended Use", another bulleted list includes "The model is intended to be used for text classification, text clustering, semantic textual similarity retrieval, cross-lingual text retrieval, etc." At the bottom, there's a section titled "Factors".



Building for Production with TFX





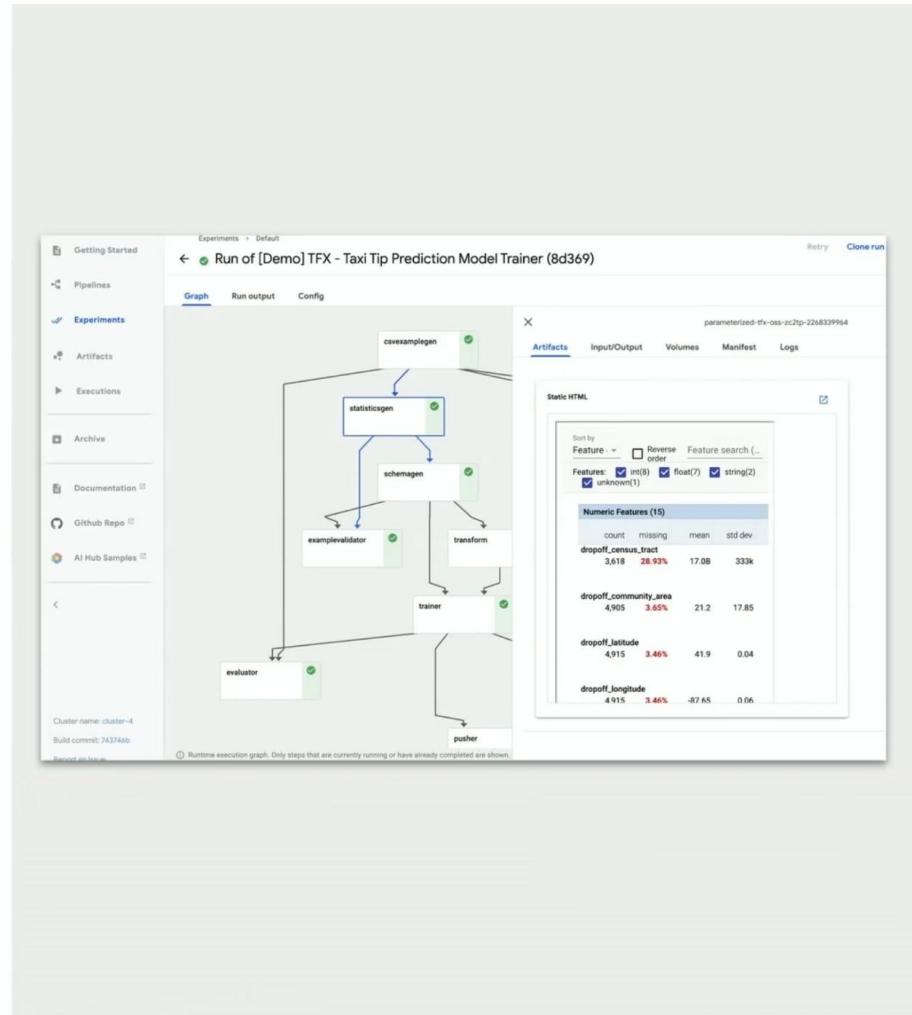
TensorFlow Extended



Google Cloud AI Platform Pipelines



TensorFlow
Enterprise

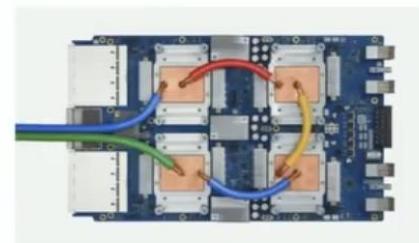




TensorFlow 2.1 supports Cloud TPUs



+





TensorFlow

Ecosystem

From research

To production

Deployed everywhere

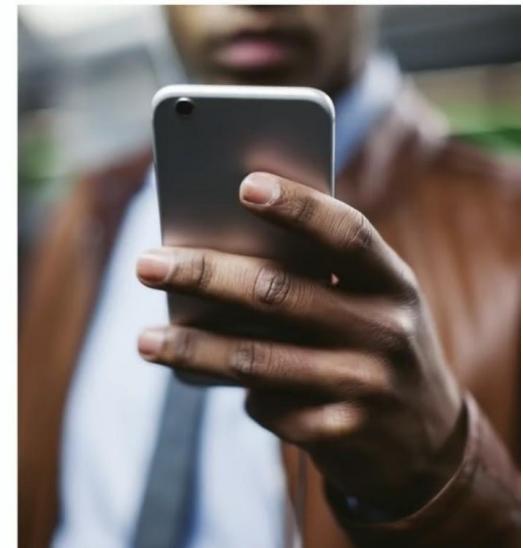
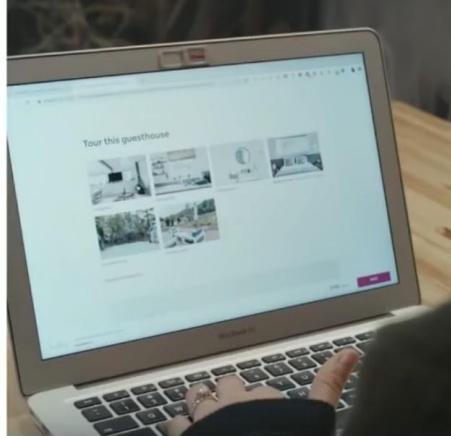
Empowering Responsible AI

Powered by the community



TensorFlow Ecosystem

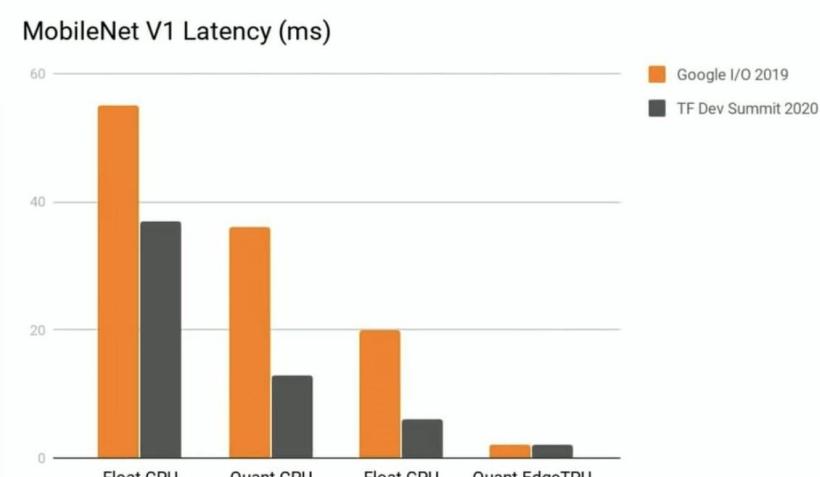
- Lower latency
- Network connectivity
- Privacy preserving





Latest in TensorFlow Lite

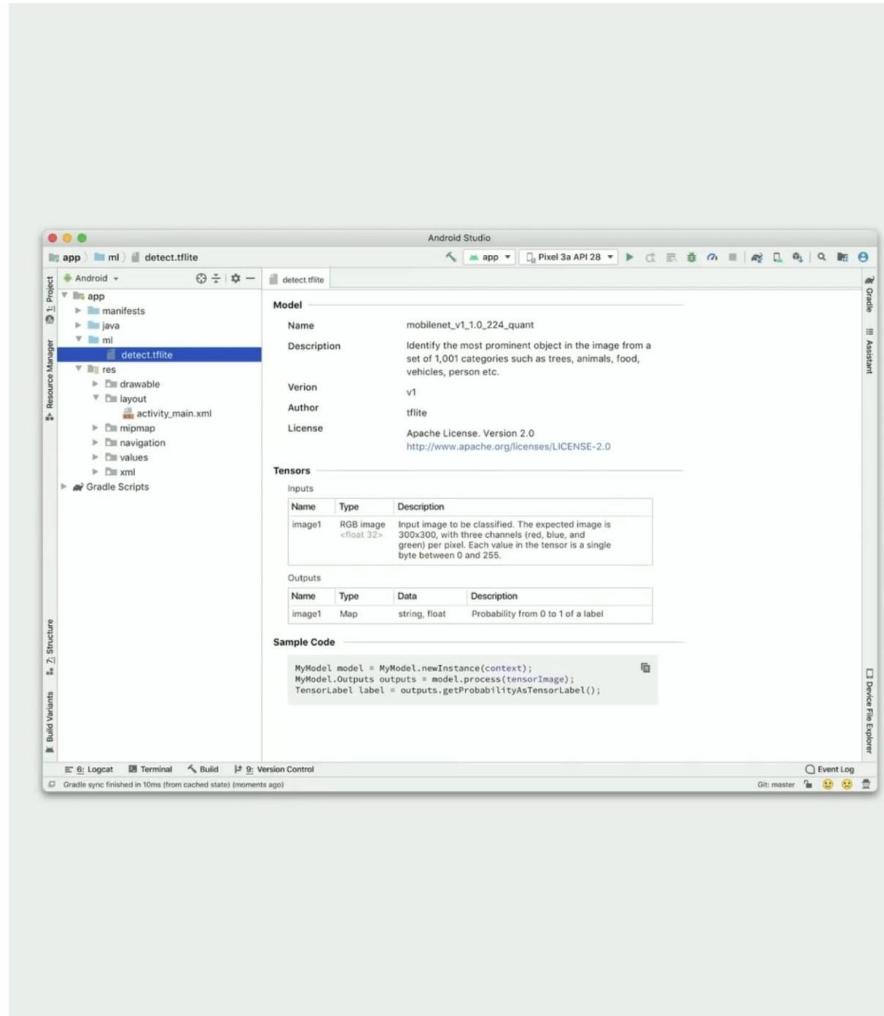
- Best in class performance acceleration





Latest in TensorFlow Lite

- Best in class performance acceleration
- Android Studio integration to Support Library





Latest in TensorFlow Lite

- Best in class performance acceleration
- Android Studio integration to Support Library
- New TF Lite Converter

MobileNet v1
MobileNet v2
MobileNet v3
EfficientNet

BlazeFace
MobileNet SSD
Mask R-CNN
Faster R-CNN

Inception v1
Inception v2
Inception v3
Inception v4

TensorFlow Lite

BERT
Mobile BERT
Smart Reply
Universal Sentence Encoder

DeepLab v3
Deep Speech v1
Deep Speech v2
Keyword Detector
PoseNet

Yolo
Yolo V3
Kinetics i3d
CycleGAN
Pix2Pix



Latest in TensorFlow.js

- New TF.js models:
 - FaceMesh
 - MobileBERT
- HuggingFace's NPM package for Question Answering in Node.js
- New WebAssembly backend for improved CPU performance

The collage consists of four images:

- A woman wearing glasses with a grid of blue dots overlaid on her face, representing FaceMesh tracking.
- A 3D wireframe mesh of a human head, likely representing a 3D reconstruction or a different type of mesh output from a model.
- A screenshot of a web application interface titled "Modify confidence score threshold." It shows a slider set to 0.71, a message about camera/lighting conditions, and a grid of 15 small images for training or testing. One image in the grid is labeled "ENDA".
- A screenshot of a video frame showing a person's hands holding a white mug with text on it, with a bounding box drawn around the mug area, demonstrating object detection or tracking.
- A screenshot of a GitHub README.md page for "Question Answering for Node.js". It includes a green button for "npm package 1.4.0", a heading "Question Answering for Node.js", and a paragraph stating "Production-ready Question Answering directly in Node.js, with only 3 lines of code!" followed by a detailed description of the package's purpose and performance.



MLIR

- Largest HW partners in the world
- 95% of the world's data-center accelerator hardware
- 4 billion mobile phones countless IoT devices
- Governance moved to LLVM



GRAPHCORE

habana



MEDIATEK



Qualcomm

SambaNova
SYSTEMS

SAMSUNG

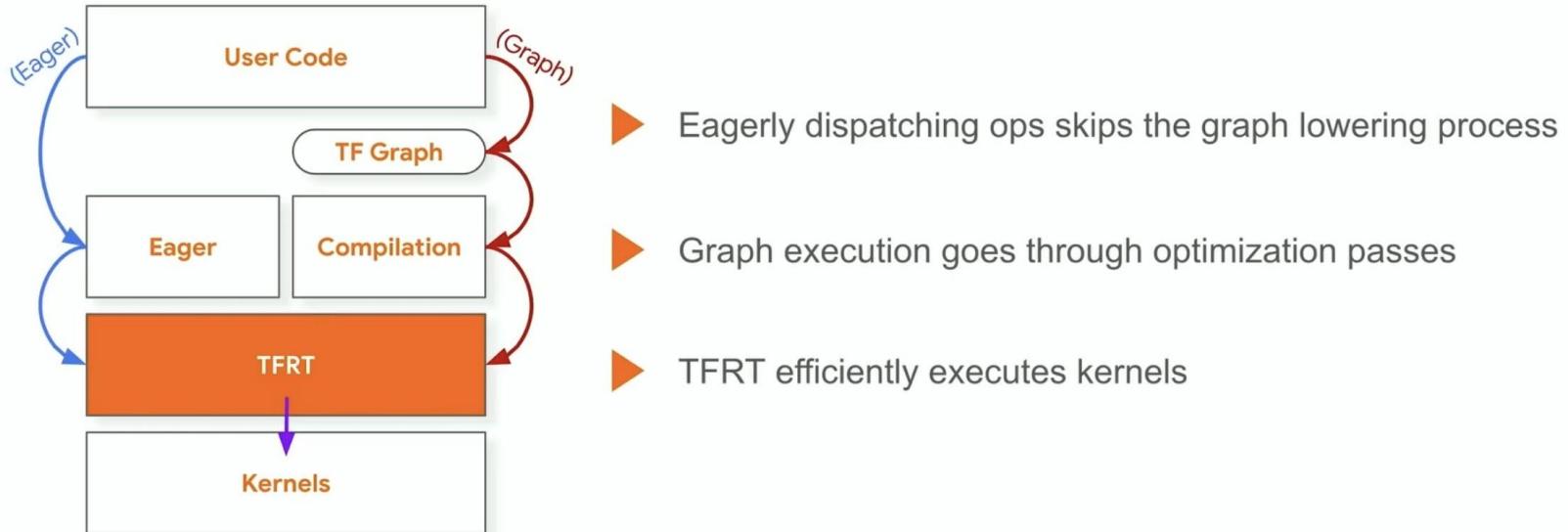


XILINX.



TensorFlow Runtime (TFRT)

Key to the model building workflow





TensorFlow

Ecosystem

From research

To production

Deployed everywhere

Empowering Responsible AI

Powered by the community



Responsible AI starts with defining it

1. General recommended practices for AI
2. Fairness
3. Interpretability
4. Privacy
5. Security



Define
Problem

Construct
and Prepare
Data

Build and
Train Model

Deploy

Iterate



Who is my ML system for?
Am I using a representative dataset?
How is my model performing?
What can I do to improve the model?

Is there real-world / human bias in my data?

Define Problem
Construct and Prepare Data
Build and Train Model
Deploy
Iterate

Are there any privacy considerations?
Where do I get relevant features in a privacy preserving way?
Are test users diverse?
How does my data affect model performance?
Should I deploy my model?
Are there complex feedback loops?



Responsible AI tools in the TensorFlow ecosystem

- Fairness Indicators
- What-If Tool
- TensorFlow Federated
- TensorFlow Privacy
- People + AI Research (PAIR) Guidebook
- Model cards
- Crowdsource app



TensorFlow

Ecosystem

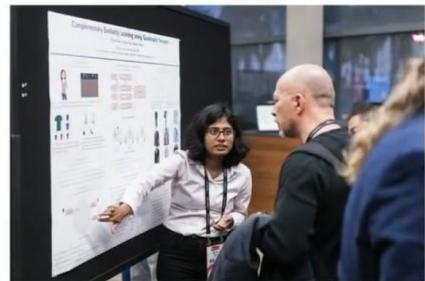
From research

To production

Deployed everywhere

Empowering Responsible AI

Powered by the community





TensorFlow User Groups



Interested in creating one? Send an email to tfug-help@tensorflow.org

Special Interest Groups

Working better together

SIG Graphics **available soon!*

SIG Keras

SIG Addons

SIG MLIR

SIG Build

SIG Swift

SIG IO

SIG TensorBoard

SIG Networking

SIG Micro

SIG JVM

SIG Rust



Community

Google Developer Experts

148 ML GDES Globally

- Host Techtalks
- Run Workshops
- Organize Doc Sprints





Community

Kaggle competition now supports 2.x

k

No setup

No provisioning

Datasets in optimal formats

TPUs and GPUs provided at no cost to
users



Community

TF 2.2 Challenge: Building AI Responsibly

A new DevPost challenge
using TF 2.x to build ML
systems

Fairness

Interpretability

Privacy

Security

responsible-ai.devpost.com



Google Code-in

A contest to introduce
pre-university students
(ages 13-17) to open source
software development.

1324 tasks completed.

769 students with completed tasks.

247 organization members,
overwhelmingly from the external
TensorFlow community.

54 published tasks.



Community

Google Summer of Code

A program for students to work
with an open source
organization on a 3 month
programming project

2020 Applications open March 16-31



SummerofCode.withgoogle.com



Updated: Machine Learning Crash Course

TensorFlow 2.0

We're setting students up for success with exercises using TF 2.0

Keras

Students will now build and train their models with `tf.keras`.

Improved Structure

New exercises have clearer instructions

g.co/machinelearningcrashcourse



Education

Request for Proposals

Design, develop, and teach
undergraduate or
graduate-level ML courses

Submissions are now open
through July 30th, 2020.

Contact:
tensorflow-rfp@google.com



TensorFlow: Data and Deployment

Learn how to develop and deploy machine learning models across any device or platform faster and more accurately than ever



deeplearning.ai

coursera



Education



Certificate in TensorFlow Development

TensorFlow.org/certificate

Topics:

- Basic ML programming concepts in TensorFlow
- Text Classification
- Computer Vision
- Sequences and Prediction

**Stipends available*



Community engagement

- Start a TFUG

tfug-help@tensorflow.org

- Join a Special Interest Group

tensorflow.org/community/forums

- Apply for Google Summer of Code

summerofcode.withgoogle.com

- Compete on Kaggle

kaggle.com/competitions

- Responsible AI Dev Post Challenge

responsible-ai.devpost.com

Education Resources

- Take Machine Learning Crash Course

g.co/machinelearningcrashcourse

- Take a specialization course

coursera.org/specializations/tensorflow-data-and-deployment

- Submit a proposal for an ML course

Tensorflow-rfp@google.com

- Get certified

TensorFlow.org/certificate

Antibiotic pellet identification





TensorFlow