Experts

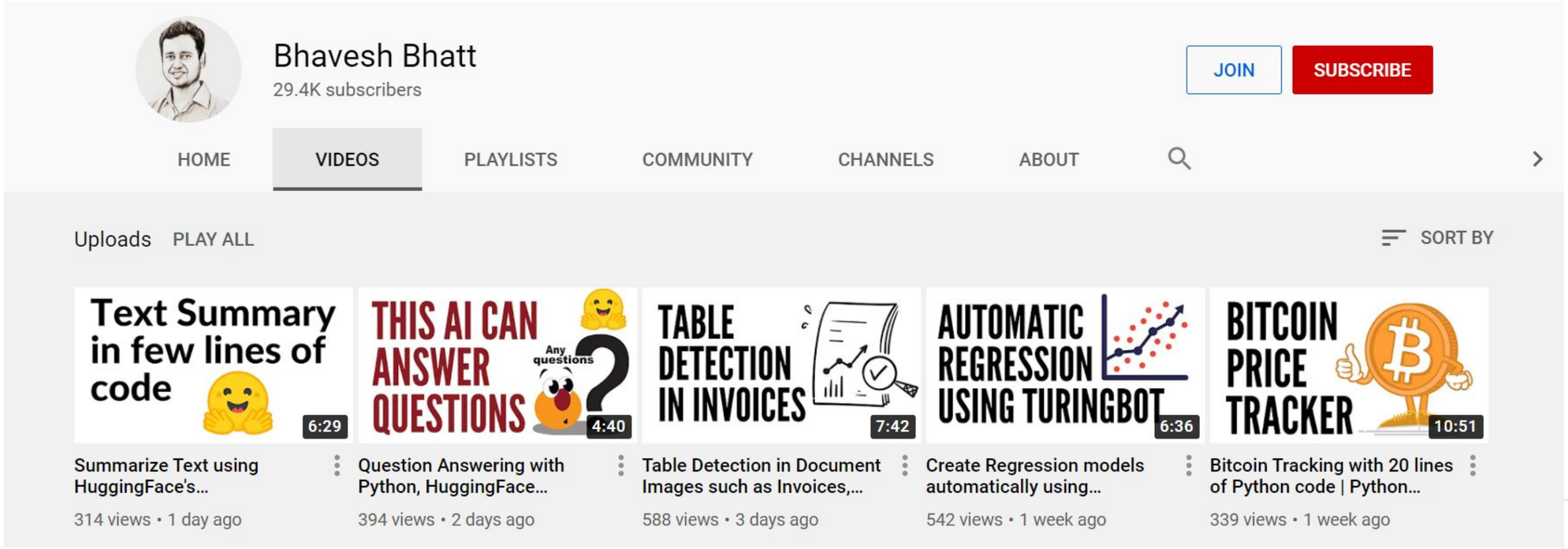
Introduction to TensorFlow Hub

Bhavesh Bhatt
@_bhaveshbhatt

Google Developers

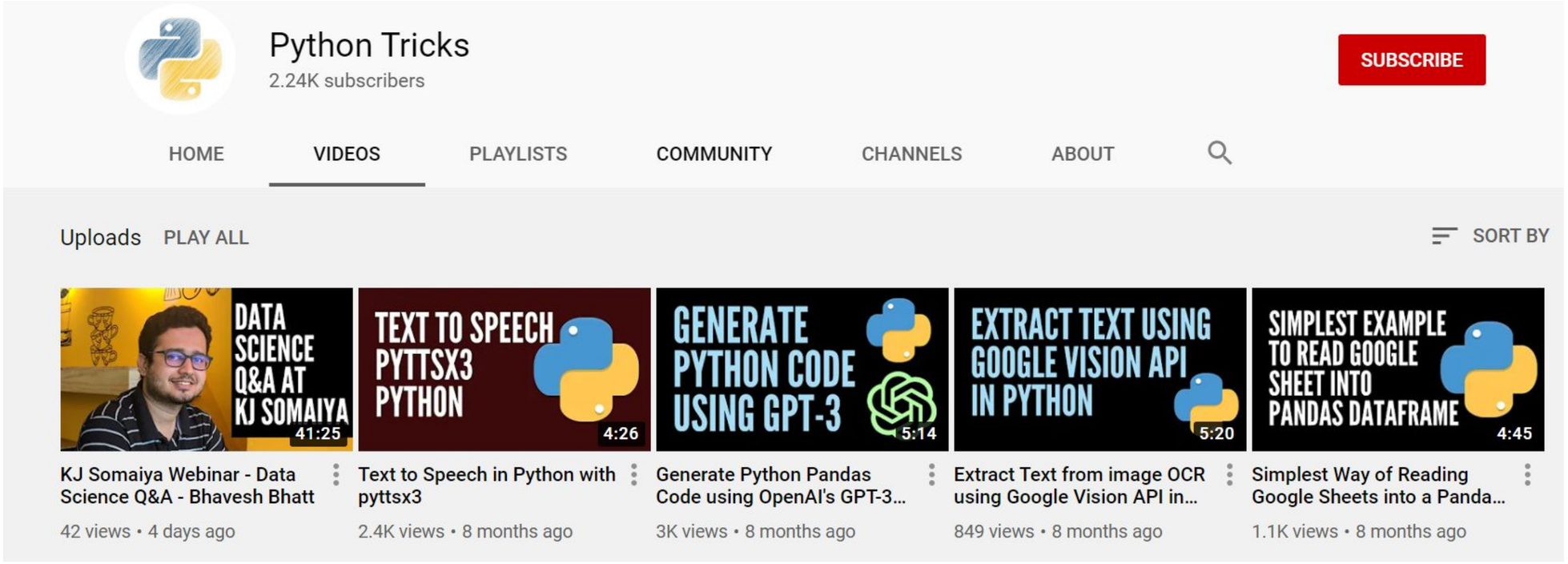
echo $(whoami)

- [Data Science/Machine Learning YouTube Channel]



echo $(whoami)

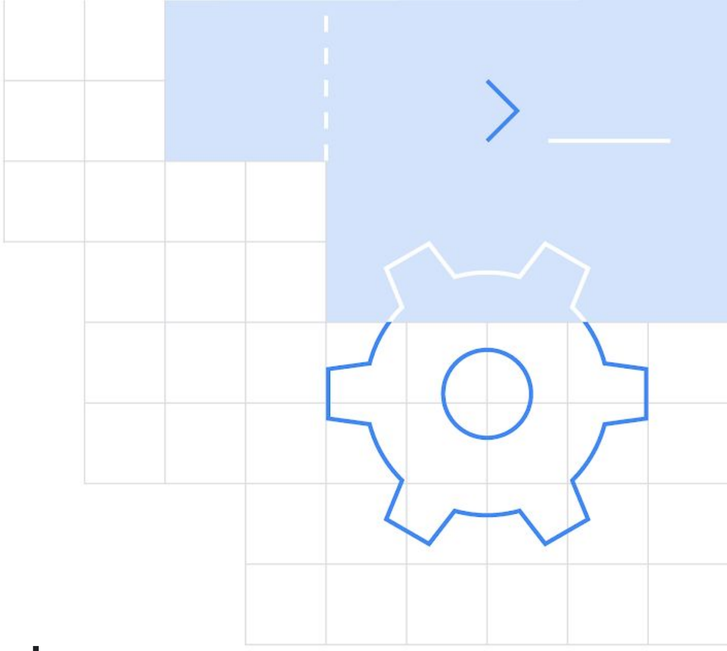
- [Python Channel]



echo $(whoami)

- Google Developer Expert (Machine Learning)

- Awarded the prestigious 40 Under 40 Data Scientist award by Analytics India Magazine in January 2020.



Ideal audience

- ML Developers that have worked with TensorFlow and Keras

Experts

Agenda

- What is TensorFlow Hub?
- Why you should use it?
- TF-Hub Code Walk along
- Q&A

Challenge

arXiv:1810.04805v2 [cs.CL] 24 May 2019

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**
Google AI Language
{jacobdevlin,mingweichang,kentonl,kristout}@google.com

Abstract

We introduce a new language representation model called **BERT**, which stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. Unlike recent language representation models (Peters et al., 2018a; Radford et al., 2018), BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.

BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE score to 80.5% (7.7% point absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), SQuAD v1.1 question answering Test F1 to 93.2 (1.5 point absolute improvement) and SQuAD v2.0 Test F1 to 83.1 (5.1 point absolute improvement).

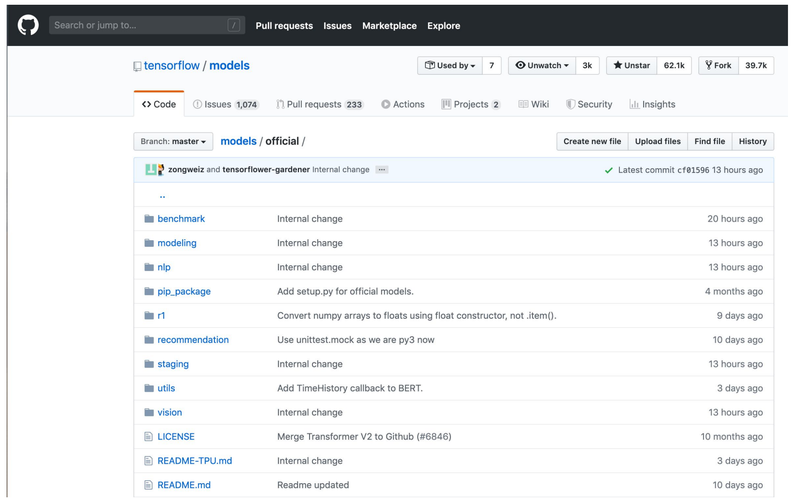
1 Introduction

Language model pre-training has been shown to be effective for improving many natural language processing tasks (Dai and Le, 2015; Peters et al., 2018a; Radford et al., 2018; Howard and Ruder, 2018). These include sentence-level tasks such as natural language inference (Bowman et al., 2015; Williams et al., 2018) and paraphrasing (Dolan and Brockett, 2005), which aim to predict the relationships between sentences by analyzing them holistically, as well as token-level tasks such as named entity recognition and question answering, where models are required to produce fine-grained output at the token level (Tjong Kim Sang and De Meulder, 2003; Rajpurkar et al., 2016).

There are two existing strategies for applying pre-trained language representations to downstream tasks: *feature-based* and *fine-tuning*. The feature-based approach, such as ELMo (Peters et al., 2018a), uses task-specific architectures that include the pre-trained representations as additional features. The fine-tuning approach, such as the Generative Pre-trained Transformer (OpenAI GPT) (Radford et al., 2018), introduces minimal task-specific parameters, and is trained on the downstream tasks by simply fine-tuning *all* pre-trained parameters. The two approaches share the same objective function during pre-training, where they use unidirectional language models to learn general language representations.

We argue that current techniques restrict the power of the pre-trained representations, especially for the fine-tuning approaches. The major limitation is that standard language models are unidirectional, and this limits the choice of architectures that can be used during pre-training. For example, in OpenAI GPT, the authors use a left-to-right architecture, where every token can only attend to previous tokens in the self-attention layers of the Transformer (Vaswani et al., 2017). Such restrictions are sub-optimal for sentence-level tasks, and could be very harmful when applying fine-tuning based approaches to token-level tasks such as question answering, where it is crucial to incorporate context from both directions.

In this paper, we improve the fine-tuning based approaches by proposing BERT: **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. BERT alleviates the previously mentioned unidirectionality constraint by using a "masked language model" (MLM) pre-training objective, inspired by the Cloze task (Taylor, 1953). The masked language model randomly masks some of the tokens from the input, and the objective is to predict the original vocabulary id of the masked



How do I use it?

Is it safe?

Is it fair?

Is it the latest version?

What is TensorFlow Hub?

A collection of SoTA* pre-trained models published by different teams as well community contributors.

*State of The Art

TensorFlow Hub

A comprehensive collection of models

Image

Text

Video

Audio

Model Diversity

Text

- Classification
- Embeddings
- Generation
- Question Answering
- ... your new model

Image

- Classification
- Object detection
- Semantic segmentation
- Generators
- Style transfer
- Embeddings / feature vectors
- Augmentation
- Pose detection
- ... your new model

Audio

- Classification
- Embeddings / Features
- ... your new model

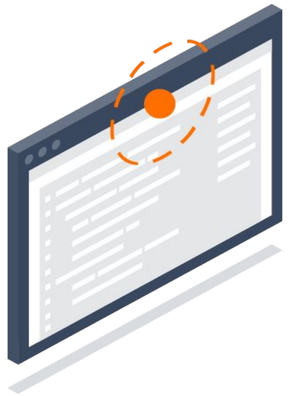
Video

- Classification
- Generation
- ... your new model

Ready to use

Pre-trained models ready for transfer learning on your own datasets and deployable anywhere you want

TensorFlow Extended



TensorFlow .JS

TensorFlow Lite

Coral

TensorFlow Hub

Search for anything, from models to data sets

Filters Clear all

- Problem domain
- Saved model format
- Architecture
- Publisher
- Dataset
- Language

Text embedding
nnlm-en-dim128-with-normalization
Published by: Google Updated: 10/24/2019
Token based text embedding trained on English Google News 200B corpus.
NNLM | Google News

Collection
universal-sentence-encoder
Published by: Google Updated: 10/24/2019
Collection of universal sentence encoders trained on variety of data.
DAN,Transformer

Image pose detection .JS
posenet/mobilenet/quantized/1/100
Published by: TensorFlow Updated: 10/24/2019
PoseNet model for pose esimation.

Image classification
imagenet/mobilenet_v1_025_224...
Published by: Google Updated: 10/24/2019
Imagenet (ILSVRC-2012-CLS) classification with MobileNet V1 (depth multiplier 0.25).
MobileNet V1 | ImageNet (ILSVRC-201...

Image classification .JS
imagenet/mobilenet_v2_075_224...
Published by: Google Updated: 10/24/2019
Imagenet (ILSVRC-2012-CLS) classification with MobileNet V2 (depth multiplier 0.75).
MobileNet V2 | ImageNet (ILSVRC-201...

Image classification .JS
imagenet/mobilenet_v1_025_128...
Published by: Google Updated: 10/24/2019
Imagenet (ILSVRC-2012-CLS) classification with MobileNet V1 (depth multiplier 0.25).
MobileNet V1 | ImageNet (ILSVRC-201...

Image segmentation .JS
bodypix_050
Published by: TensorFlow Updated: 10/24/2019
A person segmentation model.

Other
llr-pretrain-adv/latents
Published by: DeepMind Updated: 10/24/2019
LLR-pretrained ResNet-152 on ImageNet then adversarially trained for the unrestricted adversarial
Other | Other

Image generator
compare_gan/s3gan_20_128x128
Published by: Google Updated: 10/24/2019
S3GAN trainend on ImageNet with 20% labels.
Other | ImageNet (ILSVRC-201...

Image pose detection .JS
posenet/mobilenet/quantized/2/050
Published by: TensorFlow Updated: 10/24/2019
PoseNet model for pose esimation.

Text embedding
universal-sentence-encoder-xling/en-fr
Published by: Google Updated: 10/24/2019
English and French language-agnostic text encoder.
Transformer

Publisher
tensorflow
TensorFlow

Image classification
imagenet/pnasnet_large/classification
Published by: Google Updated: 10/24/2019
Imagenet (ILSVRC-2012-CLS) classification with PNASNet-5 (large).
PNASNet-5 (large) | ImageNet (ILSVRC-201...

Image feature vector .JS
imagenet/mobilenet_v1_025_192...
Published by: Google Updated: 10/24/2019
Feature vectors of images with MobileNet V1 (depth multiplier 0.25) trained on ImageNet (ILSVRC-2012-
MobileNet V1 | ImageNet (ILSVRC-201...

Image generator
biggan-deep-256
Published by: DeepMind Updated: 10/24/2019
BigGAN-deep image generator trained on 256x256 ImageNet.
Other | ImageNet (ILSVRC-201...

Image pose detection .JS
posenet/mobilenet/float/050
Published by: TensorFlow Updated: 10/24/2019
PoseNet model for pose esimation.

Image feature vector
vae
Published by: Vtab Updated: 10/24/2019
Visual representation obtained by training a VAE on ImageNet.
Other | ImageNet (ILSVRC-201...

Image generator
compare_gan...
Published by: Google Updated: 10/24/2019
ResNet19 trained on CelebA HQ (128x128) (FID: 35.85).
Other | CelebA HQ

Text embedding
bert_en_cased_L-24_H-1024_A-16
Published by: TensorFlow Updated: 10/24/2019
Bidirectional Encoder Representations from Transformers (BERT).
Transformer | Wikipedia and BooksC...

Text embedding
tf2-preview/nnlm-de-dim50-with-...
Published by: Google Updated: 10/24/2019
Token based text embedding trained on German Google News 30B corpus.
NNLM | Google News

Image feature vector .JS
imagenet/mobilenet_v2_100_160...
Published by: Google Updated: 10/24/2019
Feature vectors of images with MobileNet V2 (depth multiplier 1.00) trained on ImageNet (ILSVRC-2012-
MobileNet V2 | ImageNet (ILSVRC-201...

Text embedding
nnlm-ja-dim128-with-normalization
Published by: Google Updated: 10/24/2019
Token based text embedding trained on Japanese Google News 6B corpus.
NNLM | Google News

Image classification .JS
imagenet/mobilenet_v2_100_224...
Published by: Google Updated: 10/24/2019
Imagenet (ILSVRC-2012-CLS) classification with MobileNet V2 (depth multiplier 1.00).
MobileNet V2 | ImageNet (ILSVRC-201...

Image classification
unsupervised-adversarial-training...
Published by: DeepMind Updated: 10/24/2019
UAT++ adversarially trained WRN-106 (wide residual network) model using 80m@200K unlabeled data and
Other | CIFAR-10

Image feature vector
imagenet/resnet_v2_152/feature_vector
Published by: Google Updated: 10/24/2019
Feature vectors of images with ResNet V2 152 trained on ImageNet (ILSVRC-2012-CLS).
ResNet V2 152 | ImageNet (ILSVRC-201...

Image generator
biggan-deep-512
Published by: DeepMind Updated: 10/24/2019

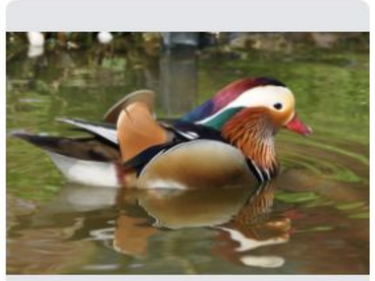
Image feature vector
rotation
Published by: Vtab Updated: 10/24/2019

Image feature vector
imagenet/mobilenet_v1_100_128...
Published by: Google Updated: 10/24/2019

Image classification TFLite
mobilenet_v1_0.25_224
Published by: TensorFlow Updated: 10/24/2019

Image classification TFLite
mobilenet_v1_0.75_192
Published by: TensorFlow Updated: 10/24/2019

Birds V1.1



Drag/click to upload image | Type/paste i



Published by: Google | Terms of Service

Results:

Type	Score
Aix galericulata	90.8%
Aix sponsa	0.9%
Dendrocygna bicolor	0.7%
Alopochen aegyptiaca	0.3%

Embed

Disease-classification.1

Drag/click to upload image | Type/paste



Published by: AgriPredict | Terms of Service

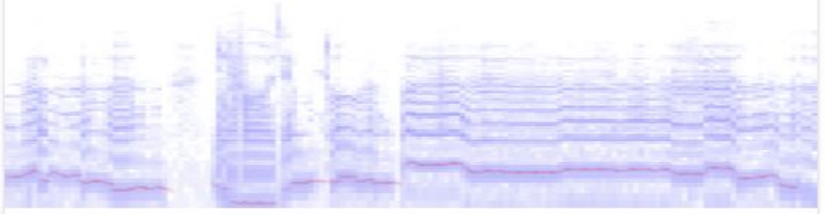
Results:

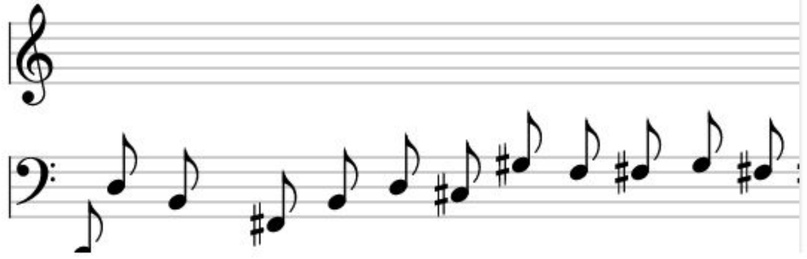
Type	Score
Soy Downy Mildew	99.3%
Soy Frogeye Leaf Spot	0.7%
Soy Healthy	0%
Tomato Yellow Leaf Curl Vi...	0%

Embed

SPICE Demo

132 Hz | C | 97 %

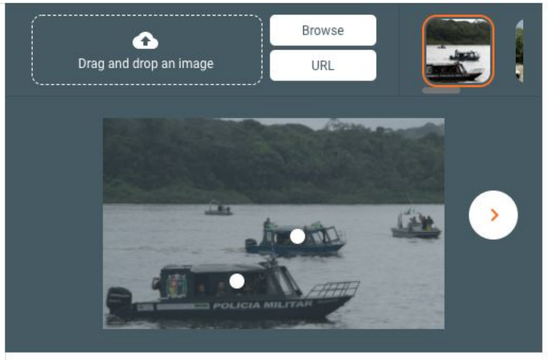




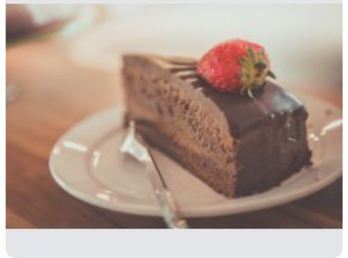
SSD Openimages v4

Results

Label: Boat



Food V1.1



Drag/click to upload image | Type/paste i



Published by: Google | Terms of Service

Results:

Type	Score
Sachertorte	82.1%
Black Forest gateau	2.8%
Devil's food cake	2.3%
Chocolate brownie	1.4%

Embed

iMet Collection Attribute Classifier

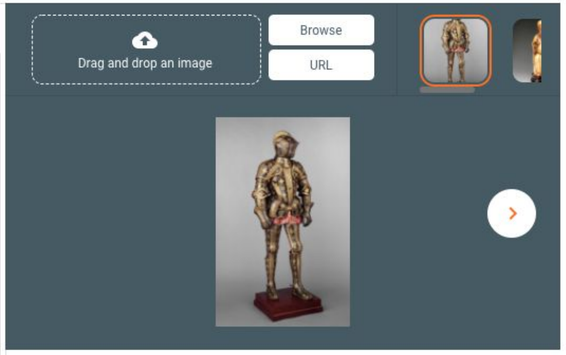
Results

armors
Score: 0.87

culture::italian
Score: 0.87

culture::milan
Score: 0.38

shields
Score: 0.26



Privacy Policy | Terms of Service

BETA version

Problem domain

Image feature vector

Architecture

MobileNet V2

Publisher

Google

Dataset

ImageNet (ILSVRC-2012-CLS)

Format: TF2.0 Saved Model

Fine tunable: Yes

License: Apache-2.0

Last updated: 2020-02-20

Model formats

Saved Model | .JS (v1, default) | .JS (v2, default) | .JS (v3, default)

Want to use this model?

To use this model, take a look at the example code, or at our user guide.

You can also try out the associated Colab.

Copy URL to clipboard

Download Model

Open Colab Notebook

Asset size: 2.62MB

TF2 SavedModel

This is a SavedModel in TensorFlow 2 format. Using it requires TensorFlow 2 (or 1.15) and TensorFlow Hub 0.5.0 or newer.

Overview

MobileNet V2 is a family of neural network architectures for efficient on-device image classification and related tasks, originally published by

- Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen: "Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation", 2018.

Mobilenets come in various sizes controlled by a multiplier for the depth (number of features) in the convolutional layers. They can also be trained for various sizes of input images to control inference speed.

Code Walkthrough

> "Yeah I can do machine learning"
>
> `import tensorflow`
>
> "You're hired."

Q&A


THANK YOU FOR ATTENTION
Q&A SESSION
Spongebob Squarepants vector trace

Slides & Code available here -

Experts

Experts

Thank You!

Bhavesh Bhatt
@_bhaveshbhatt