Model Card Version: 1.0_2023
License: Apache 2.0

Guideline Line Segmentation Model

Model Page:
https://github.com/google-research/project-guideline/tree/main/project_guideline/vision/models

A DeepLabv3+ based model predicting purple guiding lines in certain outdoor environments (e.g. street, park, trail)). The model is lightweight (0.5MB size) and runs super-real-time (~500FPS on Pixel 7 Edge TPU, ~70FPS on Pixel 7 CPU with XNNPACK, ~50FPS on Pixel 7 GPU).

Model Snapshot

Model Overview

MODEL ARCHITECTURE	INPUT(S)	OUTPUT(S)
DeepLabV3+ with MobileNetV3-small backbone.	3-channel RGB image of size [1, 513, 513, 3] in uint8 ([0, 255]). The image may contain a purple line to be detected of approximate reference color RGB [144, 99, 205].	Segmentation mask of size [1, 513, 513, 2] in float32 . The first channel is the probability belonging to *background* class. The second channel is the probability of the *purple line* class.

Usage

APPLICATION	BENEFITS	KNOWN CAVEATS
Used in Project Guideline for vision impaired navigation.	The model is trained to detect a purple line in a color specifically chosen to have high-contrast on most road surfaces and be distinguishable from other common road markings used for traffic and pedestrian control.	The model assumes a single solid purple line in the range of 3-6 inch on the ground. It may fail to detect the line under certain environmental conditions (adverse weather, abnormal lighting, etc), and may incorrectly classify incidental

	The model is lightweight and very fast (~2ms latency on Pixel 7 Edge TPU), making it ideal for low-power mobile applications.	objects of similar purple color as part of the line. Post-processing can be used to further refine the line estimation, including aggregating results across multiple frames.

Model Creators

MODEL CONTACT	MODEL AUTHOR(S)	CITATION
Xuan Yang, Google (xuanyang@google.com)	Xuan Yang, Google; Kimberly Wilber, Google; Liang-chieh Chen, Google; Fan Yang, Google	Unavailable

System Type

SYSTEM DESCRIPTION	UPSTREAM DEPENDENCIES	DOWNSTREAM DEPENDENCIES
Intended to be used as part of the Project Guideline system, but can also be used as a standalone model for related use cases.	Camera image resized to input resolution 513 x 513, using center-crop for best results.	Argmax needs to be performed on the last axis if downstream requires semantic class prediction.

Implementation Frameworks

HARDWARE & SOFTWARE FOR TRAINING	HARDWARE & SOFTWARE FOR DEPLOYMENT
• GPU (NVidia A100) • TensorFlow v2	• Pixel 6+ device, Edge TPU • MediaPipe v0.10 Image Segmenter • TensorFlow Lite v2

Compute Requirements

COMPUTE REQUIREMENTS FOR FINE-TUNING*		COMPUTE REQUIREMENTS FOR INFERENCE*	
Number of Chips	4	Number of Chips	1
Training Time (days)	<0.1	Training Time (days)	N/A
Total Computation (floating pt operations)	Unavailable	Total Computation (floating pt operations)	Unavailable
Measured Performance (TFLOPS/s)	Unavailable	Measured Performance (TFLOPS/s)	Unavailable
Energy Consumption (MWh)	Unavailable	Energy Consumption (MWh)	Unavailable

*Modeled after Patterson, David, et al. "Carbon emissions and large neural network training." arXiv preprint arXiv:2104.10350 (2021).

Model Characteristics

MODEL INITIALIZATION		MODEL STATUS		MODEL STATS	
The model is initialized from a DeepLabV3+ checkpoint trained on Cityscapes dataset.		The model is static and not updated regularly.		The model is composed of a MobileNetV3-small based encoder and a lightweight decoder, connected in a U-Net fashion.	
Training Epochs	28	**Dataset Name**	Internal Guideline Synthetic Data, Internal Guideline Real Data	**Size**	0.5MB
Base Learning Rate	0.01	**Version**	1.0	**Weights**	940K
Method	Follow the default training configuration of DeepLabv3 for Cityscapes.	**Release Date**	July 2023	**Layers**	There are 13 convolution blocks in the MobileNet backbone, followed by an ASPP module to perform feature extraction. The decoder fuses the multi-scale features by a simple sum operation.
Loss	Softmax for pixel classification	**Update Cadence**	N/A	**Latency**	~2ms (Edge TPU)

PRUNING		QUANTIZATION		DIFFERENTIAL PRIVACY
No		Yes, uint8		None
Methods	N/A	**Methods**	quantization-aware training	
Structuring	N/A	**Pre-quantized Representation**	fp32	
Sparsity Level	N/A	**End Bit Representation**	uint8	

Number of Params at Sparsity	N/A	**Hardware**	Edge TPU/GPU
Accuracy at Final Sparsity after Training	N/A		
Perplexity at Final Sparsity after Training	N/A		

Data Overview

TRAINING DATASET SNAPSHOT	DATASET MAINTENANCE & VERSIONS	INSTRUMENTATION
There are 2 sources of training data: • Internal real data collected from mobile phone videos in outdoor environments, manually annotated with ground truth line masks. • Internal synthetic data generated from virtual 3d scenes containing lines and adversarial objects, automatically annotated with ground truth line masks as part of the render pipeline.	The dataset is static.	No notable instrumentation used in collection or preprocessing of data.

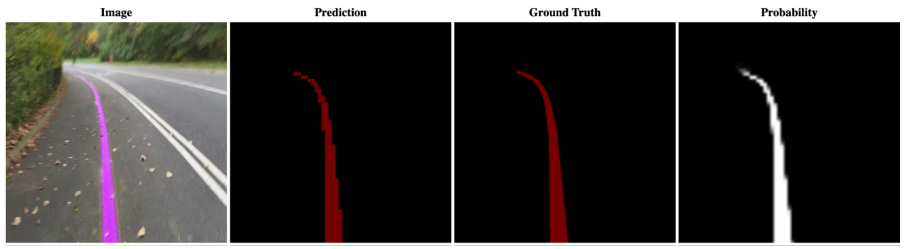
Dataset Size	~114K images	**Current Version**	1.0	**Instrumentation Criteria**	
Number of Instances	N/A	**Update Cadence for Online Data**	N/A	**Focal spot size**	N/A
Number of Fields	2 (RGB image, line mask)	**Sampling methods**	Image frames are sampled from source videos at 6fps.	**Cooling method**	N/A
Labeled Classes	2 (line, background)	**Validation methods**	Manual	**Avg Adult Effective Dose (mSv)**	N/A
Number of Labels	1	**Processing methods**	N/A	**Operational voltage range**	N/A

Average labels per instance	1	**Annotation methods**	Synthetic data annotated automatically through render pipeline. Real data annotated manually.	
Missing Labels	0			

DATA PRE-PROCESSING	DEMOGRAPHIC GROUPS	EVALUATION DATA
None	Data does not contain labeled groups or demographic attributes.	Evaluation data is a split of the training dataset.

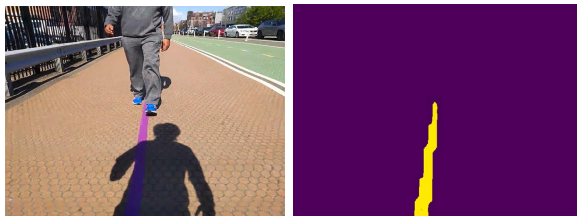
Evaluation Results

Aggregate Evaluation Results

EVALUATION PROCESS	EVALUATION RESULTS
Metrics: mean intersection-over-union (mIoU) **Evaluation Set:** The evaluation set consists of ~40% of the dataset (real) withheld from training. **Process:** The fine-tuned model is run on the evaluation set and mIoU is computed by comparing the class prediction of each pixel between the output mask and the ground truth mask. 	93% mIoU across eval dataset (real images)

Model Usage & Limitations

SENSITIVE USE	LIMITATIONS	ETHICAL CONSIDERATIONS & RISKS
No sensitive deployment cases identified.	**Input conditions:** The model is tuned for a variety of outdoor scenarios including parks, streets, and trails with daytimelighting and no inclement weather. The purple line is expected to be of approximate reference color RGB [144, 99, 205] with a non-reflective finish, applied to a typical road surface (concrete, asphalt). The performance may be impacted when deviating from these conditions. **Output Caveats:** The output mask may not accurately represent the line even if a purple line exists in the input image. There may be extraneous segments due to similar colors in the image, or missing segments due to occlusions, reflections, or lighting conditions.  Similar colored shirt picked up in output mask. (mIoU=65%)	• The model is beneficial for human accessibility applications, however is not appropriate as a safety-critical component.



Output mask includes shadowed segment of line, but not occluded. (mIoU=91%)