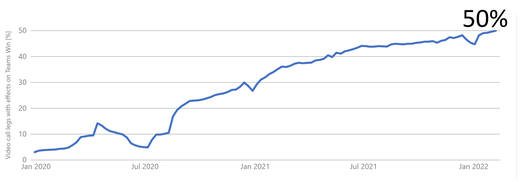
Person Segmentation in Teams

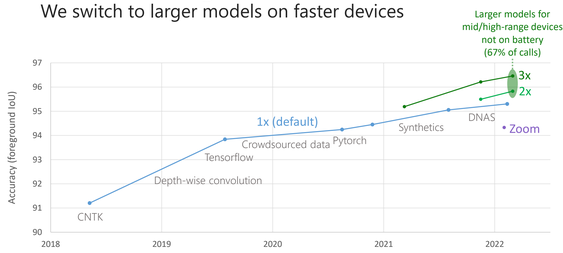
Usage

50% of all video calls in Teams Win use background effects



Accuracy

We switch to larger models on faster devices



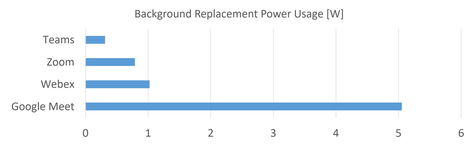
Processing time

Better models, inference, and HW have halved the processing time

Telemetry from all platforms. Including segmentation and blur/replacement processing.

Power consumption

Industry-leading power consumption of background replacement



Pre-join screen. 1x model. Plugged in. October 2021.
Surface Pro 7, i5-1135G7, 2.4 GHz, 16 GB, 30W limit

Microsoft

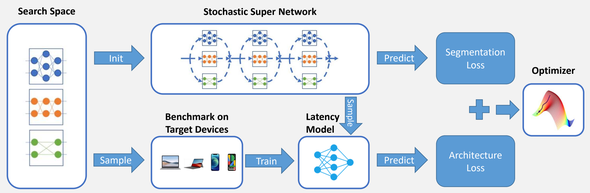
Latency Constrained DNAS

Automatically search for top-performing model architectures that run within a fixed time limit on a specific device.

Motivation

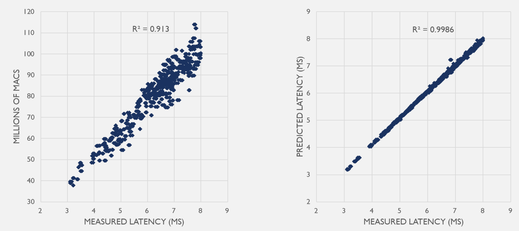
- Scalable approach of generating platform-optimized architectures.
- Fast search in single training without the need for target device in the loop.
- Avoid conditioning architecture search on hardware agnostic metrics (MACs, FLOPs) as they are poor performance proxies.

Method



[1812.03443] FBNet: Hardware-Aware Efficient ConvNet Design via Differentiable Neural Architecture Search (arxiv.org)

Results



Model	IoU	Dark frame IoU	Crowd IoU	Latency (ms)	Size (KB)
1x	95.3%	95.0%	94.5%	**1.99**	**246.5**
1x + DNAS	**95.6%**	**96.0%**	**95.0%**	2.05	420.0

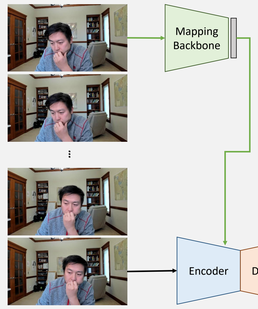
Mapping Model

Update model parameters on-the-fly to adapt to the scene.

Motivation

- Leverage temporal redundancy in videos to do segmentation better/faster.
- Split additional model capacity across spatial and temporal dimensions.

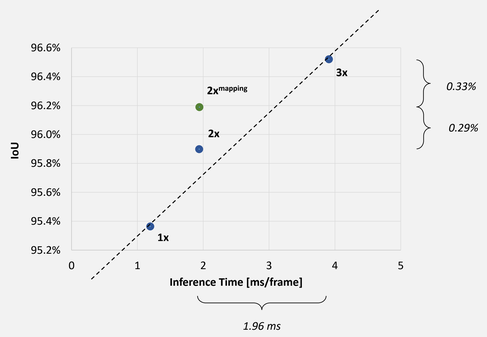
Method



Mapping Model
Latency: 1.2 ms
Runs: every 300th frame

Segmentation Model
Latency: 1.9 ms
Runs: every frame

Results



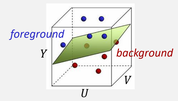
Deep Guided Filtering

Real-time segmentation at full resolution

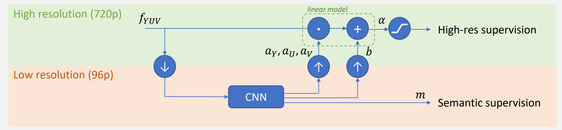
Motivation

- Inference needs to be in low resolution to keep computation low.
- In a local neighborhood, the foreground/background decision can be made by a simple linear color classifier.
- Let the low-res inference guide the lightweight high-res processing.

Method



$$\alpha = Ya_Y + Ua_U + Va_Y + b$$



Results

	No guided filtering	Classical guided filtering	Deep guided filtering
Error (MSE)	17.1	15.4	**11.8**
Compute (MAC)	75M	90M	90M
	In Teams on low-end devices	*In Teams*	*Coming soon!*

Media Intelligence Group, IC3

Varun Jain, Quan Zou, Louis Florentin, Albert Sadovnikov, Henrik Turbell