tinyML® Research Symposium

Enabling Ultra-low Power Machine Learning at the Edge

March 27, 2023

www.tinyML.org

McGill

RSSL
RELIABLE SILICON SYSTEMS LAB

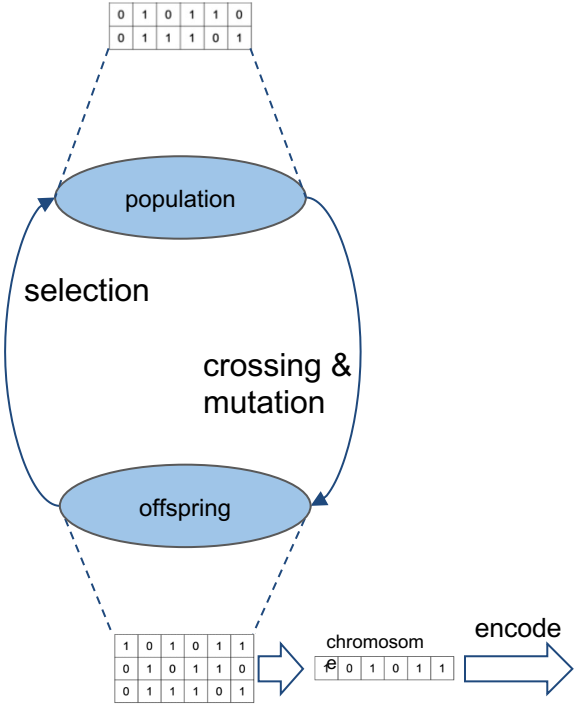
FMAS: Fast Multi-Objective SuperNet Architecture Search for Semantic Segmentation

Zhuoran Xiong, Marihan Amein, Olivier Therrien,
Warren J. Gross, Brett H. Meyer
Electrical and Computer Engineering
McGill University, Montreal
{zhuoran.xiong,marihan.amein,olivier.therrien}@mail.mcgill.ca
{warren.gross,brett.meyer}@mcgill.ca

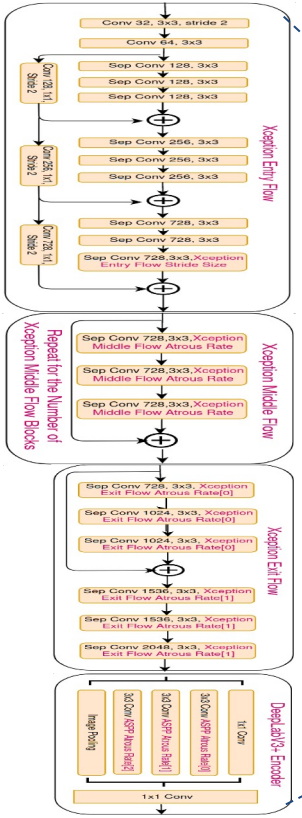
3/27/2023

Overview

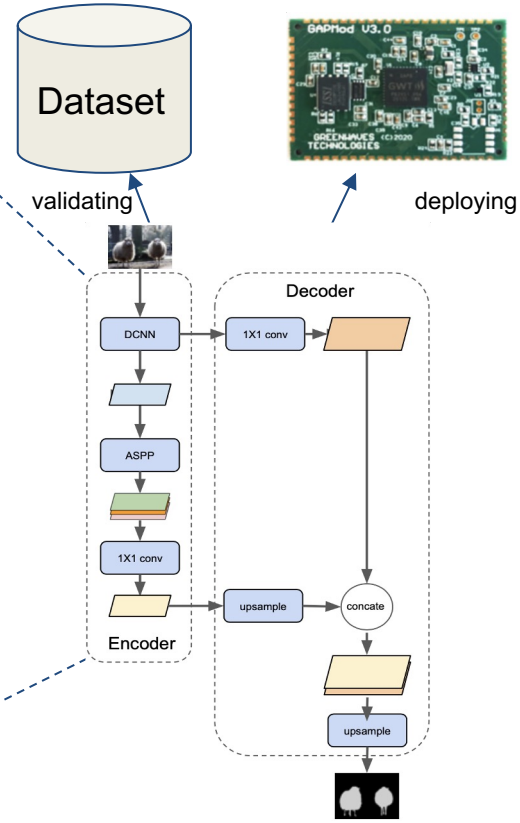
Genetic Algorithm



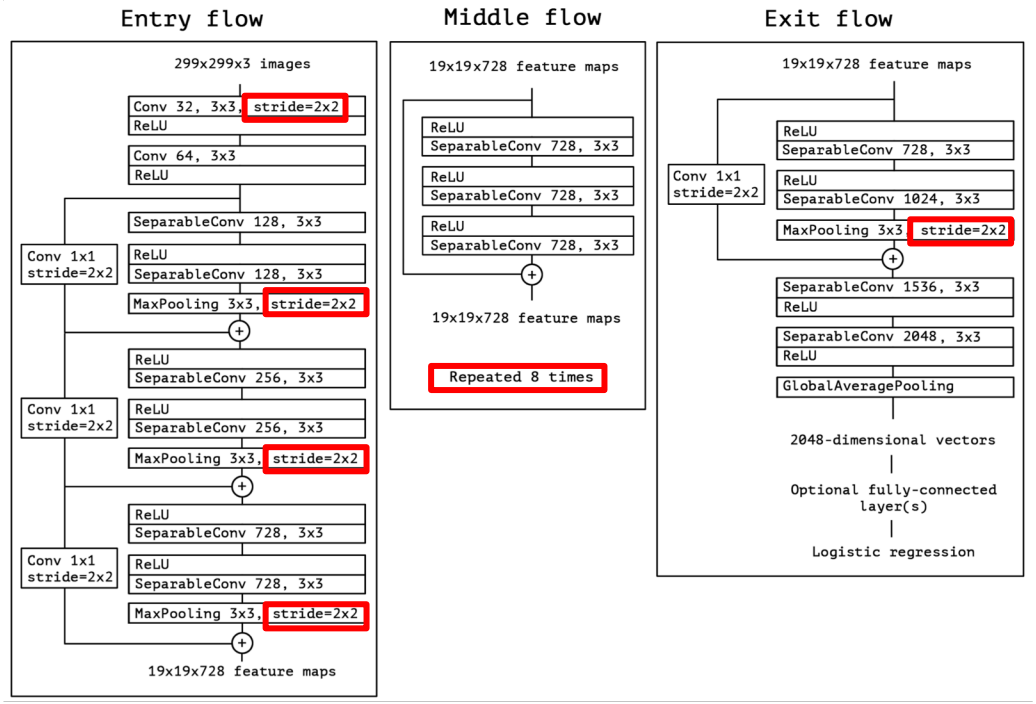
Xception Backbone



Sampled DeeplabV3+



Methodology



Xception Backbone

Table 1: Xception Hyperparameter Design Space

Hyperparameter	Possible Values
Xception Entry Flow Stride	1, 2, 3, 4
Xception Middle Flow Atrous Rate	1, 2, 3, 4
Xception Exit Flow Atrous Rates	(1, 2), (2, 4)
ASPP Atrous Rates	(6, 12, 18), (12, 24, 36)
Xception Middle Flow Blocks	$(b_1, b_2, \ldots, b_{16}), b_i \in \{0, 1\}$

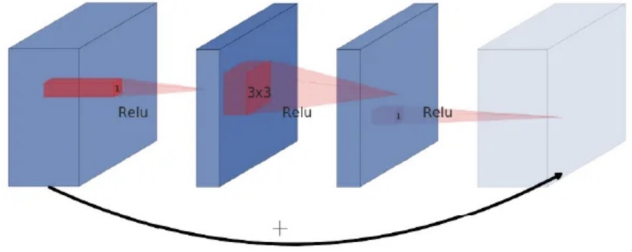
- DeepLabV3+ as a supernet
- Two backbones: Xception & MobileNetV2
- sub-sampling and reusing its pre-trained weights

Methodology

MobileNetV2 Backbone

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Bottleneck with expansion layer



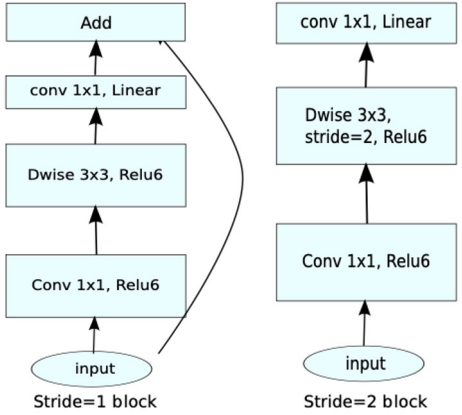
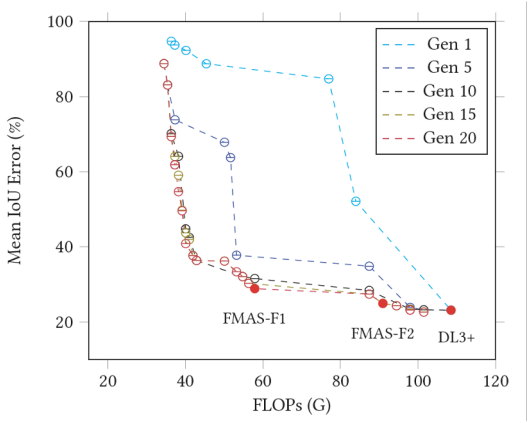


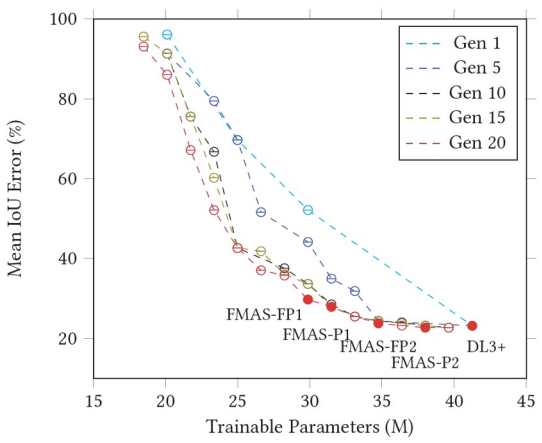
Table 2: MobileNetV2 Hyperparameters Design Space

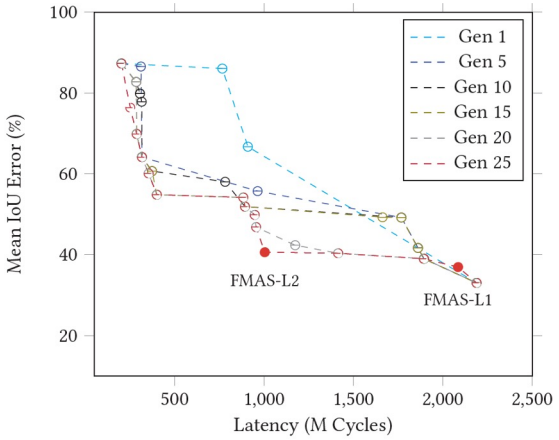
Hyperparameter	Possible Values
2^{nd} & 3^{rd} Layer Stride	2, 3
14^{th} & 17^{th} Layer Stride	1, 2
12^{th}-14^{th} Layer Dilation Rate	1, 2
15^{th}-17^{th} Layer Dilation Rate	1, 2, 3, 4
24-channel Group Layers	$(b_1, b_2), b_i \in \{0, 1\}, \sum b_i > 0$
32-channel Group Layers	$(b_1, b_2, b_3), b_i \in \{0, 1\}, \sum b_i > 0$
64-channel Group Layers	$(b_1, b_2, b_3, b_4), b_i \in \{0, 1\}, \sum b_i > 0$
96-channel Group Layers	$(b_1, b_2, b_3), b_i \in \{0, 1\}, \sum b_i > 0$
160-channel Group Layers	$(b_1, b_2, b_3), b_i \in \{0, 1\}, \sum b_i > 0$

Experiments and results

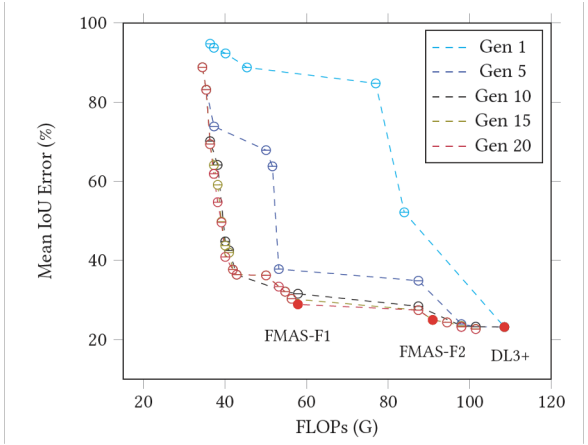
- Search by changing parameters of supernet
- evaluate on a subset of the validation dataset (20%)







Experiments and results



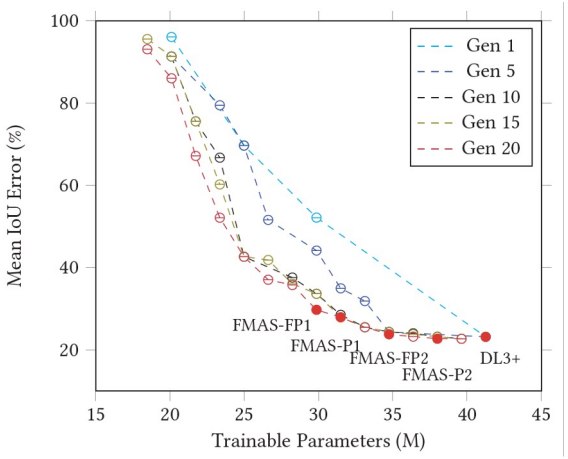


Table 3: Cost and performance with Modified Xception backbone and derived models

	Xception Architecture Parameters					Cost			MIoU Error (%)	
	Entry Stride	Middle Atrous Rate	Exit Atrous Rate	ASPP Atrous Rate	Middle Blocks	GPU Days	FLOPs (G)	Params (M)	Validation Subset (%)	Fine-tuned + Full Validation
DeepLabV3+ [11]	2	1	(1,2)	(6,12,18)	1111111111111111	-	101.47	41.26	23.14	22.71
DPC [8]	-	-	-	-	-	2600	99.96	42.70	-	**19.15**
FMAS-F1	3	1	(1,2)	(6,12,18)	1111111011011111	0.68	**57.88**	38.00	28.88	27.93
FMAS-P1	2	1	(1,2)	(6,12,18)	1111011110010100	**0.49**	87.41	31.5	27.91	26.64

Experiments and results

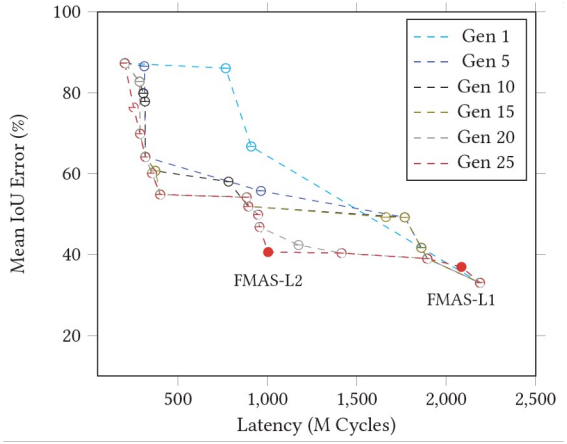


Table 4: Cost and accuracy with MobileNetV2 backbone and derived models

	MobileNetV2 Architecture Paremeters			Cost				MIoU Error (%)	
	Stride	Inverted Layers Dilation Rate	Inverted Group Layers	GPU Days	FLOPs (G)	Params (M)	Latency (M Cycles)	Validation Subset	Fine-tuned + Full Validation
MobileNetV2 [28]	(2,2,1,1)	(2,2,2,4,4,4)	1111111111	-	9.73	2.14	2189	33.03	**32.61**
FCN-VGG16 [19]	-	-	-	-	243.50	134.49	-	-	37.70
FMAS-L1	(2,2,1,2)	(2,2,1,3,4,2)	1111111111	1.46	7.88	2.14	2085	36.94	36.26
FMAS-L2	(2,3,1,1)	(2,2,2,3,2,2)	1111111111	1.46	**4.62**	2.14	**1004**	40.61	40.22

Experiments and results

Table 3: Cost and performance with Modified Xception backbone and derived models

	Xception Architecture Parameters					Cost			MIoU Error (%)	
	Entry Stride	Middle Atrous Rate	Exit Atrous Rate	ASPP Atrous Rate	Middle Blocks	GPU Days	FLOPs (G)	Params (M)	Validation Subset (%)	Fine-tuned + Full Validation
DeepLabV3+ [11]	2	1	(1,2)	(6,12,18)	1111111111111111	-	101.47	41.26	23.14	22.71
DPC [8]	-	-	-	-	-	2600	99.96	42.70	-	**19.15**
FMAS-F1	3	1	(1,2)	(6,12,18)	1111111011011111	0.68	**57.88**	38.00	28.88	27.93
FMAS-F2	2	1	(1,2)	(6,12,18)	1111111011001001	0.52	90.92	33.12	24.95	25.21
FMAS-P1	2	1	(1,2)	(6,12,18)	1111011110010100	**0.49**	87.41	31.5	27.91	26.64
FMAS-P2	2	1	(1,2)	(6,12,18)	1111111011011111	0.65	101.47	38.00	22.68	22.65
FMAS-FP1	2	1	(1,2)	(6,12,18)	1111111011001101	0.68	94.44	34.75	23.77	24.38
FMAS-FP2	2	1	(1,2)	(6,12,18)	1111011010001100	0.80	83.89	**29.87**	29.72	29.29

Thank you!
Questions ?

Copyright Notice

www.tinyml.org