



# Improving AutoTVM Efficiency by Schedule Sharing

AWS AI

Cody Yu

AWS Deep Engine Science



# Motivation

- Tuning MobileNet V2 1.0 on Nvidia V100 for 500 trails

- 31 tasks, including
- conv2d
  - depthwise conv2d

[Task 1/31]	Current/Best:	946.31/1289.20	GFL0PS	Progress: (500/500)	2104.64 s	Done.
[Task 2/31]	Current/Best:	668.17/1049.02	GFL0PS	Progress: (500/500)	1968.25 s	Done.
[Task 3/31]	Current/Best:	783.52/ 897.95	GFL0PS	Progress: (500/500)	1951.28 s	Done.
[Task 4/31]	Current/Best:	1535.61/2011.77	GFL0PS	Progress: (500/500)	2130.80 s	Done.
[Task 5/31]	Current/Best:	361.62/ 454.76	GFL0PS	Progress: (500/500)	1576.12 s	Done.
[Task 6/31]	Current/Best:	978.99/1541.32	GFL0PS	Progress: (500/500)	2365.66 s	Done.
[Task 7/31]	Current/Best:	2904.66/3478.22	GFL0PS	Progress: (500/500)	2923.23 s	Done.
[Task 8/31]	Current/Best:	1570.40/1935.42	GFL0PS	Progress: (500/500)	2422.44 s	Done.
[Task 9/31]	Current/Best:	2650.49/2888.28	GFL0PS	Progress: (500/500)	2512.69 s	Done.
[Task 10/31]	Current/Best:	5150.11/5389.75	GFL0PS	Progress: (500/500)	2796.52 s	Done.
[Task 11/31]	Current/Best:	1070.05/1403.03	GFL0PS	Progress: (500/500)	2585.96 s	Done.
[Task 12/31]	Current/Best:	1735.87/2434.35	GFL0PS	Progress: (500/500)	2648.54 s	Done.
[Task 13/31]	Current/Best:	1644.45/2162.78	GFL0PS	Progress: (500/500)	2492.33 s	Done.
[Task 14/31]	Current/Best:	2316.08/2963.59	GFL0PS	Progress: (500/500)	3015.28 s	Done.
[Task 15/31]	Current/Best:	3570.33/4043.73	GFL0PS	Progress: (500/500)	2942.17 s	Done.
[Task 16/31]	Current/Best:	1159.51/1486.77	GFL0PS	Progress: (500/500)	2554.60 s	Done.
[Task 17/31]	Current/Best:	937.18/1152.14	GFL0PS	Progress: (500/500)	2131.96 s	Done.
[Task 18/31]	Current/Best:	1206.02/1675.35	GFL0PS	Progress: (500/500)	2406.45 s	Done.
[Task 19/31]	Current/Best:	2167.02/2302.35	GFL0PS	Progress: (500/500)	2560.77 s	Done.
[Task 20/31]	Current/Best:	24.34/ 541.03	GFL0PS	Progress: (500/500)	1654.92 s	Done.
[Task 21/31]	Current/Best:	1035.45/1338.88	GFL0PS	Progress: (500/500)	2302.91 s	Done.
[Task 22/31]	Current/Best:	1054.80/2099.63	GFL0PS	Progress: (500/500)	2249.21 s	Done.
[Task 23/31]	Current/Best:	618.47/3150.01	GFL0PS	Progress: (500/500)	1863.21 s	Done.
[Task 24/31]	Current/Best:	627.86/1551.72	GFL0PS	Progress: (500/500)	2481.23 s	Done.
[Task 25/31]	Current/Best:	1000.60/1831.31	GFL0PS	Progress: (500/500)	2704.17 s	Done.
[Task 26/31]	Current/Best:	2626.91/2953.95	GFL0PS	Progress: (500/500)	2626.47 s	Done.
[Task 27/31]	Current/Best:	284.45/ 433.29	GFL0PS	Progress: (500/500)	1432.13 s	Done.
[Task 28/31]	Current/Best:	599.56/ 924.07	GFL0PS	Progress: (500/500)	1939.39 s	Done.
[Task 29/31]	Current/Best:	672.51/ 951.80	GFL0PS	Progress: (500/500)	2099.23 s	Done.
[Task 30/31]	Current/Best:	2113.59/3039.97	GFL0PS	Progress: (500/500)	2021.68 s	Done.
[Task 31/31]	Current/Best:	19.49/ 97.05	GFL0PS	Progress: (500/500)	1608.51 s	Done.

One task took about 40 mins  
→ ~19 hours in total

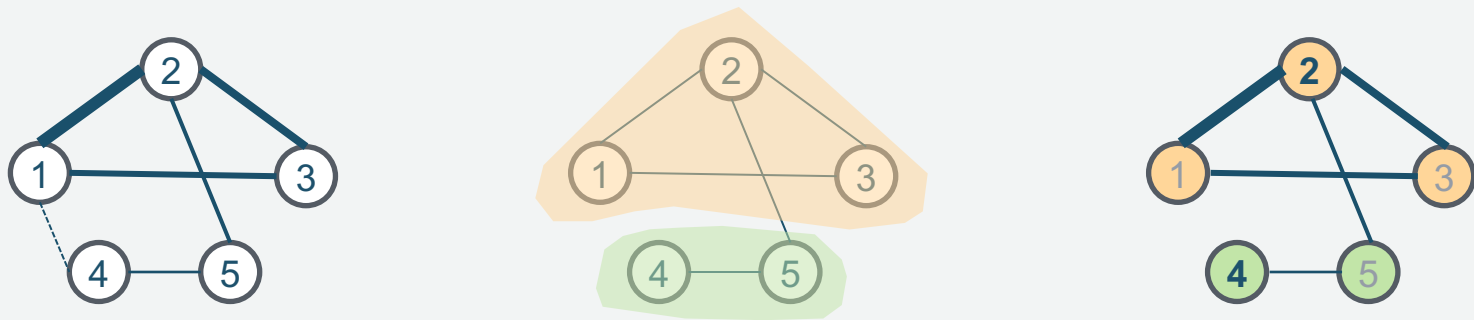


# Motivation

- Observations
  - AutoTVM tunes every task in a model from scratch
  - The high quality schedules of some tasks in a model are similar
- Proposed Solution
  - *Only tune representative tasks and share their best schedules to others*

# Schedule Sharing among AutoTVM Tasks

- Idea
  - Two tasks can share schedules if their tuning spaces are similar
- Approach
  - Cluster tasks with a distance metric
  - Identify a centric (representative) task in each cluster and only tune them
  - Share their best schedules to other tasks in the same cluster



- Node: Tasks
- Edge: Distance metric

Clustering

Determine the  
representative tasks

# Evaluation Results and RFC

- Selected 7 models from Gluon CV model zoo
- Tuned selected tasks (~5-7) for 3,000 trials on Nvidia V100
- Achieve on average 84% performance with only 28% tuning time

Model	Tuning Time w/o Sharing (mins)	Perf. w/o Sharing (ms)	Tuning Time w. Sharing (mins)	Perf. w. Sharing (ms)	Used Time	Achieve Perf.
MobileNet V2 1.0	1185	0.74	404	0.78	34%	95%
ResNet 50 V1	1666	2.27	358	3.7	21%	61%
VGG 19 BN	479	5.08	169	6.36	35%	80%
SqueezeNet 1.1	574	0.54	167	0.5	29%	108%
DenseNet 121	2670	2.99	377	3.02	14%	99%
Yolo3 MobileNet1.0 voc	2784	5.4	774	7.16	28%	75%
SSD512 ResNet50 V1 voc	3426	8.47	1150	5.65	34%	67%
Average					28%	84%



↑ RFC ↑

Welcome to Participate!

