Self-improving Learners

Min Sun

National Tsing Hua University

@2nd AII Workshop

Challenges of Modern AI

- Large-scale labelled dataset

Is 'data labeling' the new blue-collar job of the AI era?

Automation has put low-skill jobs at risk for decades. And self-driving cars, robots, and speech recognition will continue the trend. But, some experts also see new opportunities in the automated age.

Challenges of Modern AI

- Large-scale labelled dataset

Is 'data labeling' the new blue-collar job of the AI era?

Automation has put low-skill jobs at risk for decades. And self-driving cars, robots, and speech recognition will continue the trend. But, some experts also see new opportunities in the automated age.

- Talent Intensive Workforce

AI獨角獸商湯科技C輪金主浮出，高通領投目標5億美元｜數位時代
https://www.bnext.com.tw/.../qualcomm-invests-in-chinese-ai-facial... ▾ Translate this page
Nov 17, 2017 - 吳育瑞說，商湯科技現有900人中，有120位博士，專長領域包括人臉識別、圖像識別、動態影像分析等，服務客戶包括手機品牌小米、OPPO和華為， ...

Grad student descent | Science Dryad
https://sciencedryad.wordpress.com/2014/01/25/grad-student-descent/ ▾
Jan 24, 2014 - One method, common in academia, is '**grad student descent**' (a pun on gradient descent), in which a graduate student fiddles around with the ...

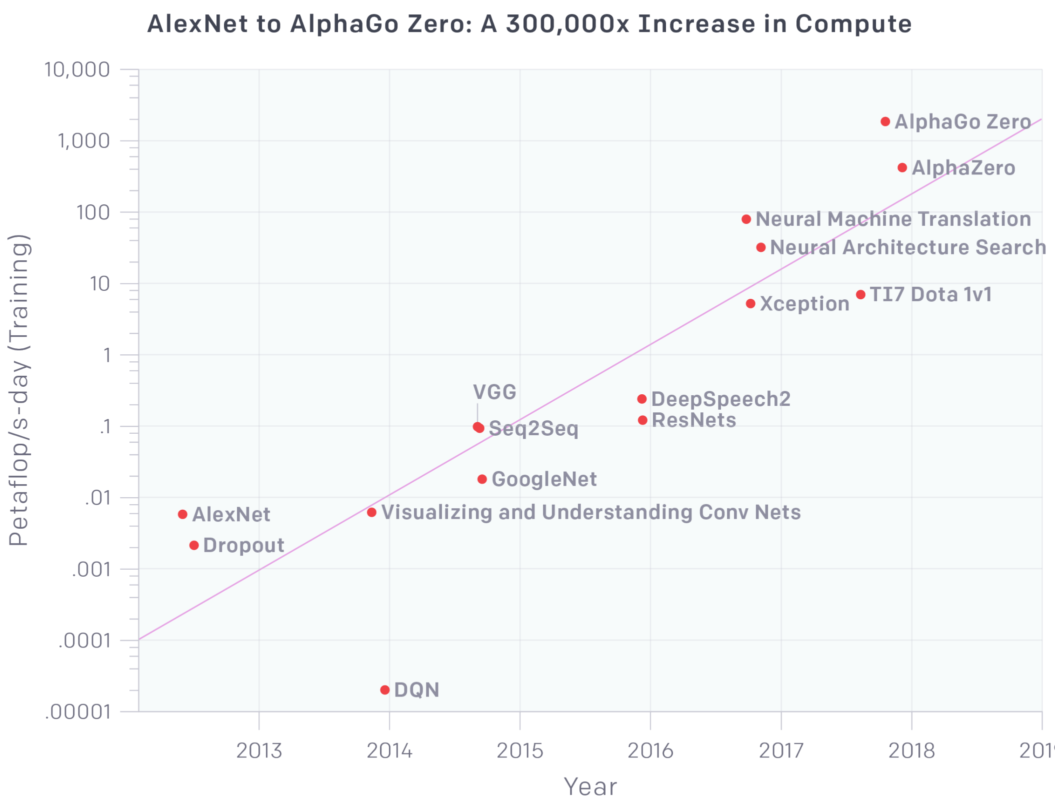
Weapons to Tackle the Challenges

- Sensory data from realistic user scenarios

Weapons to Tackle the Challenges

- Sensory data from realistic user scenarios

- Exponential trends in computing



Outline

- Self-Supervised Learning of Depth from 360° Videos (Sensory, Pitch)

- DPP-Net: Device-aware Progressive Search for Pareto-optimal Neural Architectures (Compute)

VSLab

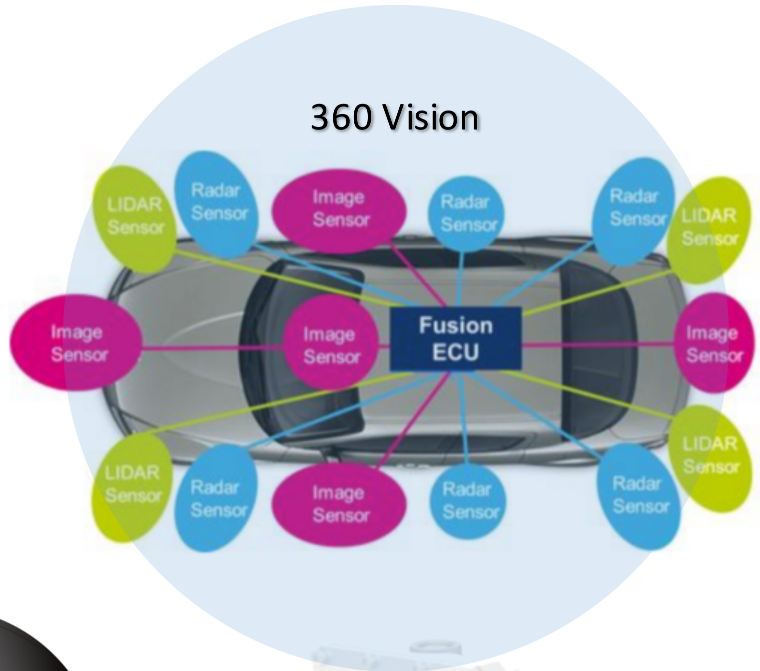
Self-Supervised Learning of Depth from 360° Videos

Min Sun

National Tsing Hua University

Under Submission

Our Goal



1. Well-Calibrated
2. Low-Cost
3. High-Resolution
4. Large FoV

$360°$ Camera

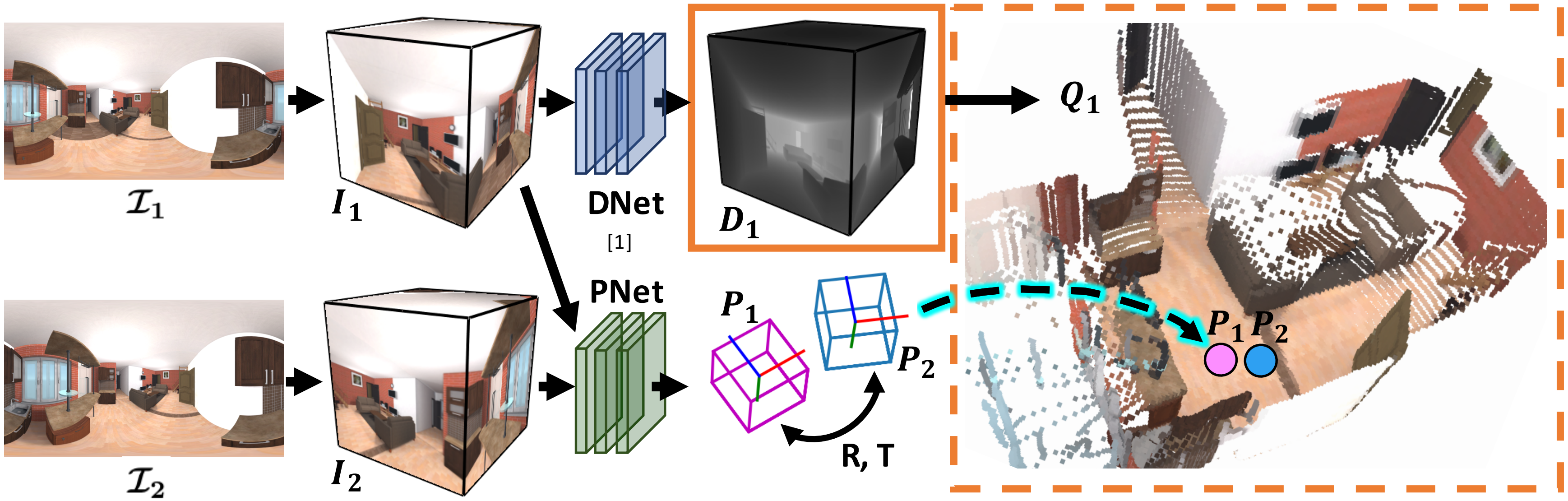
Radar

LIDAR

Image credits: https://hackernoon.com/mit-6-s094-deep-learning-for-self-driving-cars-2018-lecture-2-notes-e283b9ec10a0

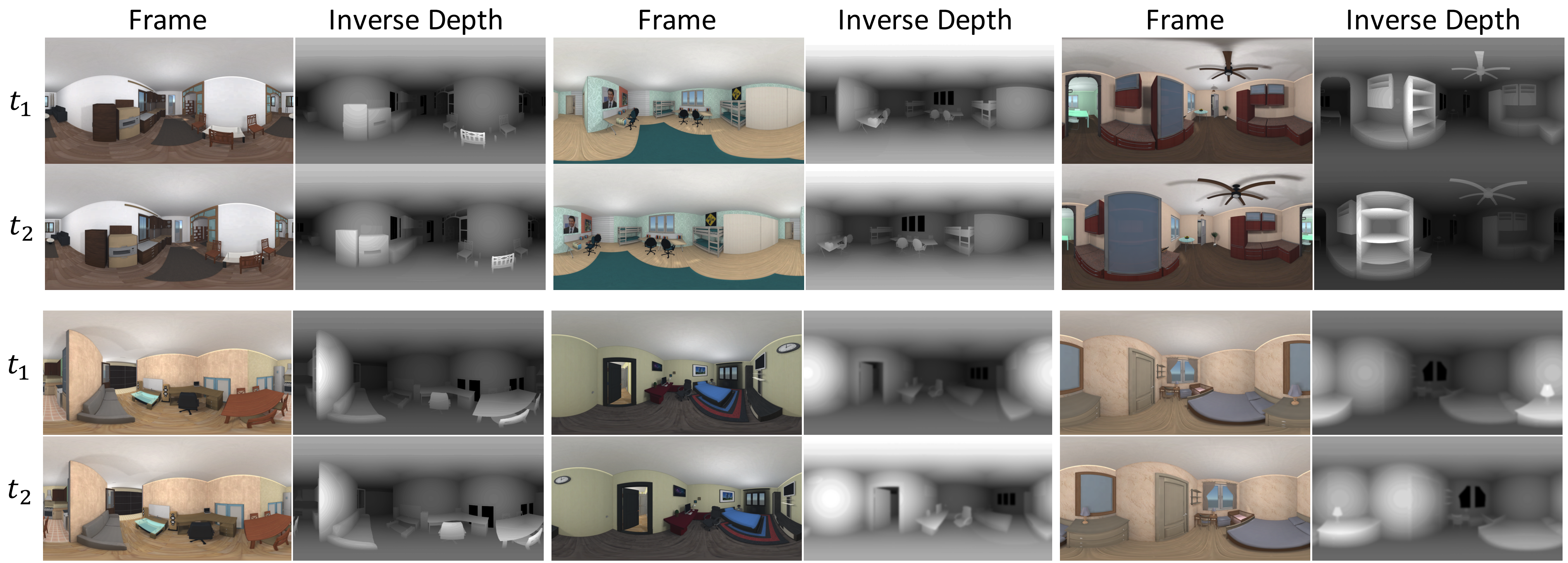
Our Model

I: Equirectangular
I: Cube
D: Depth
P: Camera motion
Q: Point Cloud



[1] Zhou et al., Unsupervised Learning of Depth and Ego-Motion from Video, CVPR 2017

Dataset – PanoSUNCG

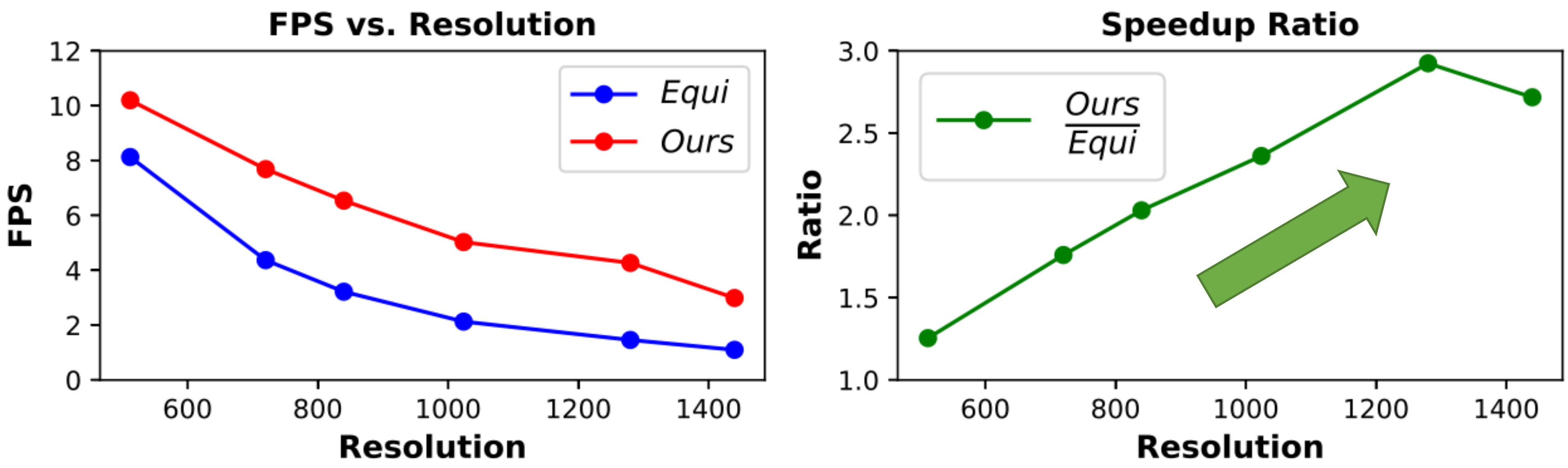


Quantitative Results – Depth

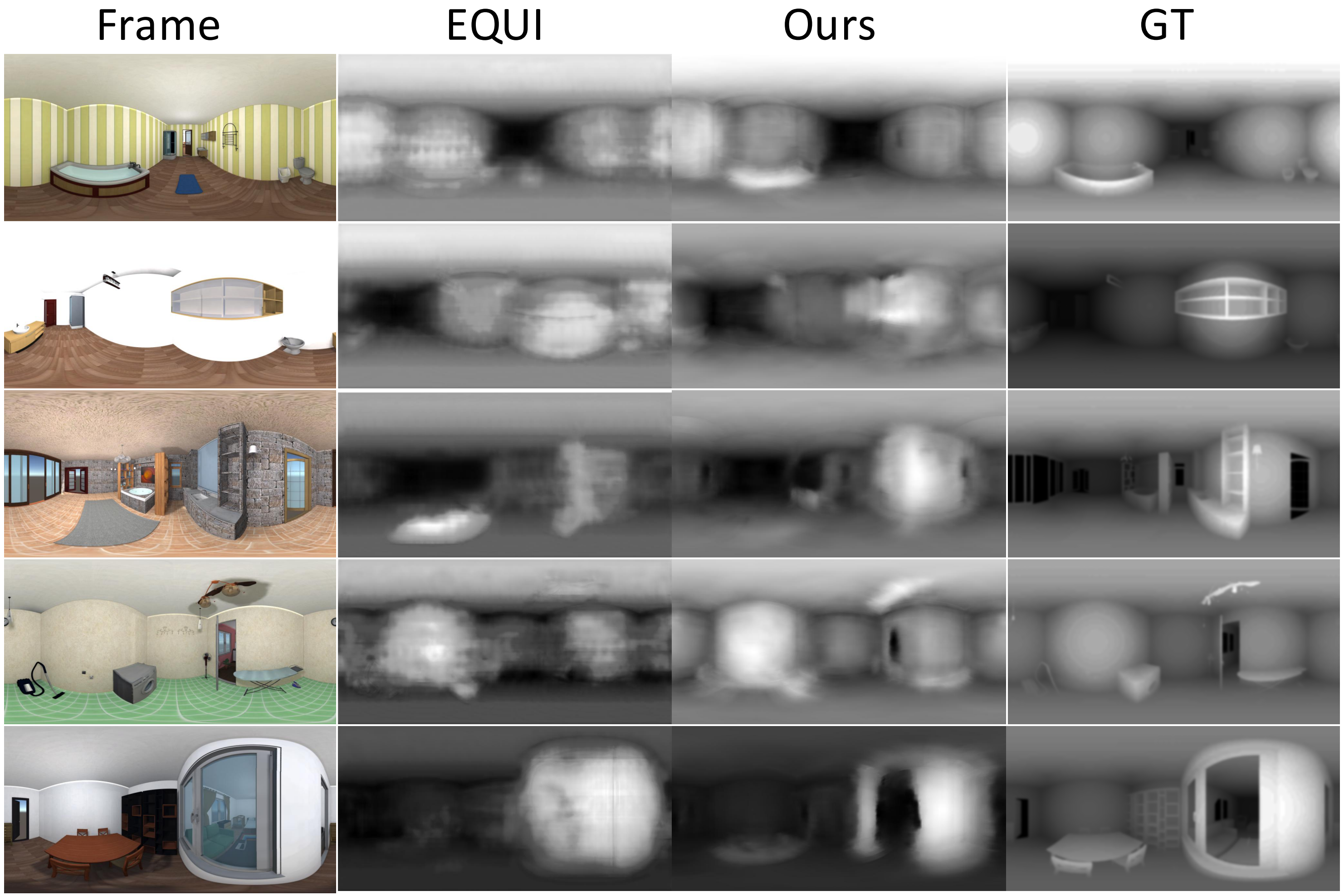
Method	Abs Rel	Sq Rel	RMSE	$RMSE_{log}$	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Ours (Full Model)	**0.337**	**4.986**	**8.589**	0.611	**0.647**	**0.829**	**0.899**
Ours w/o $\mathcal{L}_{pose}$	0.418	7.113	9.916	0.698	0.580	0.790	0.876
single pose	0.462	7.733	10.431	0.665	0.526	0.752	0.848

$$\mathcal{L}_{pose} = \sqrt{\frac{\sum_{i \in f} (P^{i'}_{t,t+1} - P^{*}_{t,t+1})^2}{6}}$$

Efficiency – Speedup Ratio

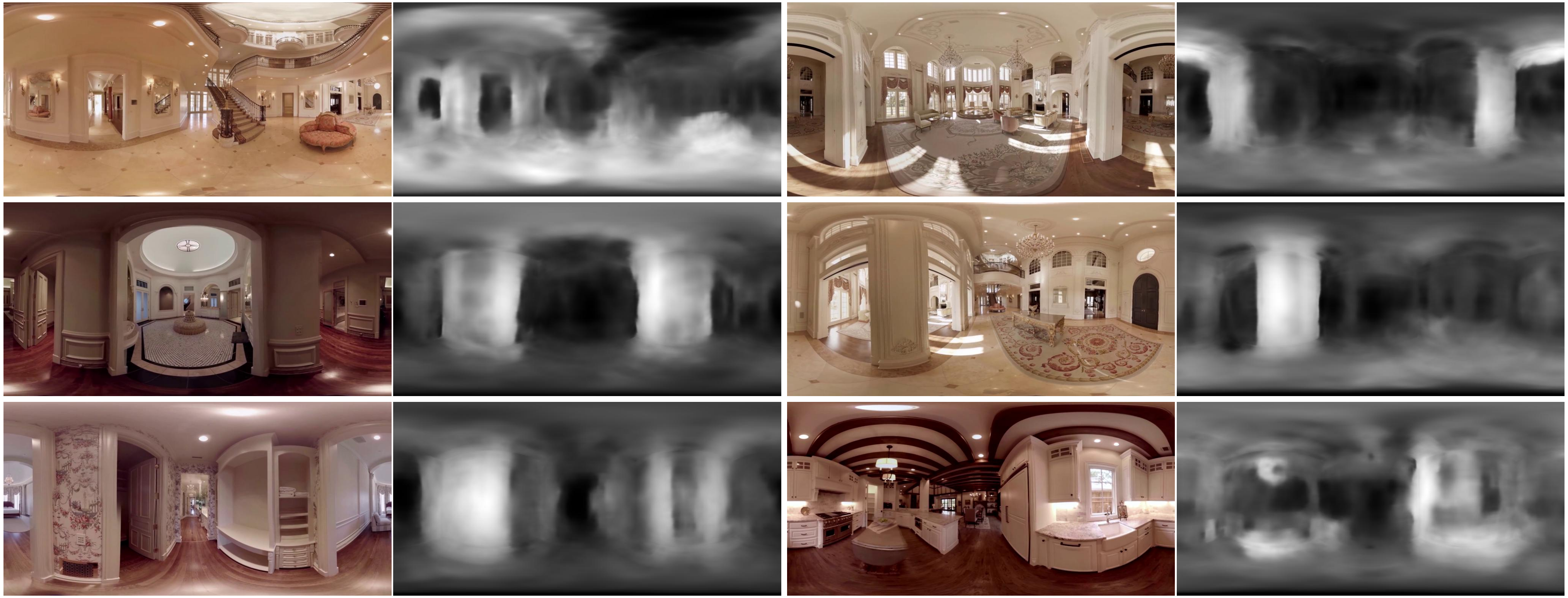


Qualitative Results – PanoSUNCG



Qualitative Results – Real-world Videos

Frame | Our prediction | Frame | Our prediction



DPP-Net: Device-aware Progressive Search for Pareto-optimal Neural Architectures

Jin-Dong (Mark) Dong[1], An-Chieh Cheng[1], Da-Cheng Juan[2], Wei Wei[2], Min Sun[1]
National Tsing-Hua University[1]
Google[2]

ICLR Workshop 2018

https://markdtw.github.io/pppnet.html

Slides by Mark : markdtw

Hot Trend - Neural Architecture Search

- Barret Zoph, et al. “Neural Architecture Search with Reinforcement Learning”, In ICLR 2017

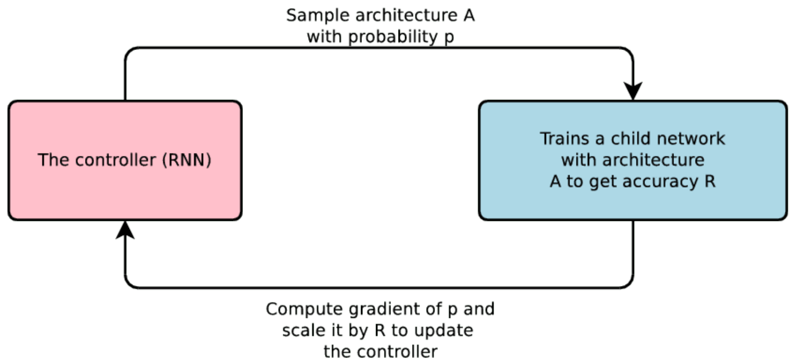
    NAS used **800 GPUs for 28days**

- Irwan Bello, et al. “Neural Optimizer Search with Reinforcement Learning”, In ICML 2017

    NASNet used **450 GPUs for 3-4 days** (i.e. 32,400-43,200 GPU hours)

- Hieu Pham, et al. “Efficient Neural Architecture Search via Parameter Sharing”, In ArXiv 2018

    ENAS used **1 GTX1080Ti for 10 hours**



What’s Missing

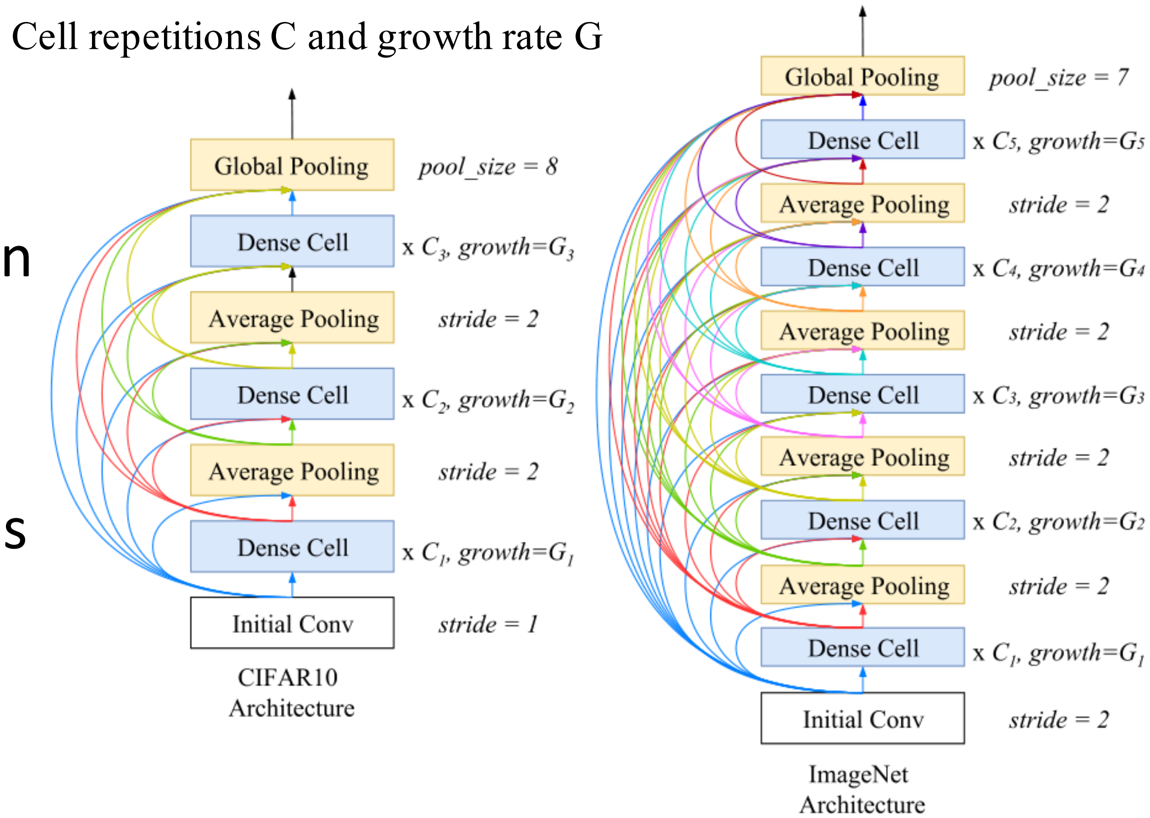
- Current works mostly focus on achieving **high classification accuracy** regardless of other factors.
  *single objective -> **multi-objectives*** (accuracy, inference time, etc)

- Demands for ubiquitous model inference is rising. However, designing suitable NNs for **all devices** (HPC, cloud, embedded system, mobile phone, etc.) remain challenging.

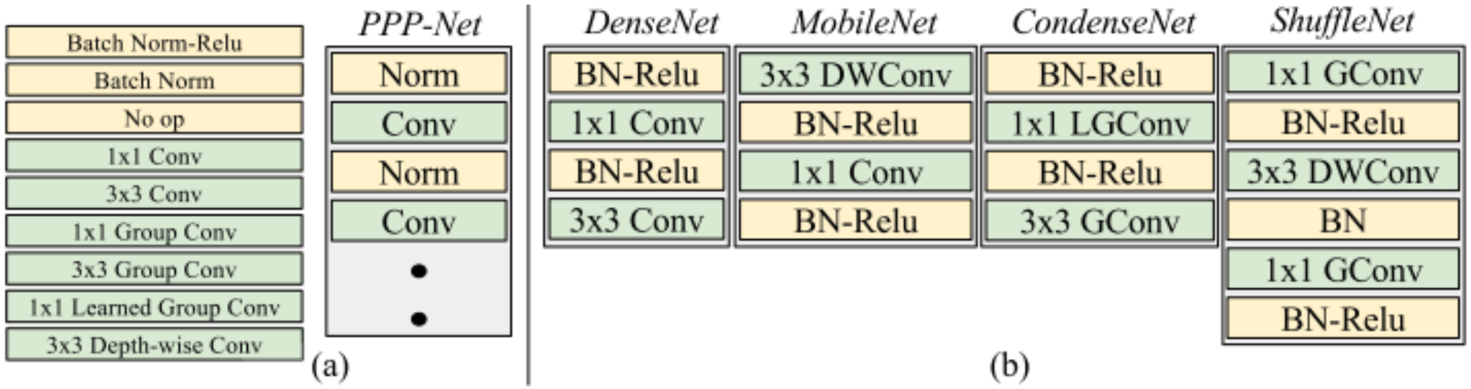
- Therefore, we aim at **automatically** design such models for **different devices** considering **multiple objectives**.

Our Approach: Search Space

- Cells are connected following *CondenseNet by Huang et al.*
  (1) layers with different resolution are also directly connected.
  (2) growth rate G **doubles** when the feature map shrinks.
- This connection scheme improves the computational efficiency.



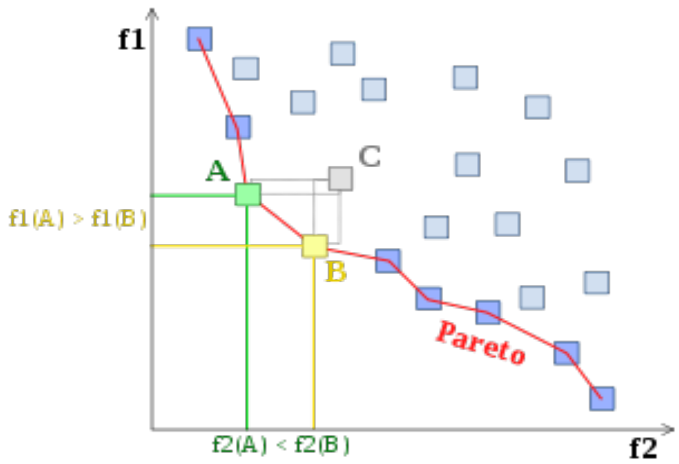
Our Approach: Search Space

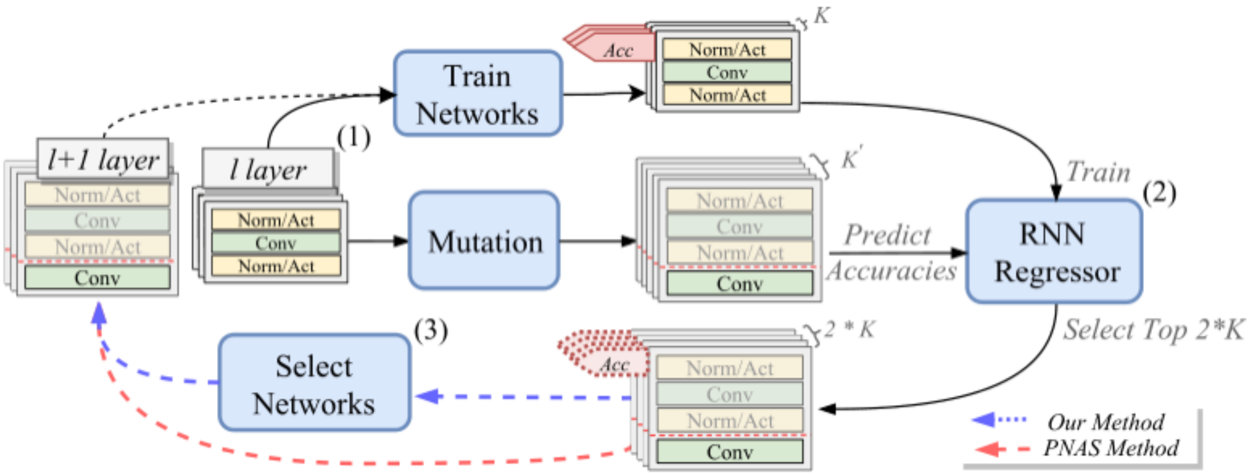


- Designed a new **cell** search space that covers famous **compact** CNNs.
- Search for a **cell** instead of a whole architecture.

Our Approach: Search Algorithm

- Sequential Model-based Optimization.
  - Sequential: **Progressively** add layers.
  - Model-based: RNN Regressor -> predict accuracy.
- Select K Networks: **Pareto** Optimality





Experiment Settings

- Test DPP-Net on 3 different **devices**.

	Workstation (WS)	**Embedded System (ES)**	**Mobile Phone (Mobile)**
Instance	Desktop PC	NVIDIA Jetson TX1	Xiaomi Redmi Note 4
CPU	Intel i5-7600	ARM Cortex-A57	ARM Cortex-A53
Cores	4	4	8
GHz	3.5	1.9	2.0
CUDA	Titan X Pascal	Maxwell 256	-
Memory	64 GB / 12 GB	4 GB	3 GB
Objectives	4	4	5

- Train on CIFAR-10.

CIFAR-10 Experiment

	Device-agnostic metrics			*Device-aware metrics*			
Model from previous works	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
Real et al. [11]	5.4	5.4M	-	-	-	-	-
NASNet-B [9]	3.73	2.6M	-	-	-	-	-
PNASNet-1 [15]	4.01	1.6M	-	-	-	-	-
DenseNet-BC (k=12) [31]	4.51	0.80M	-	-	-	0.273	79MB
CondenseNet-86 [18]	5.0	0.52M	65.8M	0.009	0.090	0.149	113MB
	Device-agnostic metrics			*Device-aware metrics*			
Model from DPP-Net	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
DPP-Net-PNAS	**4.36**	11.39M	1364M	0.013	0.062	0.912	213MB
DPP-Net-WS-A	4.78	1.00M	137M	**0.006**	0.075	0.210	129MB
DPP-Net-ES-A	4.93	2.04M	270M	0.007	**0.044**	0.381	100MB
DPP-Net-Mobile-A	5.84	**0.45M**	**59.27M**	0.008	0.065	**0.145**	**58MB**
DPP-Net-Panacea	4.58	0.52M	63.5M	0.008	0.083	0.149	104MB

CIFAR-10 Experiment

	Device-agnostic metrics			*Device-aware metrics*			
Model from previous works	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
Real et al. [11]	5.4	5.4M	-	-	-	-	-
NASNet-B [9]	3.73	2.6M	-	-	-	-	-
PNASNet-1 [15]	4.01	1.6M	-	-	-	-	-
DenseNet-BC (k=12) [31]	4.51	0.80M	-	-	-	0.273	79MB
CondenseNet-86 [18]	5.0	0.52M	65.8M	0.009	0.090	0.149	113MB
	Device-agnostic metrics			*Device-aware metrics*			
Model from DPP-Net	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
DPP-Net-PNAS	**4.36**	11.39M	1364M	0.013	0.062	0.912	213MB
DPP-Net-WS-A	4.78	1.00M	137M	**0.006**	0.075	0.210	129MB
DPP-Net-ES-A	4.93	2.04M	270M	0.007	**0.044**	0.381	100MB
DPP-Net-Mobile-A	5.84	**0.45M**	**59.27M**	0.008	0.065	**0.145**	**58MB**
DPP-Net-Panacea	4.58	0.52M	63.5M	0.008	0.083	0.149	104MB

- DPP-Net-PNAS selects the model with highest accuracy.

CIFAR-10 Experiment

	Device-agnostic metrics			*Device-aware metrics*			
Model from previous works	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
Real et al. [11]	5.4	5.4M	-	-	-	-	-
NASNet-B [9]	3.73	2.6M	-	-	-	-	-
PNASNet-1 [15]	4.01	1.6M	-	-	-	-	-
DenseNet-BC (k=12) [31]	4.51	0.80M	-	-	-	0.273	79MB
CondenseNet-86 [18]	5.0	0.52M	65.8M	0.009	0.090	0.149	113MB
	Device-agnostic metrics			*Device-aware metrics*			
Model from DPP-Net	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
DPP-Net-PNAS	**4.36**	11.39M	1364M	0.013	0.062	0.912	213MB
DPP-Net-WS-A	4.78	1.00M	137M	**0.006**	0.075	0.210	129MB
DPP-Net-ES-A	4.93	2.04M	270M	0.007	**0.044**	0.381	100MB
DPP-Net-Mobile-A	5.84	**0.45M**	**59.27M**	0.008	0.065	**0.145**	**58MB**
DPP-Net-Panacea	4.58	0.52M	63.5M	0.008	0.083	0.149	104MB

- DPP-Net-PNAS selects the model with highest accuracy.
- DPP-Net-*Device*-A runs the fastest on certain *device*.

CIFAR-10 Experiment

	Device-agnostic metrics			*Device-aware metrics*			
Model from previous works	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
Real et al. [11]	5.4	5.4M	-	-	-	-	-
NASNet-B [9]	3.73	2.6M	-	-	-	-	-
PNASNet-1 [15]	4.01	1.6M	-	-	-	-	-
DenseNet-BC (k=12) [31]	4.51	0.80M	-	-	-	0.273	79MB
CondenseNet-86 [18]	5.0	0.52M	65.8M	0.009	0.090	0.149	113MB
	Device-agnostic metrics			*Device-aware metrics*			
Model from DPP-Net	Error rate	Params	FLOPs	Time-WS	Time-ES	Time-Mobile	Mem-Mobile
DPP-Net-PNAS	**4.36**	11.39M	1364M	0.013	0.062	0.912	213MB
DPP-Net-WS-A	4.78	1.00M	137M	**0.006**	0.075	0.210	129MB
DPP-Net-ES-A	4.93	2.04M	270M	0.007	**0.044**	0.381	100MB
DPP-Net-Mobile-A	5.84	**0.45M**	**59.27M**	0.008	0.065	**0.145**	**58MB**
DPP-Net-Panacea	4.58	0.52M	63.5M	0.008	0.083	0.149	104MB

- DPP-Net-PNAS selects the model with highest accuracy.
- DPP-Net-*Device*-A runs the fastest on certain *device*.
- DPP-Net-Panacea performs relatively good on every objectives.

ImageNet Experiment

Model	Top-1	Top-5	Params	FLOPs	Time-ES	Time-Mobile	Mem
Densenet-121 [31]	25.02	7.71	-	-	0.084	1.611	466MB
Densenet-169 [31]	23.80	6.85	-	-	0.142	1.944	489MB
Densenet-201 [31]	22.58	6.34	-	-	0.168	2.435	528MB
ShuffleNet 1x (g=8)	32.4	-	5.4M	140M	0.051	0.458	243MB
MobileNetV2	28.3	-	1.6M	-	0.032	0.777	270MB
Condensenet-74 (G=4)[18]	26.2	8.30	4.8M	529M	0.072	0.694	238MB
NASNet-A (Mobile)	26.0	8.4	5.3M	564M	0.244	-	-
DPP-Net-PNAS	24.16	7.13	77.16M	9276M	0.218	5.421	708MB
DPP-Net-Panacea	25.98	8.21	4.8M	523M	0.069	0.676	238MB

- DPP-Net-Panacea outperforms CondenseNet in every objectives except number of params and memory usage.

ImageNet Experiment

Model	Top-1	Top-5	Params	FLOPs	Time-ES	Time-Mobile	Mem
Densenet-121 [31]	25.02	7.71	-	-	0.084	1.611	466MB
Densenet-169 [31]	23.80	6.85	-	-	0.142	1.944	489MB
Densenet-201 [31]	22.58	6.34	-	-	0.168	2.435	528MB
ShuffleNet 1x (g=8)	32.4	-	5.4M	140M	0.051	0.458	243MB
MobileNetV2	28.3	-	1.6M	-	0.032	0.777	270MB
Condensenet-74 (G=4)[18]	26.2	8.30	4.8M	529M	0.072	0.694	238MB
NASNet-A (Mobile)	26.0	8.4	5.3M	564M	0.244	-	-
DPP-Net-PNAS	24.16	7.13	77.16M	9276M	0.218	5.421	708MB
DPP-Net-Panacea	25.98	8.21	4.8M	523M	0.069	0.676	238MB

- DPP-Net-Panacea outperforms CondenseNet in every objectives except number of params and memory usage.
- DPP-Net-Panacea outperforms NASNet-A in every objectives

Conclusion

- Use largely available sensory data (w/o label) to self-improve your systems
- Leverage exponential increase of computation to reduce the effort of talents