2023「中技社科技獎學金」
2023 CTCI Foundation Science and Technology Scholarship
境外生研究獎學金
Research Scholarship for Overseas Students

Controllable Model Compression for Roadside Camera Depth Estimation

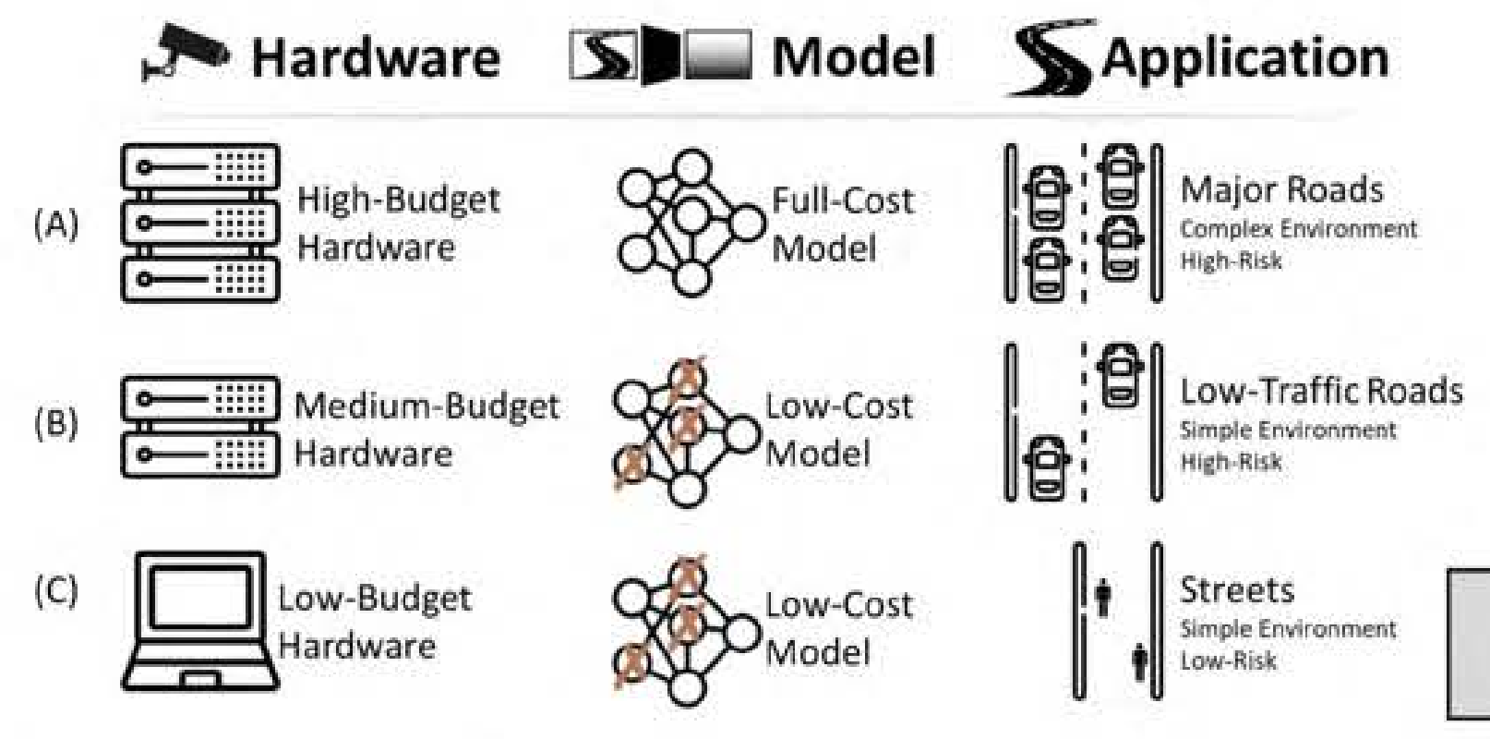
Jose Jaena Mari Ople, Kai-Lung Hua
4th-Year PhD, Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology, Taipei, Taiwan

Abstract

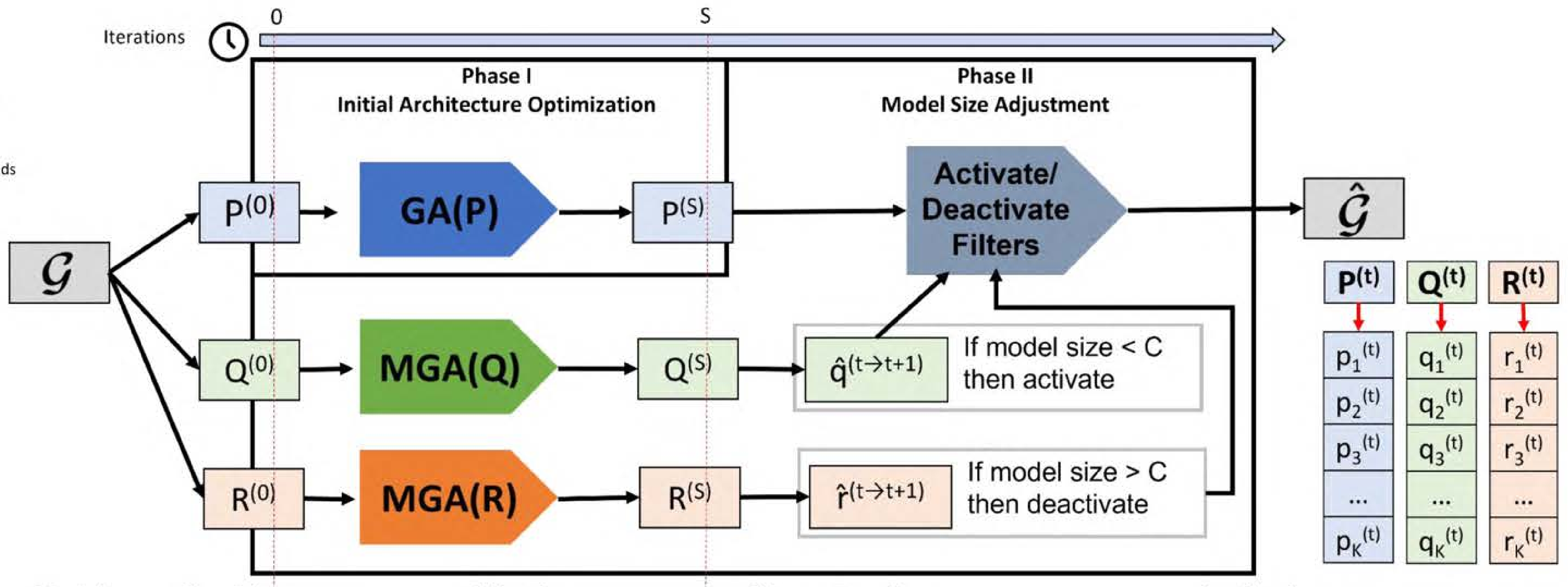
In the Cooperative Intelligent Transportation System (C-ITS) paradigm, vehicles could communicate with roadside units to augment their traffic knowledge. Smart roadside units could provide second-order information (e.g., vehicle count) from raw first-order data (e.g., visual feed, point clouds), and this “smart” feature is usually provided using deep neural network models. However, implementing these useful models implies a cost for computational complexity that could hinder the future deployment of smart roadside units needed for sustainability in transportation systems. In this paper, we propose to use model compression on deep image processing models to promote its feasibility for usage in smart sensors. We formulated a controllable convolutional model compression (CCMC) algorithm that can perform filter-wise evolutionary pruning on image processing networks, along with a predefined compression ratio. CCMC is applicable for image processing networks, which have multiple possible traffic data sources (e.g., road camera surveillance). Furthermore, CCMC has a definable target compression ratio that is useful for controlling the trade-off between resource consumption and output performance. We tested our proposed method on depth estimation, which is useful for scene understanding and mapping the locations of objects in the 3D space. Our experiments show that the pruned model has minimal performance discrepancy from the original one, supporting the sustainability features needed for intelligent transportation systems.

Index Terms— Smart sensors, neural network compression, depth estimation, genetic algorithm, sustainable solutions, intelligent transportation systems.

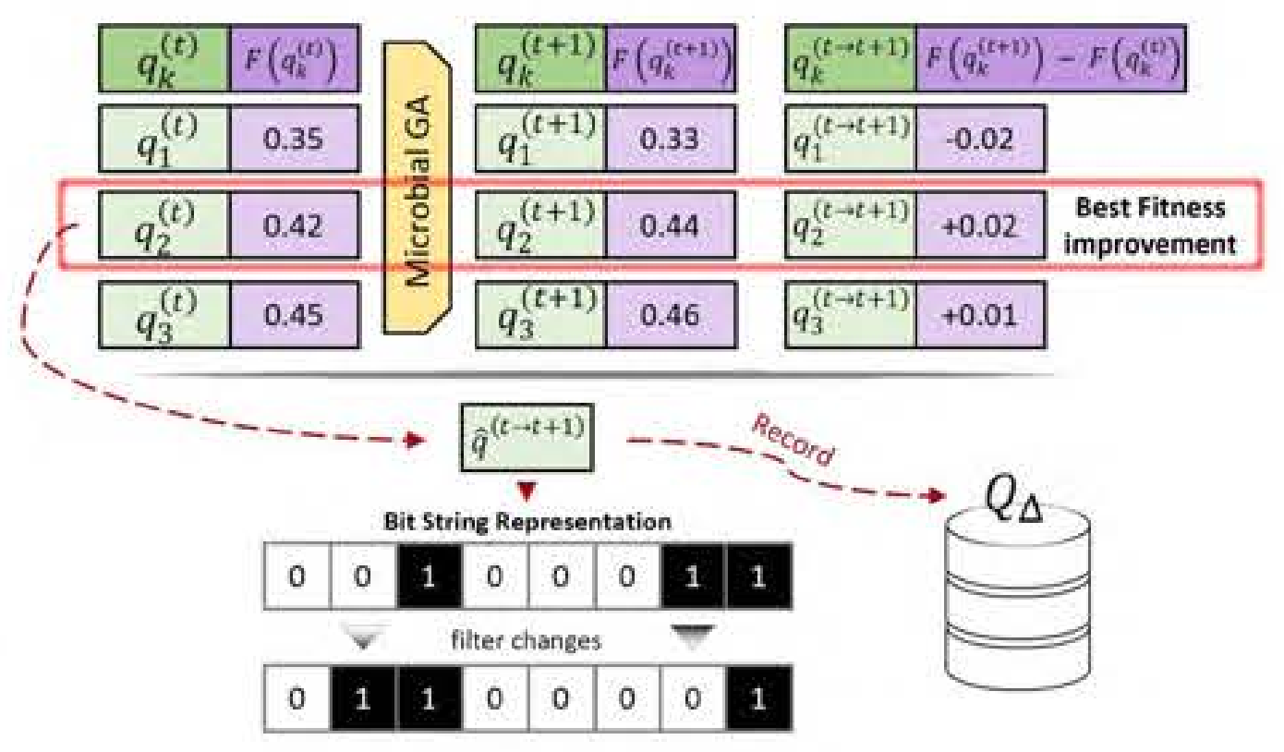
Methodology



Motivation— CCMC is used for compressing convolutional models so that they can be deployed for different types of hardware. In this work, we focus on depth estimation for smart roadside applications. And unlike previous works, our evolutionary filter-wise pruning approach allows controllable compression ratios.



Module	Algorithm	Objective	Fitness Function	Application
GA(P)	Genetic Algorithm	Generate architecture proposals with minimal parameter count and performance decrease	Minimize param count Minimize performance drop	Generate architecture proposals to be modified later
MGA(Q)	Microbial Genetic Algorithm	Find filter activations that cause a large performance increase	Minimize performance drop	Activate good-performing filters to optimally increase the memory sizes of proposed architectures
MGA(R)	Microbial Genetic Algorithm	Find filter activations that cause a large performance decrease	Maximize performance drop	Deactivate bad-performing filters to decrease the memory sizes of proposed architectures



Model Size Adjustment— For MGA(Q) and MGA(R), we record the filter changes and its corresponding change in performance.

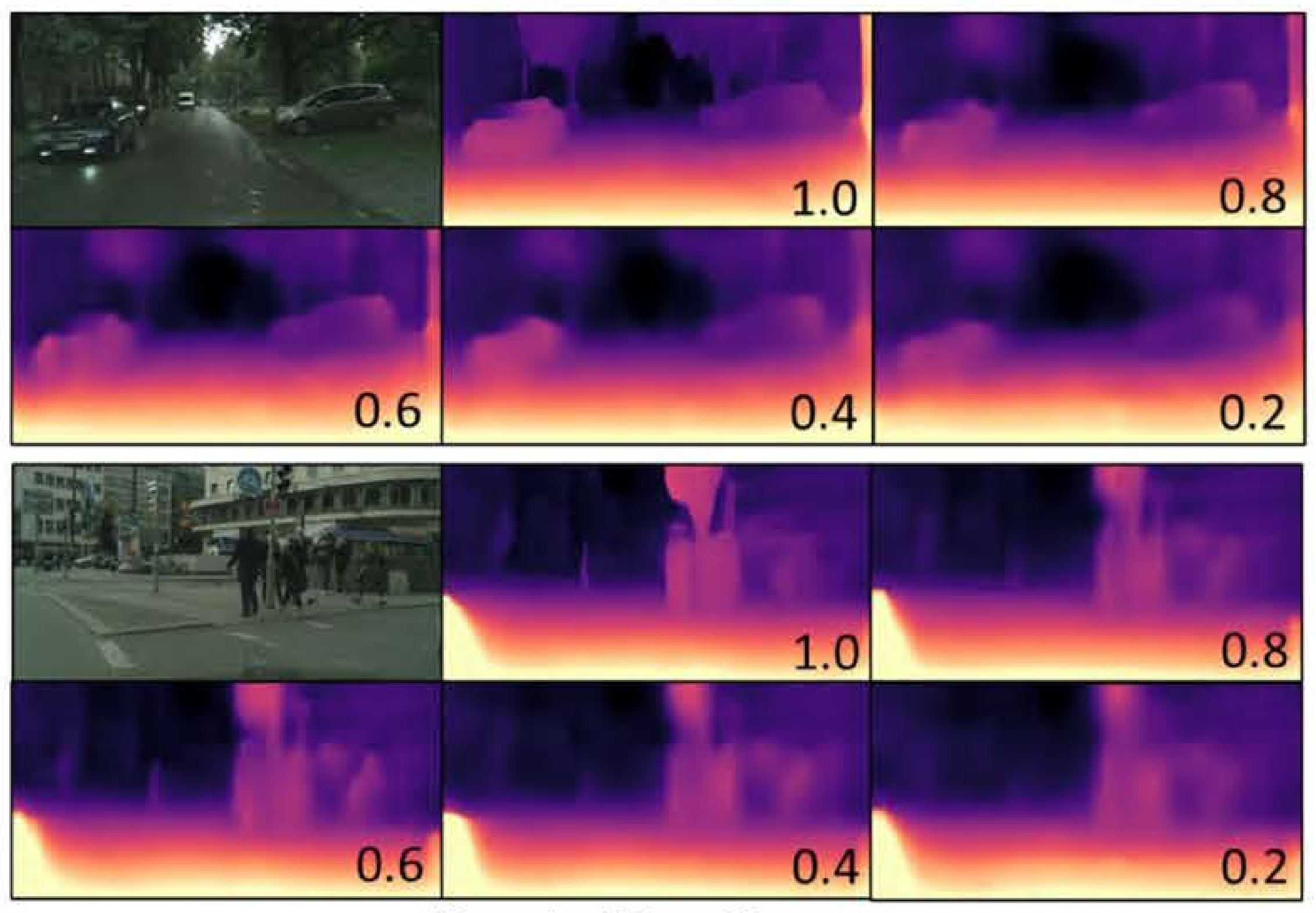
Microbial GA (MGA) is used since it preserves one-to-one correspondence between generations via mutations.

Our objective is to find optimal filter activations/deactivations.

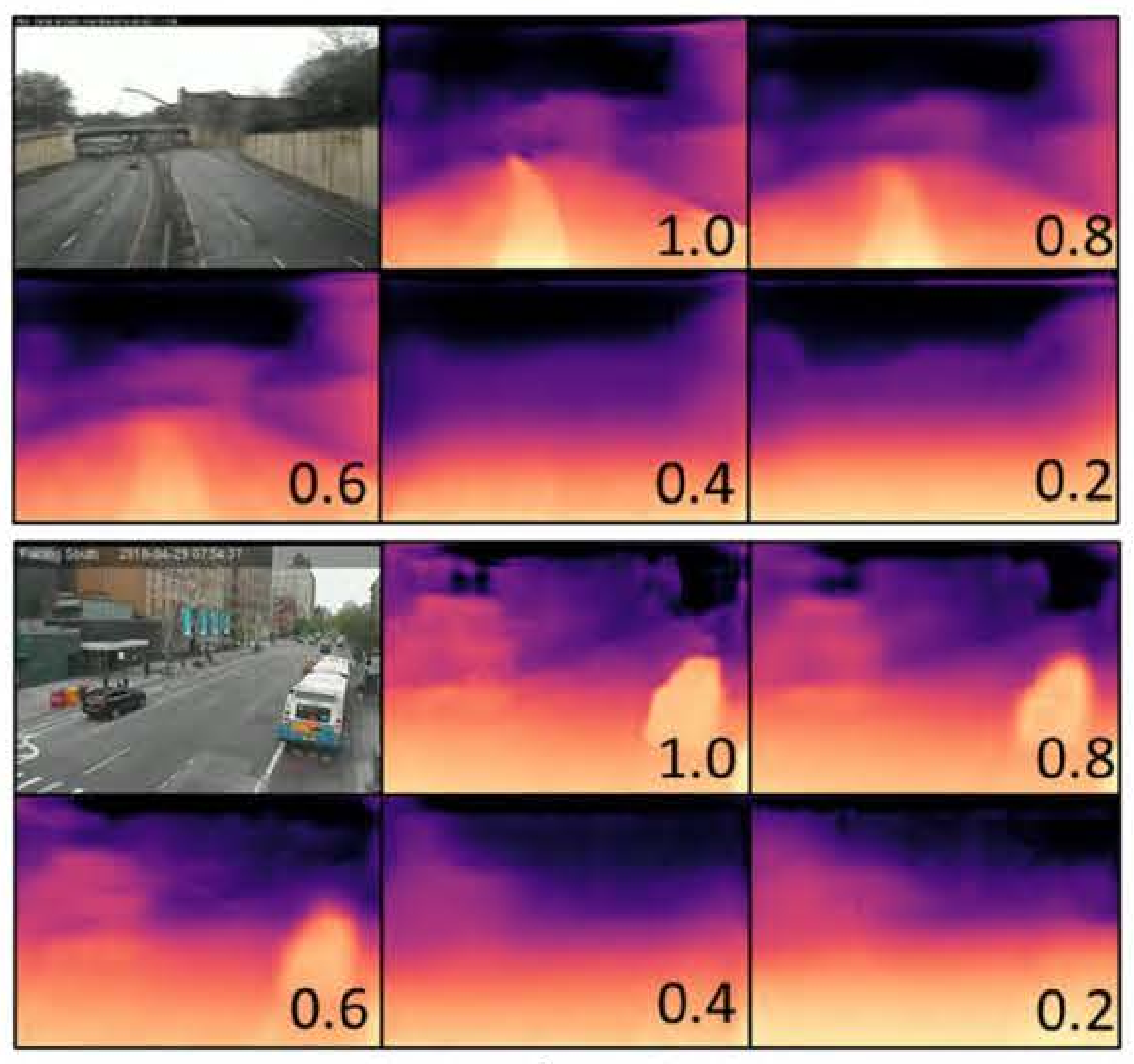
Using MGA(Q) and MGA(R) is our choice of implementation.

Overview— We have an baseline model architecture $\mathcal{G}$. We randomly initialize the populations P, Q, and R with randomly pruned filters based on $\mathcal{G}$. We concurrently run three genetic algorithms for P, Q, and R. And each has a purpose. GA(P) is a typical evolutionary filter-wise pruning algorithm that outputs architecture with good balance of performance and memory size. MGA(Q) and MGA(R) is used to increase or decrease the compression size of the potential models generated from GA(P). We combine the results from MGA(Q) and MGA(R) to GA(P) after S evolutionary iterations. See the figure to the right for more details.

Results



Monodepth2 on Cityscapes



GLPDepth on CityCam

Model		Compression			Performance Metrics						
Arch.	Dataset	Ratio	Actual Ratio	Actual Size	Abs REL	Sq REL	RMSE	RMSE log	$\delta < 1.25^1$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2	Cityscapes	0.8	0.8079	9.96 MB	0.0477	0.0001	0.0027	0.0681	0.9862	0.9987	0.9999
		0.6	0.6196	7.64 MB	0.0494	0.0001	0.0026	0.0781	0.9733	0.9996	0.9999
		0.4	0.4198	5.18 MB	0.0547	0.0002	0.0028	0.0845	0.9680	0.9985	0.9998
		0.2	0.2200	2.71 MB	0.0684	0.0003	0.0040	0.0990	0.9561	0.9953	0.9998
	CityCam	0.8	0.8013	9.88 MB	0.0545	0.0002	0.0030	0.753	0.9821	0.9992	0.9998
		0.6	0.6134	7.56 MB	0.0560	0.0002	0.0032	0.0803	0.9789	0.9993	0.9997
		0.4	0.4086	5.04 MB	0.0581	0.0002	0.0033	0.0888	0.9664	0.9979	0.9996
		0.2	0.2141	2.64 MB	0.0628	0.0002	0.0031	0.0914	0.9660	0.9971	0.9996
GLPDepth	Cityscapes	0.8	0.8132	192.29 MB	0.0833	0.0375	0.3550	0.1116	0.9823	0.9989	0.9999
		0.6	0.6054	143.16 MB	0.0570	0.0140	0.2008	0.0756	0.9457	0.9978	0.9998
		0.4	0.4113	97.26 MB	0.0992	0.0529	0.4289	0.1417	0.9073	0.9808	0.9976
		0.2	0.2189	51.76 MB	0.1172	0.0741	0.4998	0.1660	0.8872	0.9700	0.9953
	CityCam	0.8	0.8172	193.24 MB	0.0166	0.0032	0.1265	0.0267	0.9998	0.9999	0.9999
		0.6	0.6163	145.73 MB	0.0308	0.0075	0.1911	0.0407	0.9992	0.9999	0.9999
		0.4	0.4049	95.74 MB	0.0584	0.0245	0.2990	0.0795	0.9791	0.9998	0.9998
		0.2	0.2177	51.48 MB	0.0585	0.0252	0.3130	0.0817	0.9788	0.9998	0.9998

CCMC can compress depth estimation models (Monodepth2, GLPDepth) to different compression ratios (0.2, 0.4, 0.6, 0.8), and the compressed models is applied to different datasets (Cityscapes, CityCam). Qualitatively, we can see that the depth resolution degrades as the compression ratio decreases but the general depth structure is still preserved even at 20% model size. Quantitatively, the performance degradation due to compression is minimized. As presented in 0.8 and 0.6 compression ratios, their performance is similar despite a relatively large compression difference.

Publication

Ople, Jose Jaena Mari, Shang-Fu Chen, Yung-Yao Chen, Kai-Lung Hua, Mohammad Hijji, Po Yang, and Khan Muhammad. "Controllable Model Compression for Roadside Camera Depth Estimation." *IEEE Transactions on Intelligent Transportation Systems* (2022).

中技社
CTCI FOUNDATION