

DeerScale стартап компаниясының SqueezeNet жүйесі.

Құрманбаев Рахман
М092- Математика

DeerScale компаниясын 2015 жылы
Форрест Йандоло мен Курт
Койцер құрды

2018 А сериясының қаржыландыруынан 15
миллион АҚШ долларын алды. Ал 2019
жылы Тесла компаниясы сатып алды.

DeerScale-дің жасаған SqueezeNet жүйесі бұл болашақтағы автопилотпен жүретін машиналарға кеңістікте ориентация алу үшін қажет. Яғни осы жүйені дамытсақ болашақтағы автопилот жүйесіне негіз бола алады. Бұл жүйеге Uber, Tesla, Microsoft, Google компаниялары қызығушылық танытқан. 2015 жылы стартап компания ретінде құрылып, 2019 жылы Tesla компаниясының құрамына енді.

<https://arxiv.org/pdf/1602.07360.pdf>

SQUEEZE_{NET}: ALEX_{NET}-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE

**Forrest N. Iandola¹, Song Han², Matthew W. Moskewicz¹, Khalid Ashraf¹,
William J. Dally², Kurt Keutzer¹**

¹DeepScale* & UC Berkeley ²Stanford University

{forresti, moskewcz, kashraf, keutzer}@eecs.berkeley.edu

{songhan, dally}@stanford.edu

ABSTRACT

Recent research on deep convolutional neural networks (CNNs) has focused primarily on improving accuracy. For a given accuracy level, it is typically possible to identify multiple CNN architectures that achieve that accuracy level. With equivalent accuracy, smaller CNN architectures offer at least three advantages: (1) Smaller CNNs require less communication across servers during distributed training. (2) Smaller CNNs require less bandwidth to export a new model from the cloud to an autonomous car. (3) Smaller CNNs are more feasible to deploy on FPGAs and other hardware with limited memory. To provide all of these advantages, we propose a small CNN architecture called SqueezeNet. SqueezeNet achieves AlexNet-level accuracy on ImageNet with 50x fewer parameters. Additionally, with model compression techniques, we are able to compress SqueezeNet to less than 0.5MB (510× smaller than AlexNet).

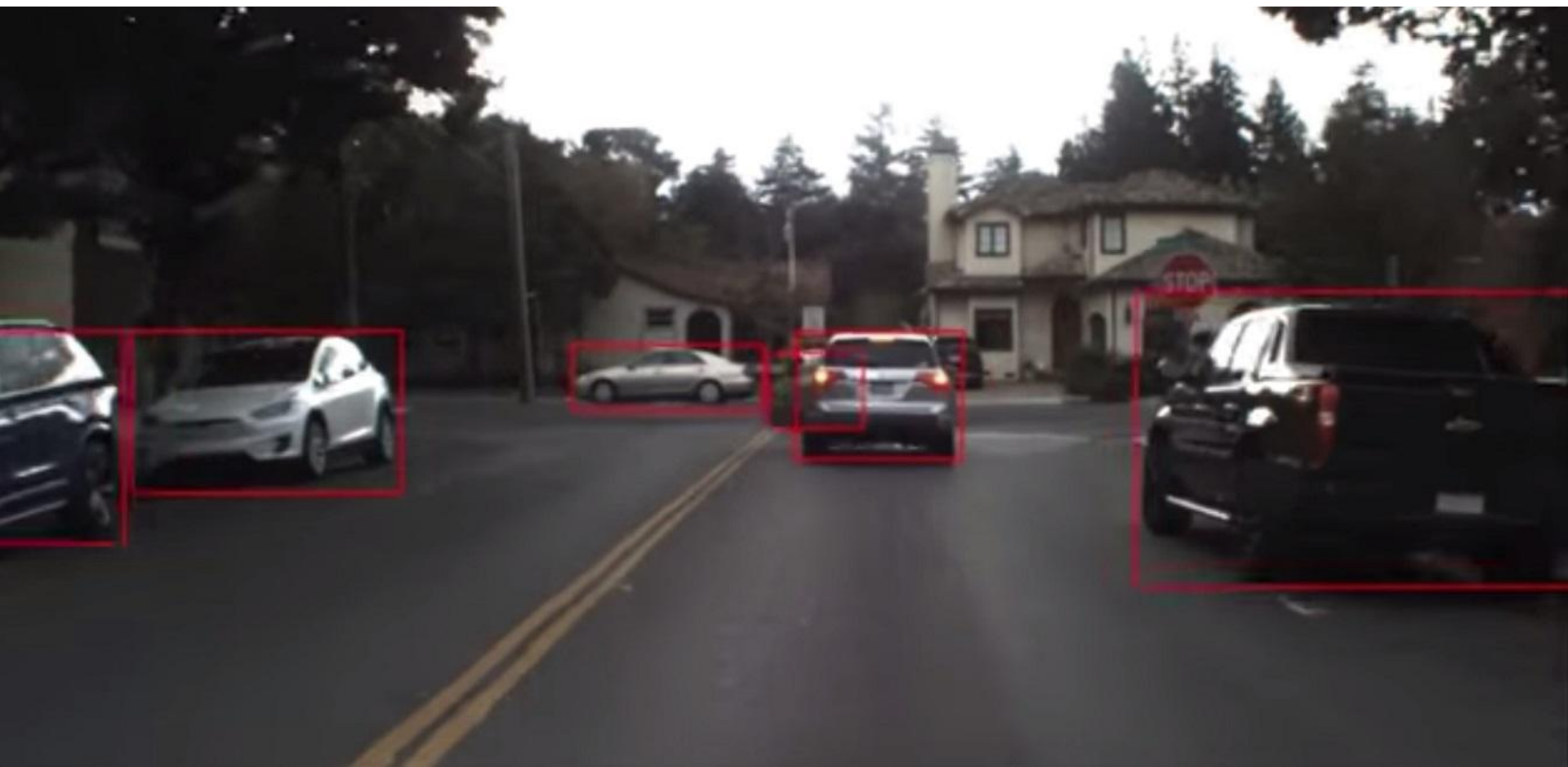
The SqueezeNet architecture is available for download here:
<https://github.com/DeepScale/SqueezeNet>

Бұл мақалада автопилотпен жүретін машиналарға үлкен есептеу қуаты бар компьютерлерді қою экономикалық тұрғыдан тиімсіз екені және үлкен компьютер қою арқылы машинаның пайдалы бөлігін азайтатыны айтылады. Қазіргі таңда қолданылатын AlexNet жүйесіне көлікті жүргізуге 240 мб жады көлемі кететін болса ал SqueezeNet жүйесі үшін 4,8 мб та жеткілікті бұл өз кезегінде автокөлікке минипроцессорларды орнатуға мүмкіндік береді.

Table 2: Comparing SqueezeNet to model compression approaches. By *model size*, we mean the number of bytes required to store all of the parameters in the trained model.

CNN architecture	Compression Approach	Data Type	Original → Compressed Model Size	Reduction in Model Size vs. AlexNet	Top-1 ImageNet Accuracy	Top-5 ImageNet Accuracy
AlexNet	None (baseline)	32 bit	240MB	1x	57.2%	80.3%
AlexNet	SVD (Denton et al., 2014)	32 bit	240MB → 48MB	5x	56.0%	79.4%
AlexNet	Network Pruning (Han et al., 2015b)	32 bit	240MB → 27MB	9x	57.2%	80.3%
AlexNet	Deep Compression (Han et al., 2015a)	5-8 bit	240MB → 6.9MB	35x	57.2%	80.3%
SqueezeNet (ours)	None	32 bit	4.8 MB	50x	57.5%	80.3%
SqueezeNet (ours)	Deep Compression	8 bit	4.8MB → 0.66MB	363x	57.5%	80.3%
SqueezeNet (ours)	Deep Compression	6 bit	4.8MB → 0.47MB	510x	57.5%	80.3%

SqueezeNet жүйесінің AlexNet жүйесінен ерекшелігі AlexNet жүйесі 1x1 өлшемді бөліктерге бөледі. Ал SqueezeNet 3x3 өлшемді бөліктерге бөледі. Сондықтан модельдің көлемі 50 есеге дейін кемиді және бұл көлікті басқаруға мүлде зиян әсерін тигізбейді.



- <https://arxiv.org/pdf/1602.07360.pdf>
- <https://en.wiki>
- <https://habr.com/ru/news/t/469793/pedia.org/wiki/DeepScale>