Review Series of Recent Deep Learning Papers:

Parameter Prediction Paper: Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction

Hyeonwoo Noh, Paul Hongsuck, Seo Bohyung Han

CVPR 2016

Reviewed by : Arshdeep Sekhon

[1]Department of Computer Science, University of Virginia

https://qdata.github.io/deep2Read/

August 25, 2018

Image QA

- Task: Image Question Answering
- Previous Methods: tasks were of the same type: one of some category of objects, etc
- less efforts on solving various recognition problems simultaneously, which is more complex and realistic

Q: What type of animal is this?
Q: Is this animal alone?

Q: Is it snowing?
Q: Is this picture taken during the day?

Q: What kind of oranges are these?

Q: What is leaning on the wall?

Motivation

1. Task defined by the question in Image QA

Motivation

1. Task defined by the question in Image QA
2. solve the heterogeneous recognition tasks using a single CNN by adapting the weights in the dynamic parameter layer

Motivation

1. Task defined by the question in Image QA
2. solve the heterogeneous recognition tasks using a single CNN by adapting the weights in the dynamic parameter layer
3. the weights in the dynamic layer are determined by the question
4. A hashing trick to reduce number of parameters

The Model

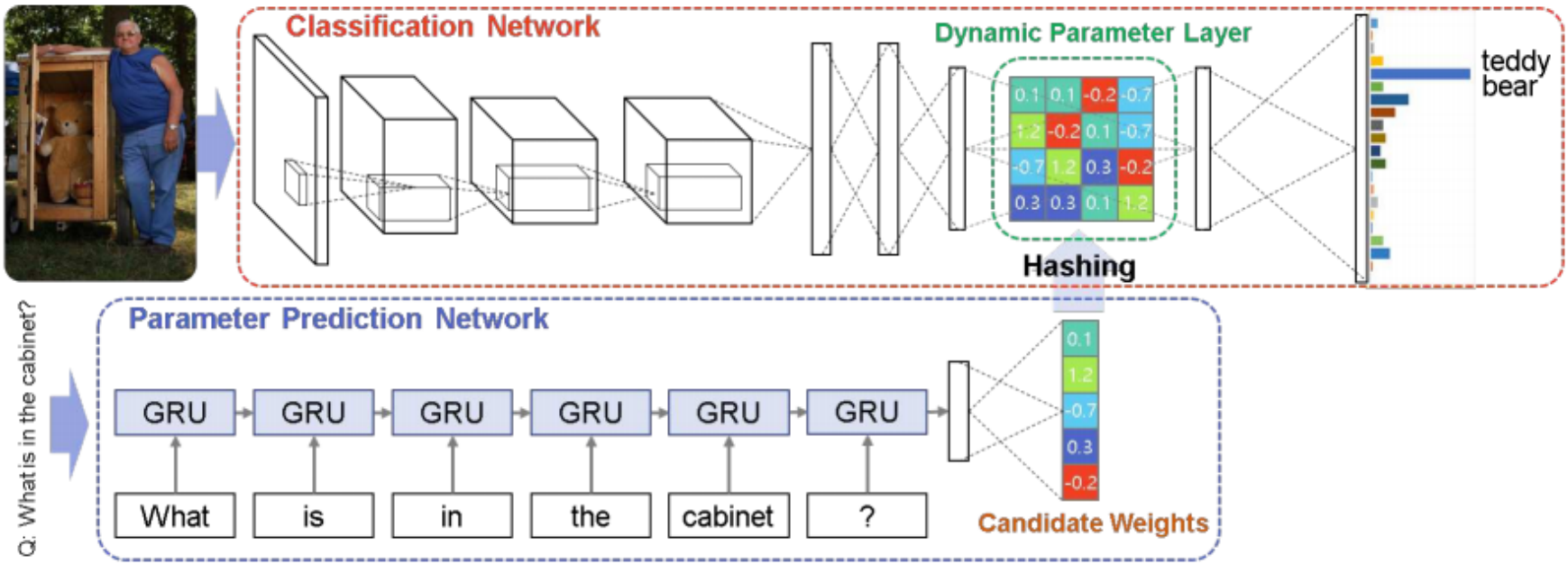
1. Conventional Approach:

$$\hat{a} =_{a \in \Omega} p(a|q, I; \theta) \tag{1}$$

2. Dynamic Parameter Prediction:

$$\hat{a} =_{a \in \Omega} p(a|I; \theta_s, \theta_d(q)) \tag{2}$$

The Model



VGG-Net pretrained on Image Net

second last layer is dynamically predicted.

The Model

1. The parameter prediction network:
2. GRU cells followed by a fully-connected layer, which produces the candidate weights

$$\boldsymbol{p} = \boldsymbol{W}_p \boldsymbol{h}_T \tag{3}$$

3. In the classification network, the dynamic layer

$$\boldsymbol{f^o} = \boldsymbol{W}_d(q)\boldsymbol{f_i} + b \tag{4}$$

Hashing

1. Instead of predicting entire $\boldsymbol{W}_d$, construct $\boldsymbol{W}_d(q)$ based on a small number of candidate weights

Hashing

1. Instead of predicting entire $\boldsymbol{W}_d$, construct $\boldsymbol{W}_d(q)$ based on a small number of candidate weights
2. random weight sharing technique based on hashing
3. a single parameter in the candidate weight vector $\boldsymbol{p}$ is shared by multiple elements of $\boldsymbol{W}_d(q)$

Hashing

1. Instead of predicting entire $\boldsymbol{W}_d$, construct $\boldsymbol{W}_d(q)$ based on a small number of candidate weights
2. random weight sharing technique based on hashing
3. a single parameter in the candidate weight vector $\boldsymbol{p}$ is shared by multiple elements of $\boldsymbol{W}_d(q)$
4. a predefined hash function that converts the 2D location in $\boldsymbol{W}_d(q)$ to the 1D index in $\boldsymbol{p}$
5. $$w_{mn}^d = p_{\psi(m,n)} . \eta(m, n) \tag{5}$$
6. $\psi m, n$ maps a key (m, n) to a natural number in $\{1, \cdots, K\}$
7. $\eta(m, n)$ also a hash function: $\mathbb{N} \times \mathbb{N} \rightarrow \{-1, 1\}$

Results

	Open-Ended				Multiple-Choice			
	All	Y/N	Num	Others	All	Y/N	Num	Others
Question [1]	48.09	75.66	36.70	27.14	53.68	75.71	37.05	38.64
Image [1]	28.13	64.01	00.42	03.77	30.53	69.87	00.45	03.76
Q+I [1]	52.64	75.55	33.67	37.37	58.97	75.59	34.35	50.33
LSTM Q [1]	48.76	78.20	35.68	26.59	54.75	78.22	36.82	38.78
LSTM Q+I [1]	53.74	78.94	35.24	36.42	57.17	78.95	35.80	43.41
CONCAT	54.70	77.09	36.62	39.67	59.92	77.10	37.48	50.31
RAND-GRU	55.46	79.58	36.20	39.23	61.18	79.64	38.07	50.63
CNN-FIXED	56.74	80.48	37.20	40.90	61.95	80.56	38.32	51.40
DPPnet	**57.22**	**80.71**	**37.24**	**41.69**	**62.48**	**80.79**	**38.94**	**52.16**