Self-Attention Based Neural Network for Few Shot Classification

Ravi Jain
Dept of Computer Scienee and Communications Engineering,
Graduate School of FSE
Waseda University,
Tokyo, Japan
vainaijr114@akane.waseda.jp

Hiroshi Watanabe
Dept of Computer Scienee and Communications Engineering,
Graduate School of FSE,
Waseda University,
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—This paper shows results of experiments carried on few shot classification using attention based neural network plus modifications to its formulation.**

***Keywords*—Stand Alone Self Attention, Few Shot Classification**

I. INTRODUCTION

In this paper we discuss results of applying stand-alone self-attention based neural network on few shot image classification, this acts as a replacement for convolutional neural networks, and we demonstrate an improvement in accuracy for 5 shot 5 way, and 1 shot 5 way tasks. Furthermore, we show results on modification of the stand-alone self-attention formulation. We use a baseline approach towards this task.

II. OVERVIEW OF STAND-ALONE SELF-ATTENTION

The concept of stand-alone self-attention, proposed in [1], overcomes the limitations of convolutions to even consider the impact each pixel has on the representation of neighboring pixel, using a query, key and value based mathematical formula.

The basic idea behind the use of this technique, is that representation of each pattern that we see depends on the representation of patterns surrounding it, so one would interpret the same pattern with a different meaning, based on what all patterns it is surrounded by. Alongside, how far each pixel is to a pixel also has an impact on each pixel's representation, this is formulated mathematically by a notion of relative positioning.

Stand Alone Self-attention is computed using this formula (for details please refer to the original paper)

$$y_{ij} = \sum_{a,b \in N_k(i,j)} softmax_{ab}\left(q_{ij}^T k_{ab} + q_{ij}^T r_{a-i,b-j}\right) v_{ab} \qquad (1)$$

III. EXPERIMENT DETAILS

We carry our experiment on few shot image classification task using CUB-200 dataset, as described in the paper CloserLookFewShot [2], we split the dataset into into 100 base, 50 validation, and 50 novel classes. The hyperparameter, training details are included in Table 1.

Table 1 – Training details

Dataset	CUB
Loss function	CrossEntropyLoss (with LabelSmoothing during test phase)
Learning rate	0.001
Optimizer	Adam (SGD during test phase)
Epochs trained for	400
Loss Type	distLinear
Technique used for all experiments	Baseline++
Training Augmentation	Enabled
Shot, Ways	5 shot 5 way, 1 shot 5 way
AttnConv4 denotes (repeated four times, we only use pooling for the first four blocks)	Stand Alone Self Attention -> BatchNorm2d -> ReLU -> MaxPool2d(2)
Kernel size	3x3
Outdim	64
Final Feat dim	1600

IV. PROPOSED CHANGES

We use a Query, Key and Value based stand alone self attention neural network, and apply a few modifications to the formulation. Instead of using a matrix multiplication between the query pixel and the key pixels, we use a scalar dot product, the summation, and multiplication with value part is taken care by einsum operator.

978-1-7281-9802-6/20/$31.00 ©2020 IEEE

We even introduce the notion of a priority matrix, considering a weighted impact of each neighboring pixel on a particular pixel. The modified formulations are listed below.

With priority matrix, (2)

$$y_{ij} = \sum_{a,b \in N_k(i,j)} softmax_{ab}\left(q_{ij} * \left(q_{ij} + k_{ab}\right) * p_{ab} + q_{ij} * r_{a-i,b-j}\right) v_{ab}$$

Without priority matrix, (3)

$$y_{ij} = \sum_{a,b \in N_k(i,j)} softmax_{ab}\left(q_{ij} * \left(q_{ij} + k_{ab}\right) + q_{ij} * r_{a-i,b-j}\right) v_{ab}$$

V. Experiment Results

Here we show comparison of results obtained after applying proposed formulations on the experiment specified above. (higher means compared to those listed in CloserLookFewShot [2]). On using the modification described above, we were able to achieve an improvement in the accuracy, as reported in Table 2, and Table 3. Further experiments which did not give an improvement in accuracy details, have also been reported. All of these experiments where carried using one NVIDIA GTX 1080 Ti

Table 2 – Results on 5 shot 5 way, using Baseline++, on CUB dataset

Backbone	Accuracy	Dropout Rate	Formulation
AttnConv6	67.01% +- 0.72%	0.3 after every AttnConv	Scalar Dot Product
AttnConv6	82.75% +- 0.58% (higher)	Only once at the end (0.3)	Scalar Dot Product with additive similarity (2)
AttnConv6	58.25% +- 0.77%	Only once at the end (0.3)	Matrix multiplication
AttnConv6 with LPPool instead of MaxPool	78.50% +- 0.67%	Only once at the end (0.3)	Scalar Dot Product with additive similarity (2)
AttnConv4	75.69% +- 0.70%	Only once at the end (0.3)	Along with Priority Matrix (3)

Table 3 – Results on 1 shot 5 way, using Baseline++, on CUB dataset

Backbone	Accuracy	Dropout rate (once at the end)	Formulation
AttnConv4	61.84% +- 0.88% (higher)	0.3	Scalar Dot Product with additive similarity (2)
AttnConv6	66.12% +- 0.94% (higher)	0.3	Scalar Dot Product with additive similarity (2)
AttnConv6	56.09% +- 0.82%	0.3	Along with Priority Matrix (3)
AttnConv4	56.45% +- 0.89%	0.3	Along with Priority Matrix initialized with Kaiming Normal (3)
AttnConv4, (Maxpool, with batch norm)	61.70% +-0.92% (higher)	0.5	Scalar Dot Product with additive similarity (2)

VI. Conclusion

In this paper we proposed a novel formula to compute stand-alone self-attention and saw results obtained on Few shot classification experiment, further experiments would involve increasing the kernel size, and carrying out these experiments on **>1** GPU using data parallelism.

VII. Acknowledgment

This research is supported by JICA

VIII. References

[1] **Stand-Alone Self-Attention in Vision Models,** Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, Jonathon Shlens, CoRR, abs/1906.05909, 2019, arXiv

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Wang, and Jia-Bin Huang. **A closer look at few-shot classification**. In International Conference on Learning Representations. CoRR, arXiv:1904.04232v1 [cs.CV] 8 Apr 2019

 978-1-7281-9802-6/20/$31.00 ©2020 IEEE