MaxDropout: Deep Neural Network Regularization Based on Maximum Output Values

Recogna Laboratory

Claudio Filipi Gonçalves dos Santos, Mateus Roder, João Paulo Papa

December 9, 2020

Universidade Federal de São Carlos (UFSCar)
São Carlos, SP - Brasil
Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)
Bauru, SP - Brasil

Summary

Introduction

Introduction

Regularization methods have been used since the very begining of Deep Learning approaches

LeNet-5 was trained used artificial data augmentation

Goal: prevent neural networks to overfit

There are different ways to perform regularization nowadays

Introduction

Some well-known regularization:

- Data augmentation
  - AutoAugmentation [1]
  - Cutout [2]
  - RandomErasing [7]
- Change neurons value during training
  - Dropout [5]
  - Variational Dropout [4]

Introduction

We performed a deeper study on Dropout

Hypothesis: could a neural network using rules on dropping values generalize better?

Proposed Approach

Introduction

Presenting MaxDropout

A regularizer that works by dropping the more active (higher values) during training!

Algorithm

Pseudocode for MaxDropout training algorithm.

while *training* **do**
    for each layer **do**
        $rate \leftarrow U(0, r)$
        $normTensor \leftarrow L2Normalize(Tensor)$
        $max \leftarrow Max(normTensor)$
        $Kept\_Idx \leftarrow Idx\_Of(normTensor, (1 - rate) * max)$
        $returnTensor \leftarrow Tensor * Kept\_Idx$
    end for
end while

Simulation

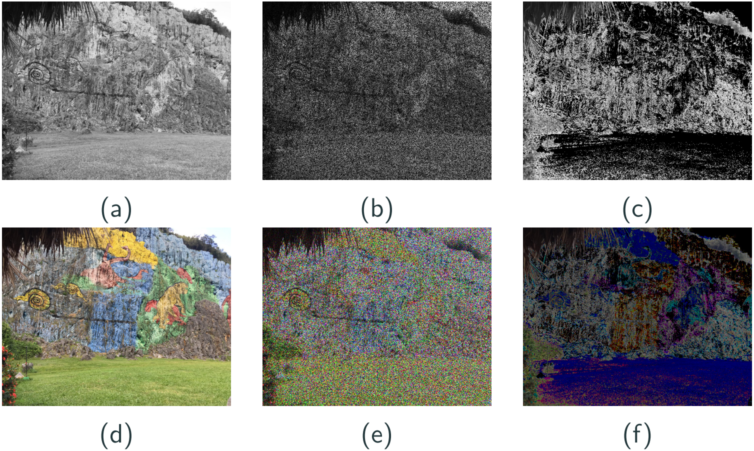

Figure 1: (a) Original grayscale image and its outcomes after (b) Dropout and (c) MaxDropout transformations, respectively, and (d) original colored image and its outcomes after (e) Dropout and (f) MaxDropout, respectively (dropout rate of 50%).

Experiments

Three direct comparisons:

- Comparison against other regularizers
- Comparison along other regularizers
- Comparison against Dropout[5]
- Ablation

All experiments were performed on CIFAR-10 and CIFAR-100

Results

Results - Training Evolution

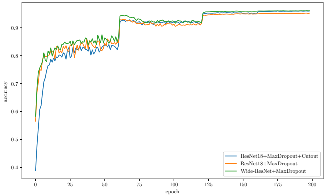

Figure 2: Convergence over CIFAR-10 test set.

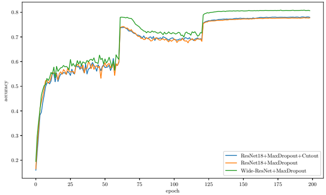

Figure 3: Convergence over CIFAR-100 test set.

Results - Comparison Against Other Regularizers

Approach	CIFAR-100	CIFAR-10
ResNet18 [3, 7]	24.50 ± 0.19	5.17 ± 0.18
ResNet18+RandomErasing [7]	24.03 ± 0.19	4.31 ± 0.07
ResNet18+Cutout [2]	21.96 ± 0.24	$\mathbf{3.99 \pm 0.13}$
ResNet18+MaxDropout	$\mathbf{21.94 \pm 0.23}$	4.63 ± 0.11

Table 1: Results of MaxDropout and other regularizers

Results

Regularizer	CIFAR-100	CIFAR-10
Cutout [2]	21.96 ± 0.24	3.99 ± 0.13
MaxDropout	21.94 ± 0.23	4.63 ± 0.11
MaxDropout + Cutout	$\mathbf{21.82 \pm 0.13}$	$\mathbf{3.76 \pm 0.08}$

Table 2: Results of the MaxDropout combined with Cutout.

Results - Working Along with Other Regularizers

Model	CIFAR-100	CIFAR-10
WRN [6]	19.25	4.00
WRN + Dropout [6]	18.85	3.89
WRN + MaxDropout	**18.81**	**3.84**

Table 3: Results of Dropout and MaxDropout over the WRN.

Results - Ablation

MaxDropout Rate (r)	CIFAR-100	CIFAR-10
05	22.05 ± 0.17	4.76 ± 0.09
10	22.06 ± 0.32	4.71 ± 0.09
15	22.16 ± 0.20	$\mathbf{4.63 \pm 0.11}$
20	21.99 ± 0.21	4.70 ± 0.08
25	$\mathbf{21.94 \pm 0.23}$	4.70 ± 0.06
30	22.08 ± 0.24	4.67 ± 0.12
35	22.10 ± 0.29	4.71 ± 0.16
40	22.17 ± 0.34	4.79 ± 0.20
45	22.31 ± 0.29	4.71 ± 0.11
50	22.33 ± 0.23	4.75 ± 0.10

Table 4: Ablation results concerning MaxDropout over ResNet18.

Conclusão

Conclusion

We introduced MaxDropout, an improved version of the original Dropout method

Experiments show that:

- it can be incorporated into existing models
- it works along with other regularizers
- it can replace the standard Dropout with some accuracy improvement

Future work

Two main tasks:

- re-implement in other frameworks (TensorFlow, MXNet)
- test in other problems (such as NLP)

Source Code

https://github.com/cfsantos/MaxDropout-torch

References

[1] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le.
Autoaugment: Learning augmentation policies from data.
arXiv preprint arXiv:1805.09501, 2018.

[2] T. DeVries and G. W. Taylor.
Improved regularization of convolutional neural networks with cutout.
arXiv preprint arXiv:1708.04552, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun.
Identity mappings in deep residual networks.
In *European conference on computer vision*, pages 630–645. Springer, 2016.

[4] D. P. Kingma, T. Salimans, and M. Welling.
Variational dropout and the local reparameterization trick.
In *Advances in Neural Information Processing Systems*, pages 2575–2583, 2015.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov.
Dropout: a simple way to prevent neural networks from overfitting.
The journal of machine learning research, 15(1):1929–1958, 2014.

[6] S. Zagoruyko and N. Komodakis.
Wide residual networks.
arXiv preprint arXiv:1605.07146, 2016.

[7] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang.
Random erasing data augmentation.
In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

Questions?

Thank you all!

RECOGNA
LABORATORY