Is Deep-Paper-Gestalt A Valid Way to Review Paper

EE447 Project Final Report Qi Zhenlin 516030910531 2020.05.20 Main Contents

 \rightarrow Problem background

 \rightarrow Existing work

 \rightarrow My experiment

 \rightarrow Conclusion

What is Gestalt?

As used in Gestalt psychology, the German word *gestalt* is interpreted as "pattern" or "configuration".



Gestalt psychologists emphasized that organisms perceive entire patterns or configurations, not merely individual components. [1]

Can we review scientific paper via a Gestalt process ?

▲ A fact: The number of the paper submissions to top-tier computer vision conferences has been increased dramatically over the past few years. [2]



In order to reduce the pressure of reviewers, we are willing to leverage the super power of AI, especially the Deep Neural Networks, to judge a scientific paper.

But question is, is that valid? Will there be any good paper mistakenly rejected by an ignorant AI? Main Contents

 \rightarrow Problem background

7

 \rightarrow Existing work

 \rightarrow My experiment

 \rightarrow Conclusion

There is some existing research using the SOTA cv model, ResNet-18, to conducting prediction on good/bad paper.

Cornell University

arXiv.org > cs > arXiv:1812.08775

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 20 Dec 2018]

Deep Paper Gestalt

Jia-Bin Huang

Recent years have witnessed a significant increase in the number of paper submissions to computer vision conferences. The sheer volume of paper submissions and the insufficient number of competent reviewers cause a considerable burden for the current peer review system. In this paper, we learn a classifier to predict whether a paper should be accepted or rejected based solely on the visual appearance of the paper (i.e., the gestalt of a paper). Experimental results show that our classifier can safely reject 50% of the bad papers while wrongly reject only 0.4% of the good papers, and thus dramatically reduce the workload of the reviewers. We also provide tools for providing suggestions to authors so that they can improve the gestalt of their papers.

Comments: Project page: this https URL

Subjects: Computer Vision and Pattern Recognition (cs.CV) Cite as: arXiv:1812.08775 [cs.CV] (or arXiv:1812.08775v1 [cs.CV] for this version) Search... Help

Researchers state that: Our classifier safely rejects the number of bad paper submissions by half, while only sacrificing 0.4% of the good paper submission. [2]



Q

About the Dataset, [2]

We collect positive examples (good papers) from the list of accepted papers in top-tier computer vision conferences (CVPR, ICCV).

For negative examples, as we do not have access to papers that were rejected from these conferences, we use workshop papers as an approximation.

A glimpse on the Dataset

	area	百日社社		arear 41		
22 () III					4 boto	
			A CONTRACTOR OF A CONTRACTOR O			5500 <u>6550</u> (ma ₆) (+
						••••••••••••••••••••••••••••••••••••••
	t t				***-i) :	
					••••••••••••••••••••••••••••••••••••••	

A glimpse on the Dataset



Workshop papers

Conference papers

A subtle problem:

The paper downloaded from *Computer Vision Foundation* (<u>https://www.thecvf.com/</u>) have different symbol for conference paper and workshop paper.

GyF

This CVPR paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the version available on IEEE Xplore.

This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the version available on IEEE Xplore.

Overview of the framework proposed by this work:



Result 1: The performance of the model presented by [2] is shown below.

▲ Is seems possible to protect all the good paper as we want to reject some of the bad ones.



Result 2: The model can give some reason (pointing out the good looking parts) for paper justified to be good.



Note the top-right region of the first page.

Main Contents

 \rightarrow Problem background

 \rightarrow Existing work

 \rightarrow My experiment

 \rightarrow Conclusion

Dataset and Model download

From the GitHub repo offered by the author of [2] we can download whole dataset and pre-trained model.

The GitHub link: <u>https://github.com/vt-vl-</u> <u>lab/paper-gestalt</u> [3] A problem:

The proposed model is based on the ResNet-18 backbone.

ResNet-18 need input as (3x224x224) tensor. But the original images in the Dataset are in (3x440x680).

We need to do some data preprocess.

Use the Python-Opencv lib to resize input images,



Question: Which interpolation method should we use ?

It is an important detail, which the original paper did not include. Because DNN is sensitive to the input values [4].

I choose to use one of the interpolation method offered by the Opencv lib, and re-train (finetune) the ResNet-18 network.

Recent test Acc is 0.9077551020408163 PS C:\Docs\FromGithub\Paper Gestalt\MyTestCode> The result I get from finetune (with early stop by validation).



Question: Can we get better accuracy without changing the backbone of model ?

The answer is data augmentation. But how ?



Maybe we can start from the definition of Gestalt psychology, which includes "inference from partial information".





When we simply dropout the title region of each input image...



Divide the whole input into 64 subregion, dropout one at a time and record the performance degradation.



Randomly use input level dropout when training, and finally use full image to train for one epoch.



Some interesting questions about this work:

<1> What kind of paper will get strong reject by this Deep Gestalt system ?

<2> Is this system self-consistent ? (To score itself).

<3> Can it be used to other fields of paper ?
(The quality of generalization).

<1> What kind of paper will get strong reject by this Deep Gestalt system?



Select one of the "worst" paper and make it better.



Rank 95%+ in test set.

Rank 40%- in test set.

<2> Is this system self-consistent?



We apply the trained classifier to this paper. Our network ruthlessly predicts with high probability (over 97%) that this paper should be rejected. We can also do some simple fixing to make it willing to accept itself (:D)



Now it only have a probability (about 74%) that this paper should be rejected.

<3> Can it be use to other field of paper? (The quality of generalization).

I just use the paper from ICML-2018, a more comprehensive machine learning conference. (https://github.com/niudd/ICML-2018-Papers)

Recent test Acc is 0.7848875255623721 PS C:\Docs\FromGithub\Paper_Gestalt\MyTestCode> If we substitute the original positive dataset with ICML-18 papers, then the result on original dataset is even worse.

Recent test Acc is 0.7731236603844124 PS C:\Docs\FromGithub\Paper Gestalt\MyTestCode>

If we combine the CVF dataset and ICML dataset and re-train the model, the accuracy is lower than both the single dataset.

Recent test Acc is 0.8967280163599182 PS C:\Docs\FromGithub\Paper_Gestalt\MyTestCode>

Main Contents

- \rightarrow Problem background
- \rightarrow Existing work
- \rightarrow My experiment
- → Conclusion

Advantage

I. The Deep Paper Gestalt model can really learn some useful representation, which can help review scientific paper.

II. Through data augmentation it can reach better result on classification accuracy.

Disadvantage

I. Actually this approach has low quality when we consider robustness, because it is easy to be cheated.

II. This DNN based model has really low generalization ability, which can not be fixed by simply enlarge dataset.

The End. Thanks.

Reference

[1] "Gestalt psychology". Britannica Concise Encyclopedia. Encyclopaedia Britannica, Inc. 2008-05-01. p. 756. ISBN 9781593394929.

[2] Huang, J. B. (2018). Deep Paper Gestalt. arXiv preprint arXiv:1812.08775.

[3] https://github.com/vt-vl-lab/paper-gestalt

[4] Su, J., Vargas, D. V., & Sakurai, K. (2019). One pixel attack for fooling deep neural networks. IEEE Transactions on Evolutionary Computation, 23(5), 828-841.