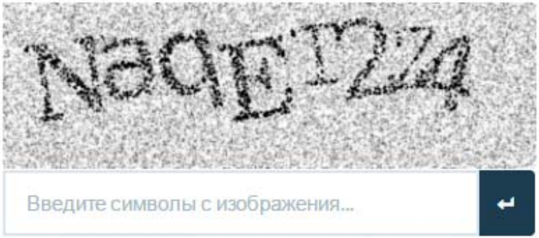
Using Domain Adversarial Learning for Text Captchas Recognition

Kushchuk D.O.
Ryndin M. A.
Yatskov A. K.
Varlamov M. I.
Ivannikov Institute for System Programming of the RAS

25 September 2020

Introduction

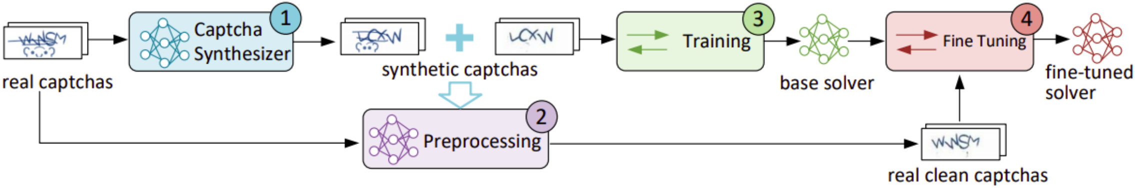
- Captcha - Completely Automated Public Turing test to tell Computers and Humans Apart.
- In our work, captcha is understood as an image with a noisy background consisting of a priory unknown number of letters and Arabic numerals.



Relevance

- Automatic collection of information from sites
- The usefulness of finding and investigating vulnerabilities of public Turing tests for their improvement
- The existence of a number of limitations in modern methods of recognition of noisy images

State-of-the-Art [1]

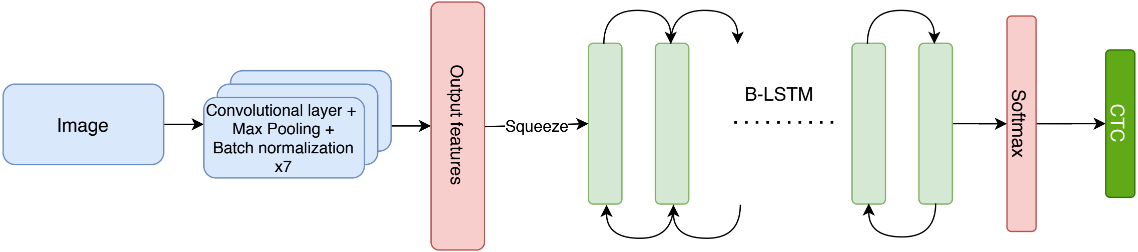


[1] [Guixin Ye et al., 2019, Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach]

Problem statement

- Create a method for recognizing captcha images with variable text's length on it
- Use a small amount of labeled real data for training
- Reduce time for learning and human work

One-step algorithm CRNN[2]



[2] [Baoguang Shi et al. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition]

Results on synthetic data

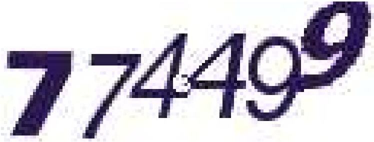
19000 synthetic generated captchas were used for learning

	Time of learning, min	Accuracy, %	Average time of recognition 1 captcha, sec
CRNN	85	97	0,0016

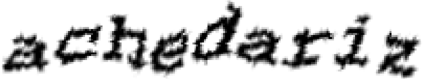
Example of synthetic data

Real data

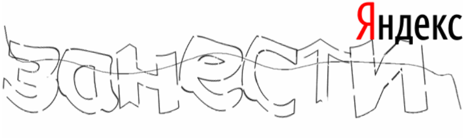
3 set of real data images were collected in the amount of 9000 each and only 1500 of it were labeled – 500 for training and 1000 - for testing



eBay



Wikipedia

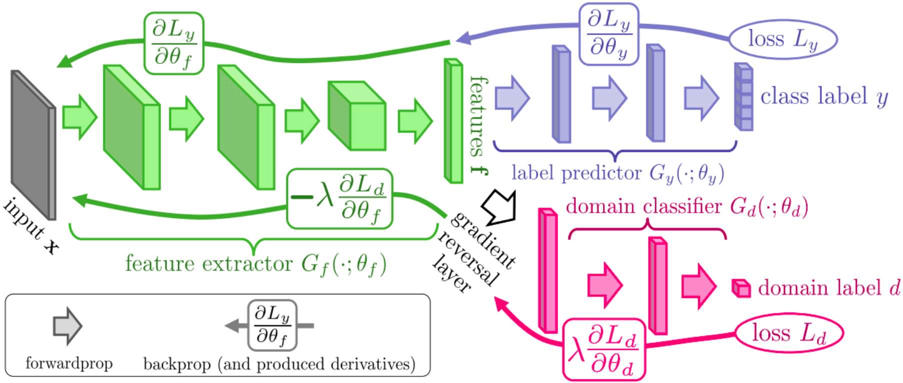


Yandex

Treating our task as domain adaptation task

- The accuracy of the CRNN model, which was trained on synthetic data, on real data is zero
- To train the model specially for real data, it takes about 19 thousand labeled images, which is a lot
- Our task is similar to the task of domain adaptation – adaptation the model on real data to decrease the amount of labeled real examples for training
- We are going to apply an algorithm from this area of knowledge to our task

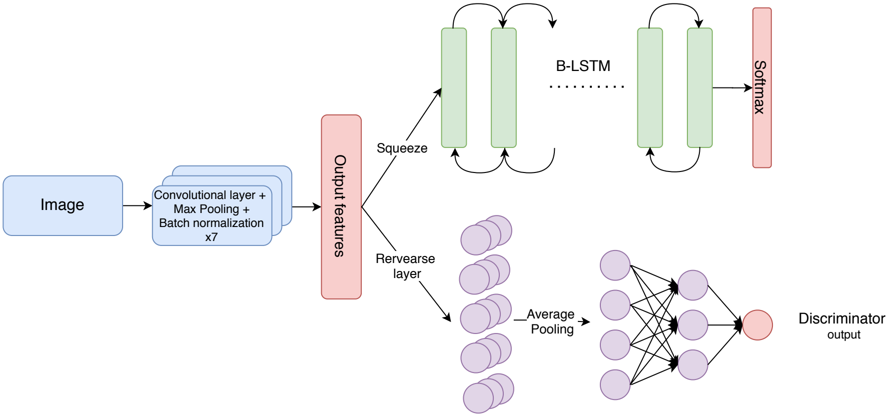
DANN architecture[3]



$\frac{\partial L_y}{\partial \theta_f}$
$\frac{\partial L_y}{\partial \theta_y}$
loss L_y
class label y
features $\mathbf{f}$
label predictor $G_y(\cdot;\theta_y)$
input $\mathbf{x}$
$-\lambda \frac{\partial L_d}{\partial \theta_f}$
feature extractor $G_f(\cdot;\theta_f)$
gradient reversal layer
domain classifier $G_d(\cdot;\theta_d)$
domain label d
loss L_d
$\lambda \frac{\partial L_d}{\partial \theta_d}$
forwardprop
$\frac{\partial L_y}{\partial \theta_f}$
backprop (and produced derivatives)

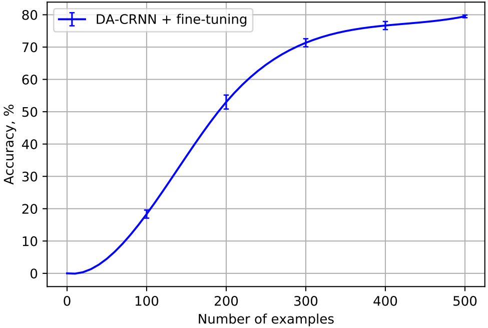
[3] [Ganin et al., Domain-Adversarial Training of Neural Networks]

DA-CRNN architecture

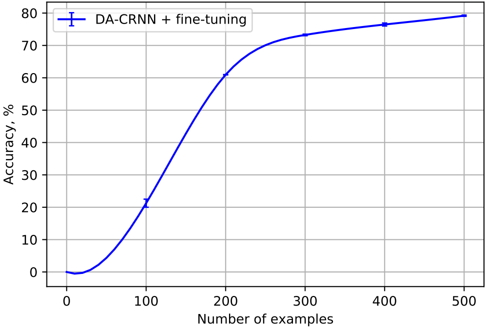


Integration architecture CRNN in DANN. New model is named DA-CRNN

Results DA-CRNN on real data (1)

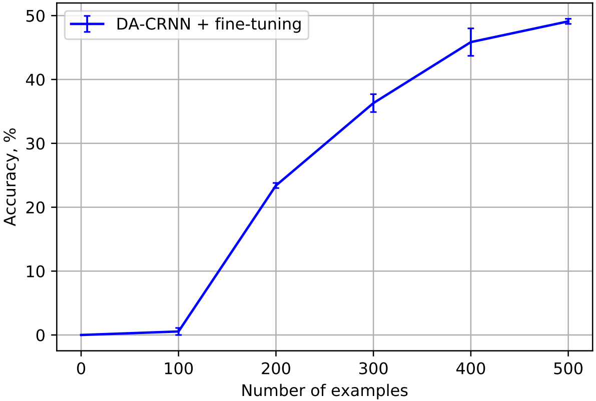


Accuracy on Wikipedia



Accuracy on eBay

Results DA-CRNN on real data (2)



Accuracy on Yandex

Comparison of the developed method DA-CRNN with modern analogs

Method	Wikipedia	Yandex	eBay
Haichang Gao et al.[3]	23.8%	-	58.8%
Elie Bursztein et al. [2]	25%	-	43%
Elie Bursztein et al. [1]	28%	-	51.39%
M. Tang et al. [4]	-	56.0%	-
Sheng Tian et al. [5]	66.5%	63.2%	**91.5%**
Guixin Ye et al. [6]	78%	-	86.6%
DA-CRNN	**79%**	**71.26%**	79.3%

Time for training and evaluation

Метод	Ресурсы	Время на обучение, ч	Время разгадывания 1 капчи, мс
Ref. [6]	4 x NVIDIA Tesla P40 GPU, 256GB of RAM	53	50
Ref. [5]	2 x NVIDIA Tesla P40 GPU	-	6
DA-CRNN	NVIDIA GeForce GTX 1080, 64GB of RAM	**12**	**1,6**

Conclusion

- An experiment of the CRNN model was carried out on synthetic data, showing high results of captcha recognition with an variable text's length on artificial images.
- A set of the real data from Wikipedia, eBay and Yandex was collected and labeled.
- It has been shown experimentally that the usage of CRNN algorithm in conjunction with the adversarial learning method DANN allows one to achieve comparable or higher quality on real data, using the low number (200-500) of labeled examples for training.

To sum up, the DA-CRNN architecture was developed, which achieves the quality similar to modern effective methods, requires a minimum amount of labeled real data for training and requires less time for learning and labeling.

Список литературы I

[1] Elie Bursztein, Jonathan Aigrain и Angelika Moscicki. “The End is Nigh: Generic Solving of Text-based CAPTCHAs”. в: 2014.

[2] Elie Bursztein, Matthieu Martin и John Mitchell. “Text-Based CAPTCHA Strengths and Weaknesses”. в: New York, NY, USA, 2011, 125–138. ISBN: 9781450309486. DOI: 10.1145/2046707.2046724. URL: https://doi.org/10.1145/2046707.2046724.

[3] Haichang Gao и др. “A Simple Generic Attack on Text Captchas”. в: *NDSS*. 2016.

[4] M. Tang и др. “Research on Deep Learning Techniques in Breaking Text-Based Captchas and Designing Image-Based Captcha”. в: *IEEE Transactions on Information Forensics and Security* 13.10 (2018), с. 2522—2537.

[5] Sheng Tian и Tao Xiong. “A Generic Solver Combining Unsupervised Learning and Representation Learning for Breaking Text-Based Captchas”. в: New York, NY, USA: Association for Computing Machinery, 2020, 860–871. ISBN: 9781450370233. DOI: 10.1145/3366423.3380166.

Список литературы II

[6] Guixin Ye и др. “Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach”. в: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. 2018, с. 332—348.