CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Poster

Deep Learning Based Communication: an Adversarial Approach

Yousef Emami

Rahim Taheri

CISTER-TR-190605

Deep Learning Based Communication: an Adversarial Approach

Yousef Emami, Rahim Taheri

CISTER Research Centre

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail:

https://www.cister-labs.pt

Abstract

Deep learning based communication using autoencoder have revolutionized the design of physical layer inwireless communication. In this paper, we propose an adversarial autoencoder to mitigate vulnerability ofautoencoder against adversarial attacks. Results confirm the effectiveness of adversarial training by reducingblock error rate (BLER) from 90 percent to 56 percent.

Deep Learning Based Communication: an Adversarial Approach

Yousef Emami[1], Rahim Taheri[2]

[1] Faculteldade de Engenharia, Universidade do Porto,4200-465 PORTO, Portugal (up201809175@fe.up.pt)

[2]Department of Computer Engineering and IT, Shiraz University of Technology,Shiraz, Iran(r.taheri@sutech.ac.ir)

Abstract

Deep learning based communication using autoencoder have revolutionized the design of physical layer in wireless communication. In this paper, we propose an adversarial autoencoder to mitigate vulnerability of autoencoder against adversarial attacks. Results confirm the effectiveness of adversarial training by reducing block error rate(BLER) from 90 percent to 56 percent.

.

Author Keywords. Deep learning, autoencoder, adversarial autoencoder, white-box attacks.

1. Introduction

Deep learning based communication using autoencoder made possible to learn the best way to communicate over combinations of hardware and channel effects through end-to-end optimization which is defined as jointly optimizing the full chain of physical layer signal processing from the transmitter to receiver(Ben Hilburn, O'Shea, Tim, n.d.).The idea of autoencoder based communication ({Hoydis} 2017) cannot deal with hardware imperfection in an effective way further gradient of the channel is not accessible during training hence ({Brink} 2018) extended this idea to overcome such challenges. However, the salient drawback of the this idea is its vulnerability against adversarial attacks as revealed in ({Larsson} 2019). In this paper we do adversarial training to increase the robustness of autoencoder proposed in ({Hoydis} 2017) against adversarial attacks, the results confirm the effectiveness of our approach. Our main contribution is proposing an adversarial autoencoder which increases the robustness of autoencoder against adversarial attacks.

2. Proposed approach

In order to increase the robustness of the autoencoder against adversarial attacks, we train the autoencoder using adversarial training. In this regard, an adversarial autoencoder is proposed as depicted in figure 1. Adversarial training is a simple and effective method proved to increase the robustness of the model against adversarial attacks. It is notable this approach is only effective when the attack method used to generate the adversarial example is the same as the method used by the attacker({Larsson} 2019). In the proposed approach, The channel is represented by an additive noise layer with a fixed variance $\beta = (2RE_b/N_0)^{-1}$ where E_b/N_0 denotes the energy per bit (E_b) to noise power (N_0).

In the second step, we conduct a white box attack against adversarial autoencoder when trained with normal data and measure BLER. In the third step we train adversarial autoencoder according to its functionality, we feed generator with perturbation and feed discriminator with real data and in this case conduct a white-box attack and again measure

BLER.

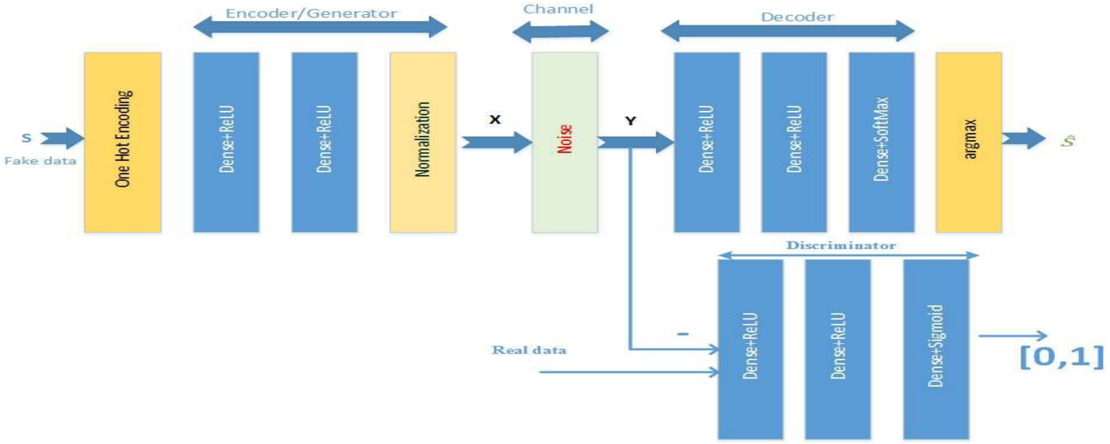

Figure 1. The structure of the proposed adversarial autoencoder

3. Discussion

Figure 2 demonstrates the results of our experiment, the dashed line is BLER of white-box attack before adversarial training which is 90 percent. The solid line is BLER of white-box attack after adversarial training which leveled off in 56 percent.

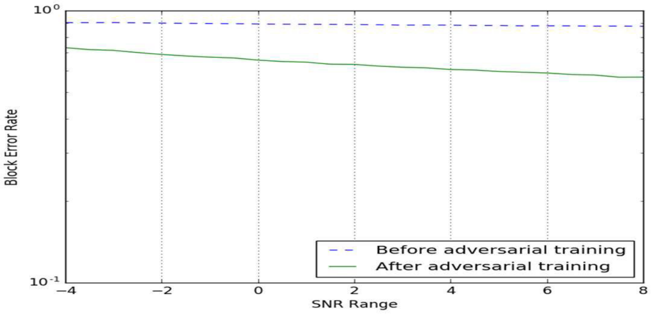

Figure 2. White-box attack before adversarial training and after adversarial training

4. Conclusions

We did adversarial training by proposing an adversarial autoencoder. The results confirm the effectiveness of adversarial training in increasing robustness however attacker still can be successful by utilizing an attack model different that of one utilized for adversarial training. As a future work, we plan to conduct a black-box attack.

References

{Brink}, S. {Dörner} and S. {Cammerer} and J. {Hoydis} and S. t. 2018. "Deep Learning Based Communication Over the Air." *IEEE Journal of Selected Topics in Signal Processing* 12 (1): 132–43.

{Hoydis}, T. {O'Shea} and J. 2017. "An Introduction to Deep Learning for the Physical Layer." *IEEE Transactions on Cognitive Communications and Networking* 3 (4): 563–75.

{Larsson}, M. {Sadeghi} and E. G. 2019. "Physical Adversarial Attacks Against End-to-End Autoencoder Communication Systems." *IEEE Communications Letters* 23 (5): 847–50.

Hilburn, O'Shea, Tim, Tamoghna Roy Ben. n.d. "DeepSig: Deep Learning for Wireless Communications."https://devblogs.nvidia.com/deepsig-deep-learning-wireless-ommunications/.

5. Acknowledgement

This work was partially supported by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit (UID/CEC/04234).