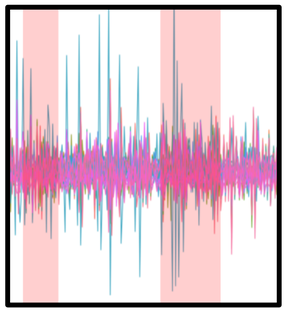
GAN-Based Data Augmentation for Epileptic Seizure Annotation

CS4180 Deep Learning 2018/2019

Noor Ul Sehr Zia, Srinath Sudharsan,
Niels de Bruin, Georgios Andreadis

1 Introduction

- Epilepsy affects **>50M individuals** worldwide
- Manual analysis takes **hours** per patient
- Automating it means **more time for patients**

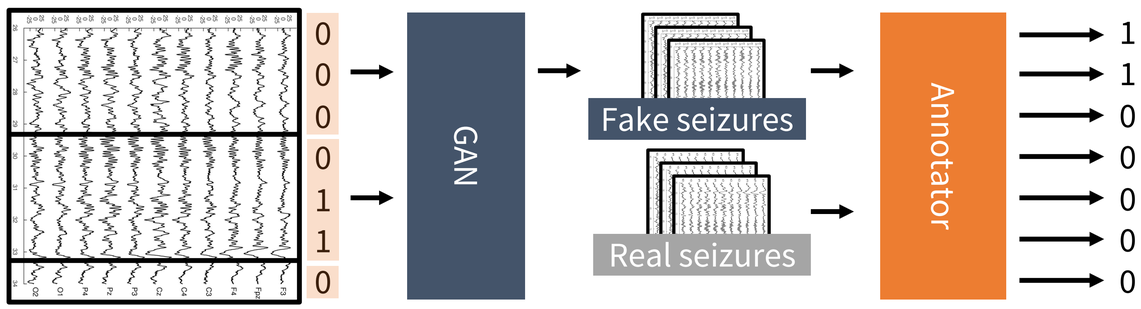


Dataset: Annotated EEGs of seizure patients [1]
Challenge: Annotate types of seizures per timestamp

2 Approach

Problem: Seizure events relatively rare in recordings
Our approach:

- Augment dataset with GAN-generated fake seizures (based on [2])
- Implement and train LSTM and CNN models with Keras
- Use baseline comparison model to test GAN

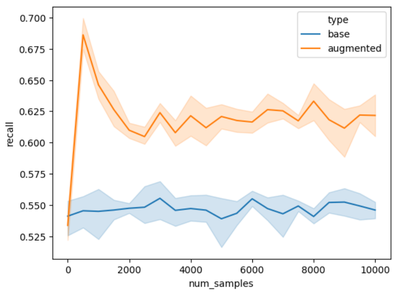


3 Research Questions

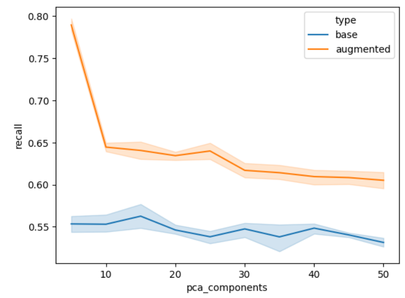
1. How do existing methods compare against each other, in terms of **precision** and **recall**?
2. Can GAN-based **data augmentation** improve performance?

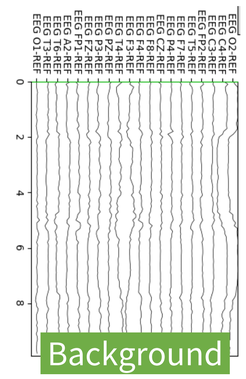
4 Experiments

1. Num. Fake Seizures

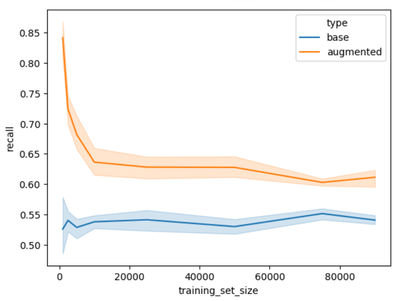


2. Num. PCA Components



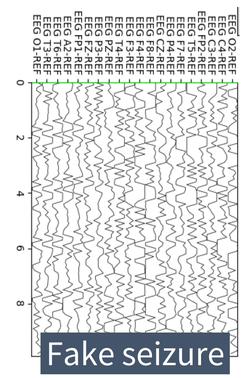


3. Num. Training Samples



Key insights

- Adding generated seizures improves performance
- Data augmentation especially strong with small training set



5 Discussion

- Risk of overfitting
- Single window size vs. variable-length seizures
- Generate different types of seizures (instead of binary)

6 Conclusion

- Annotating with raw data is challenging (2k+ features in window)
- Rarity of seizure events can be compensated with GANs
- Data augmentation improves performance significantly

[1] Harati et al. "The TUH EEG CORPUS: A big data resource for automated EEG interpretation." IEEE, 2014.
[2] Radford et al. "Unsupervised representation learning with deep convolutional generative adversarial networks." Preprint on arXiv:1511.06434, 2015.