Resources

442.022 Nonlinear Signal Processing, Practical
Summer Term 2021

Christian Toth, christian.toth@tugraz.at
Signal Processing and Speech Communication Laboratory
Graz University of Technology

Last updated: June 25, 2021

Overview

This document shall provide you with some references and resources that are relevant to this course and will help you to complete the assignments and deepen your understanding of certain topics. Of course you are not expected to read through all these materials - consider it rather as a buffet of knowledge where you can taste many different things from. I will update this document as the course progresses.

It should be understood that this by far is not an exhaustive list of (good) resources. In case you discover additional material that you think might be relevant to other participants, please let me know so that I can maybe include them in this list.

General Resources

Oppenheim and Schafer 2009 is a standard book on discrete time signal processing. The full text is provided online from the TU Graz Library and there are also many print versions available.

Zhang et al. 2020 is a fairly recent open-source book that provides a nice introduction to deep learning with many examples in PyTorch (which we also use in this course). It is definitely worth checking out!

Goodfellow, Bengio, and Courville 2016 is a very popular book on deep learning. Print versions available at TU Graz Library or online at www.deeplearningbook.org.

A comprehensive introduction to machine learning from a probabilistic perspective is provided by Murphy 2012. Again, some exemplars are available in the TU Graz library. The author currently prepares a new version Murphy 2021 with online draft freely available.

Normalizing Flows

Brubaker 2020 gives a nice introduction to normalizing flows based on their survey (Kobyzev, Prince, and Brubaker 2020). Papamakarios et al. 2019 provide another, more elaborate review of recent work on normalizing flows.

Convolutional Neural Networks

Personally, I think the NYU Deep Learning course LeCun, Canziani, and Goldstein 2020 is very comprehensible and well presented. For an introduction to CNNs I would recommend the following video lectures thereof.

- Canziani 2020a or a motivation of CNNs from the perspective of natural signal characteristics
- Canziani 2020b for an introduction to the linear algebra involved in CNNs
- Canziani 2020c for a very nice explanation of 1d and 2d convolutions and channels

WaveNet

Accompanying the paper (Oord, Dieleman, et al. 2016) there is a research blog post (Oord and Dieleman 2016) from DeepMind that gives an overview of what WaveNet does, including some sound samples generated with WaveNet. The idea of the fast generation algorithm is presented in (Paine et al. 2016). Note that already Kubin 1996 proposed a recurrent nonlinear architecture for speech synthesis based on radial basis functions

If you got interested in this topic, some recent works on generative models for speech/audio synthesis based on normalizing flows include (S. Kim et al. 2019; H. Kim et al. 2020; Prenger, Valle, and Catanzaro 2019). With the knowledge acquired in this course you should be able to get the gist of them.

References

Brubaker, Marcus (2020). *Introduction to Normalizing Flows (ECCV2020 Tutorial)*. URL: https://www.youtube.com/watch?v=u3vVyFVU_lI.

Canziani, Alfredo (2020a). *1D multi-channel convolution and autograd*. URL: https://www.youtube.com/watch?v=eEzCZnOFU1w.

— (2020b). *Listening to convolutions*. URL: https://www.youtube.com/watch?v=OrBEon3VlQg.

— (2020c). *Natural signals properties and CNNs*. URL: https://www.youtube.com/watch?v=kwPWpVverkw.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. The MIT Press. URL: https://www.deeplearningbook.org/.

Kim, Hyeongju et al. (2020). "WaveNODE: A Continuous Normalizing Flow for Speech Synthesis". In: arXiv: 2006.04598.

Kim, Sungwon et al. (2019). "{F}lo{W}ave{N}et : A Generative Flow for Raw Audio". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, pp. 3370–3378.

Kobyzev, Ivan, Simon Prince, and Marcus Brubaker (2020). "Normalizing Flows: An Introduction and Review of Current Methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1. DOI: 10.1109/TPAMI.2020.2992934.

Kubin, G. (1996). "Synthesis and coding of continuous speech with the nonlinear oscillator model". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 1. IEEE, pp. 267–270. DOI: 10.1109/ICASSP.1996.540409.

LeCun, Yann, Alfredo Canziani, and Mark Goldstein (2020). *Deep Learning (Course from the NYU Center for Data Science)*. URL: https://atcold.github.io/pytorch-Deep-Learning/.

Murphy, Kevin P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

— (2021). *Probabilistic Machine Learning: An introduction*. MIT Press. URL: probml.ai.

Oord, Aaron van den and Sander Dieleman (2016). *Introduction to Normalizing Flows (ECCV2020 Tutorial)*. URL: https://deepmind.com/blog/article/wavenet-generative-model-raw-audio.

Oord, Aaron van den, Sander Dieleman, et al. (2016). "WaveNet: A Generative Model for Raw Audio". In: arXiv: 1609.03499.

Oppenheim, Alan V. and Ronald W. Schafer (2009). *Discrete-Time Signal Processing*. 3rd. Prentice Hall Press.

Paine, Tom Le et al. (2016). "Fast Wavenet Generation Algorithm". In: arXiv: 1611.09482.

Papamakarios, George et al. (Dec. 2019). *Normalizing Flows for Probabilistic Modeling and Inference*. arXiv: 1912.02762.

Prenger, Ryan, Rafael Valle, and Bryan Catanzaro (2019). "Waveglow: A Flow-based Generative Network for Speech Synthesis". In: *ICASSP 2019 - 2019 IEEE International Con-*

ference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 3617–3621. DOI: 10.1109/ICASSP.2019.8683143.
Zhang, Aston et al. (2020). *Dive into Deep Learning*. https://d2l.ai.