

Seminars

proposed topics:

1. understanding deep neural networks
2. overview on other ML techniques
3. natural language processing
4. advanced computer vision

time limit: 20 minutes + 10-minute discussion

Seminars: understanding deep neural networks

1. Techniques that allow understand more on how and why deep architectures work

- <https://distill.pub/2020/circuits/early-vision/>
- <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>
- <https://distill.pub/2017/feature-visualization/>
- [D. Erhan, Y. Bengio, A. Courville, P. Vincent. , Visualizing higher-layer features of a deep network \(2009\).](#)
- <http://arxiv.org/pdf/1506.02753.pdf>
- <https://arxiv.org/pdf/1312.6034v2.pdf>
- <https://github.com/tensorflow/lucid>

2. Fooling deep neural networks: deep learning models are highly sensitive to carefully prepared adversarial attacks

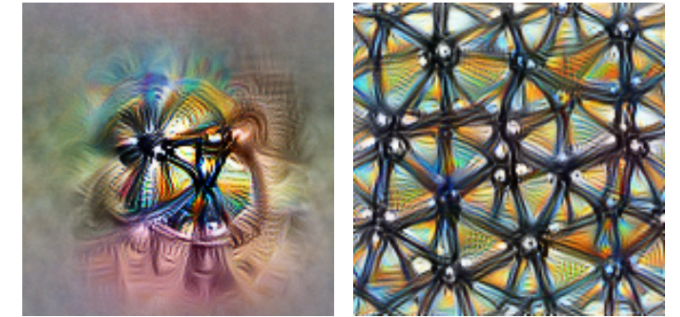
How to generate adversarial examples? Can we use them to improve network stability?

- Szegedy, Christian, et al. “Intriguing properties of neural networks” *arXiv:1312.6199* (2013).
- Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples” *arXiv:1412.6572* (2014).
- Papernot, Nicolas, et al. “The limitations of deep learning in adversarial settings.” *2016 IEEE European symposium on security and privacy (EuroS&P)*. IEEE (2016).
- Su, Jiawei, Danilo V. Vargas, and Kouichi Sakurai. “One pixel attack for fooling deep neural networks.” *IEEE Transactions on Evolutionary Computation* 23.5 828 (2019).

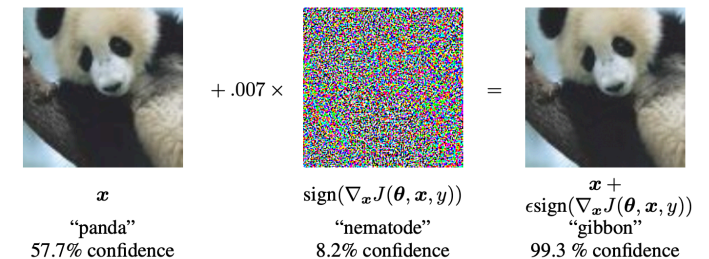
3. Defense strategies against adversarial attacks

Overview on various techniques for defending against adversarial examples for attacking deep neural networks.

- Szegedy, Christian, et al. “Intriguing properties of neural networks” *arXiv:1312.6199* (2013).
- Xie, Cihang, et al. “Mitigating adversarial effects through randomization.” *arXiv:1711.01991* (2017).
- Das, Nilaksh, et al. “Keeping the bad guys out: Protecting and vaccinating deep learning with jpeg compression.” *arXiv:1705.02900* (2017).
- Xie, Cihang, et al. “Feature denoising for improving adversarial robustness.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019).
- Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples” *arXiv:1412.6572* (2014).



Feature visualization answers questions about what a network—or parts of a network—are looking for by generating examples.



Seminars: overview on other ML techniques

1. Reinforcement Learning

Basics and applications

- <https://arxiv.org/pdf/cs/9605103.pdf>
- <https://mpatacchiola.github.io/blog/2016/12/09/dissecting-reinforcement-learning.html>
- <https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/>

2. Hidden Markov model

Definition and example applications

- <https://jonathan-hui.medium.com/machine-learning-hidden-markov-model-hmm-31660d217a61>
- <http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf>

3. Recommender systems

Idea, applications and the Netflix Price

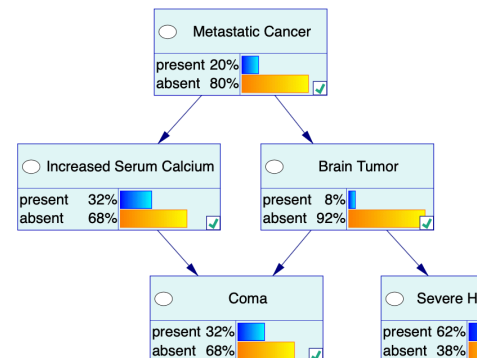
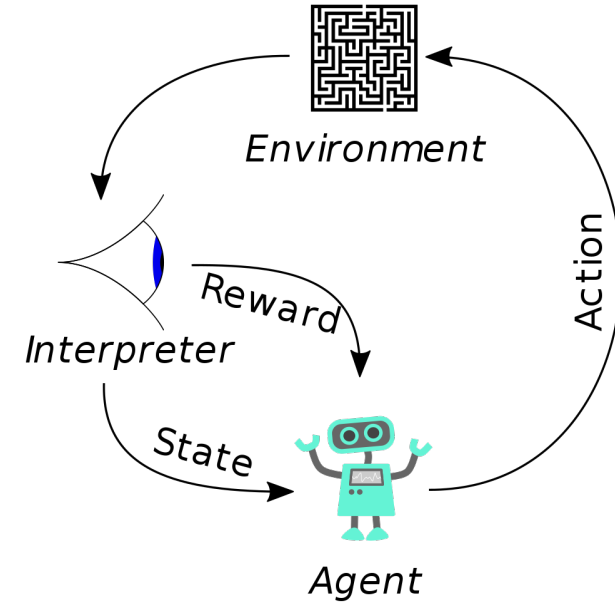
- <https://arxiv.org/pdf/1203.4487.pdf>
- <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>

4. Bayesian networks

- <https://repo.bayesfusion.com/bayesbox.html>
- <http://www.eng.tau.ac.il/~bengal/BN.pdf>
- <http://www.niedermayer.ca/node/35>

5. Evolutionary optimization methods

- differential evolution
- particle swarm optimization



Seminars: natural language processing, NLP

1. Optical character recognition (OCR)

- problem definition
- classical or neural network approach: an overview
- presenting one particular engine, e.g. state-of-the-art *Tesseract* (what deep learning model is used inside?)
- <https://tesseract-ocr.github.io/tessdoc/>

2. Automatic speech recognition (ASR)

- problem definition
- possible applications
- databases (e.g. *Librispeech*)
- an overview on single specific algorithm, e.g. *ContextNet* or *Deep Speech*
- <https://arxiv.org/pdf/2005.03191.pdf>
- <https://arxiv.org/pdf/1512.02595.pdf>

3. Machine Translation

- problem definition, example methods

4. Question Answering

- problem definition, example methods

5. BERT natural language model

- overview, importance

Seminars: computer vision

1. Point Feature Matching

- problem definition
- classical and deep learning approaches
- *SuperGlue* algorithm: <https://arxiv.org/abs/1911.11763>

2. Pose estimation

- problem definition and deep-learning solution, e.g.:
- <https://arxiv.org/pdf/1803.08225.pdf>

3. Super-Resolution

- idea, applications, algorithms
- <https://arxiv.org/abs/1809.00219>

4. Any other problem with machine learning solution