Machine learning for inverse problems, learning from noisy data, and DNA storage

Reinhard Heckel

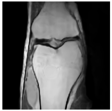
reinhard.heckel@tum.de

Research

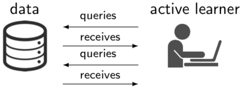
Machine learning, statistics, and signal processing.

Current focus:

i. Deep learning for inverse problems



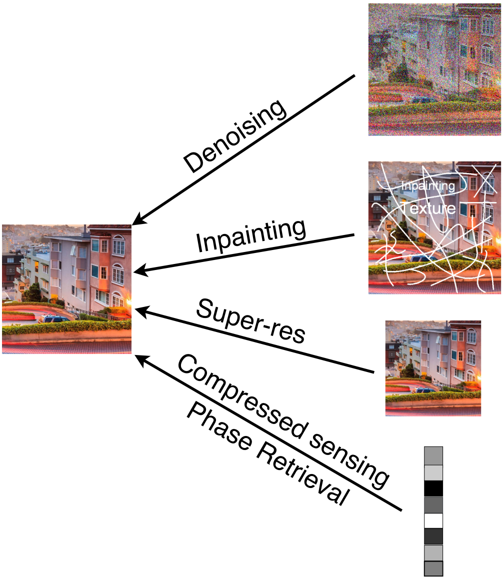
ii. Learning from unlabeled data or noisy labels



iii. DNA data storage

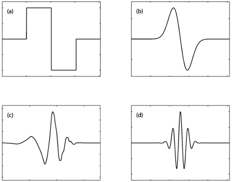
i. Deep learning for inverse problems

Example of inverse problems



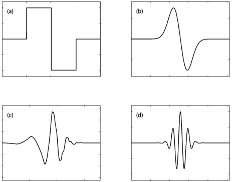
Inverse problems

Traditionally solved with handcrafted models like wavelets/sparsity

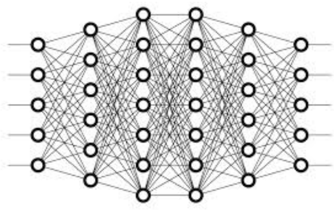


Inverse problems

Traditionally solved with handcrafted models like wavelets/sparsity

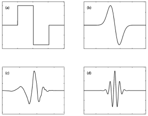


Now state-of-the-art based on image-generating deep neural networks



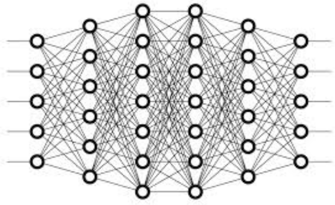
The deep decoder

handcrafted

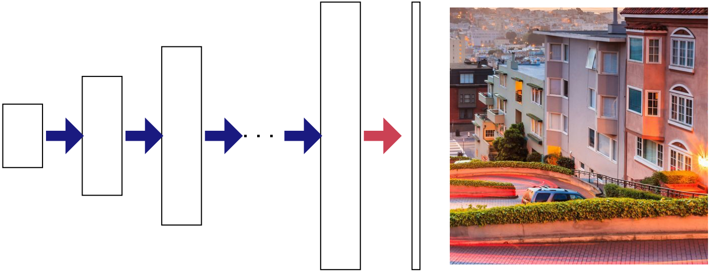


+

deep neural networks



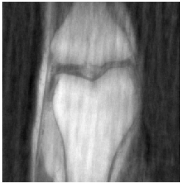
The deep decoder: handcrafted neural network



An image generating network that is

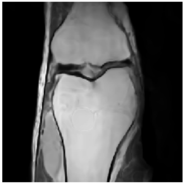
- not trained
- yields state-of-the-art compression and image restoration performance, for example for MRI imaging
- is underparameterized

Deep decoder for MRI



LS
25.82dB

L1-Wav
29.04dB



DD
30.08dB

ii. Learning from unlabeled data or noisy labels

Learning from examples

1. Collect candidate examples for example via google image search
2. Labeling the candidate images
3. Training a deep network

An AI company - what do these people do?

An AI company - what do these people do?

Yan Cong for The NYT. “Workers at the headquarters of Ruijin Technology Company. They identify objects in images to help artificial intelligence make sense of the world.”

Labeling is the most expensive step

Labeling is the most expensive step

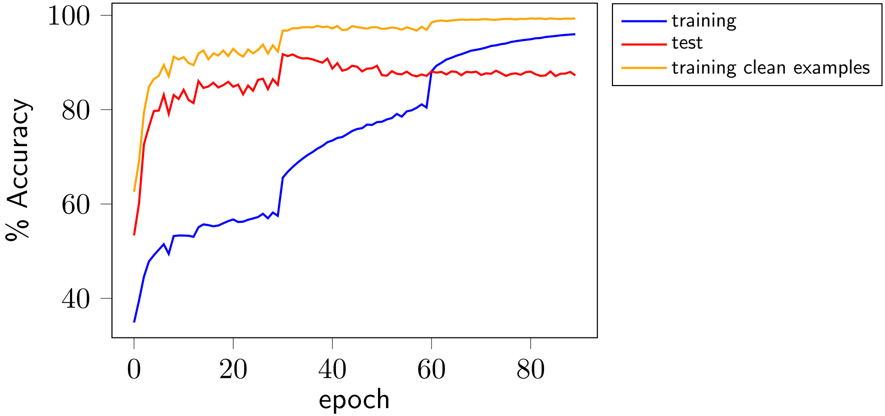
Labeling is the most expensive step

- How to use the human workers most efficiently with active learning?

Labeling is the most expensive step

- How to use the human workers most efficiently with active learning?
- Don't label - learn from noisy candidates

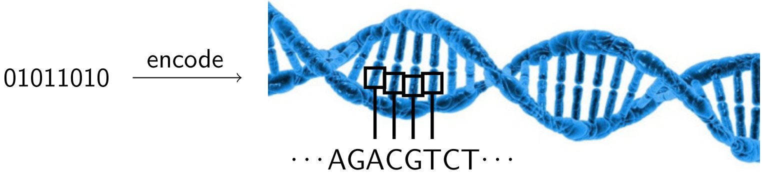
Training a deep network on noisy candidates



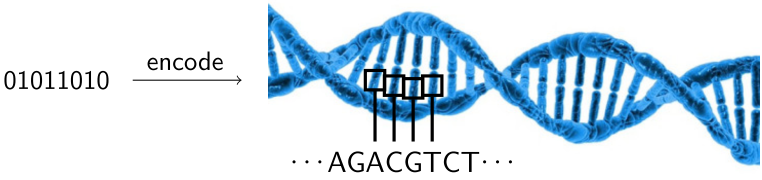
- Deep nets fit correct examples faster than wrong ones
- Early stopping enables training on noisy examples!

ii. DNA data storage

DNA data storage



DNA data storage



Leads to interesting coding/clustering/reconstructions problems


A Journal of the Gesellschaft Deutscher Chemiker
Angewandte
GDCh
International Edition
Chemie
www.angewandte.org
2015–54/8
Oil/Water Separation
Review by S. Seeger et al.
ACIEFS 54 (8) 2293–2562 (2015) · ISSN 1433-7851 · Vol. 54 · No. 8
WILEY-VCH

A commercial application: Storing information for eternity?

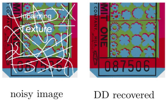
Our first customer: Massive attack


MATT BLACK
MEZZANINE
DNA
CONTAINS
1 million copies of the
Mezzanine album encoded[1]
in 901'065 DNA sequences,
each 146 basepairs long
and encapsulated in silicone
particles for long-term
storage stability[2]
[1] Grass et al. Angew. Chem. Int. Ed. 54, 2552 (2015).
[2] Paunescu et al. Nat. Protoc. 8, 2440 (2013).

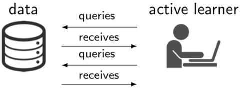
Research focus

Machine learning, statistics, and signal processing.

i. Deep learning for inverse problems



ii. Learning from few or noisy labels



iii. DNA data storage

Thank you!