# Deep Learning Antipatterns

**Agoston Torok**
***Advisor, Neural Machines***

Budapest ML Forum

May 3-4, 2022

# whoami

PhD - Virtual Reality, Neuroscience, Sensor Fusion

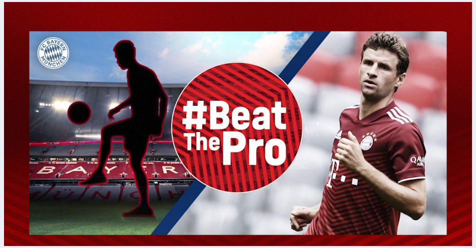
Research - RCNS, SZTAKI, UCL, Technion, UT Austin etc.

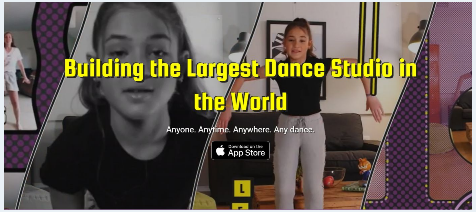
Data Science Lead - Synetiq (ad attribution)

Director Data Science - AGT International (IoT for sports)

Director Data Science - Coach-AI (mobile computer vision)

Advisor, speaker, blogger - Neural Machines




TDDCOMICS

Neural Machines Ltd.

# Design Patterns for Data Science

**Design patterns** are typical solutions to commonly occurring problems in software design (e.g. adapter, facade, proxy)

An **AntiPattern** is a literary form that describes a commonly occurring solution to a problem that generates decidedly negative consequences.

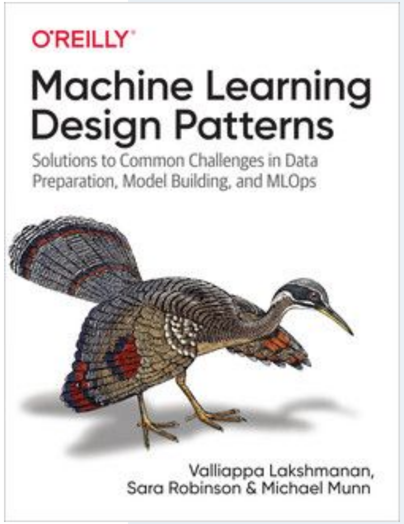
Shared:

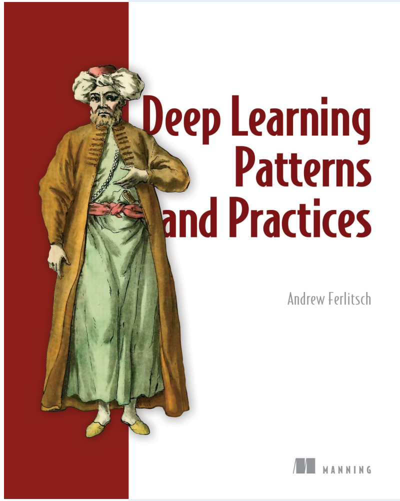
- Good intent
- Solution-oriented
- “Agile”

Different:

- Craftsmanship
- Effect

https://refactoring.guru/design-patterns/what-is-pattern
https://sourcemaking.com/antipatterns
https://towardsdatascience.com/design-patterns-in-machine-learning-b73eea4882cd




chines Ltd.

# Why specific to deep learning?

- Performance gains on big, unstructured data
  - Model and data parallelism
- End-to-end learning
  - No feature extraction
- Transfer learning
  - Domain adaptation
- Focus on loss functions
  - e.g. style, content, adversarial, perceptual
- Booming research
  - e.g. self-supervised learning

# What is specific to deep learning?

- Performance gains on big, unstructured data
    - Model and data parallelism
- **End-to-end learning**
    - No feature extraction
- **Transfer learning**
    - Domain adaptation
- Focus on loss functions
    - e.g. style, content, adversarial, perceptual
- Booming research
    - e.g. self-supervised learning

# The ultimate decision

I can
design the right
**architecture**!

I can deliver
**faster**!

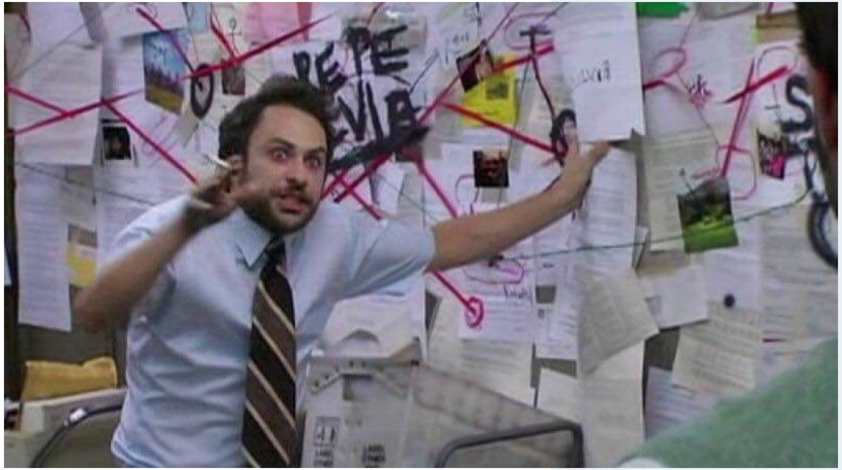
**End-to-end learning**

**Transfer learning**

# Why are these antipatterns?

Well not both of them…

At least, not at the same time…



# The effect of these practices

## End-to-end learning

- Time spent on architecture design, loss functions, data ingestion, training

- Ideal outcome is a fast model

- Risk is that we end up with a poor model because of lack of data, expertise, time

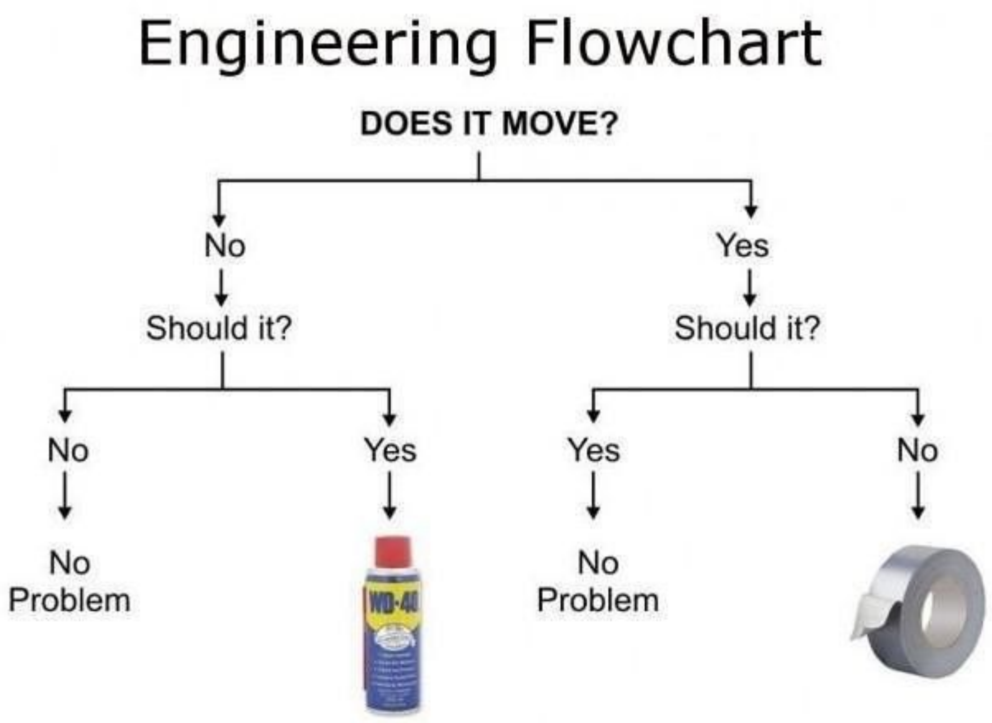
## Transfer learning

- Time spent on research, trying, evaluation, benchmarking

- Ideal outcome is a model fast

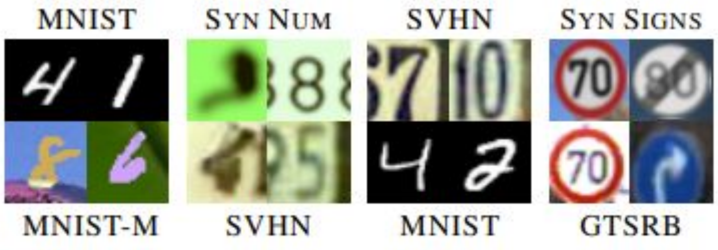
- Risk is that we end up with a poor model because of too large difference in domains/tasks

## Antipattern emerges when a good solution is applied to the wrong problem



# Definitive questions

Do you have enough data in the target domain?




Do you have enough labels for the specific task?

Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180-1189). PMLR.

# What’s big enough data?

Heuristics

- 10x the degrees of freedom
- 1000 examples per class

Complexity-based decision

- Linear vs. non-linear
- Structural risk minimization, AIC, BIC
- PAC learning

# What does the required sample size depend on?

- Representational power of the model
- Complexity of the problem
- Probability of successful training
- Accuracy of the prediction
- Way the data is presented
- Way the data is selected

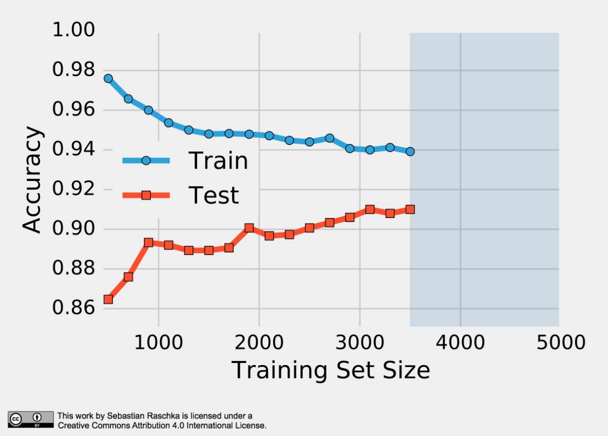
$$m \geq \frac{1}{\varepsilon}(ln \mid H \mid + ln\frac{1}{\delta})$$

$$m \geq \frac{1}{\varepsilon}(8\, VC(H) log_2 \frac{13}{\varepsilon} + 4 log_2 \frac{2}{\delta})$$

Computational Learning Theory by Littman & Isbell, Udacity

# Then how come we still can train NN?

- It depends on “only” model complexity, error, failure rate
  - Nothing about data selection, presentation
  - Nothing about inductive bias (CNN vs FCN)
- Connection between structure and *n*
  - Structure optimization for MNIST
  - 500 samples → 91.13% accuracy
- Learning curve analysis
  - Dataset size
  - Epoch





D’souza, R. N., Huang, P. Y., & Yeh, F. C. (2020). Structural analysis and optimization of convolutional neural networks with a small sample size. *Scientific reports*, *10*(1), 1-13.

Du, S. S., Wang, Y., Zhai, X., Balakrishnan, S., Salakhutdinov, R. R., & Singh, A. (2018). How many samples are needed to estimate a convolutional neural network?. *Advances in Neural Information Processing Systems*, *31*.

# When can you do transfer learning?

Anytime…

Inductive transfer learning

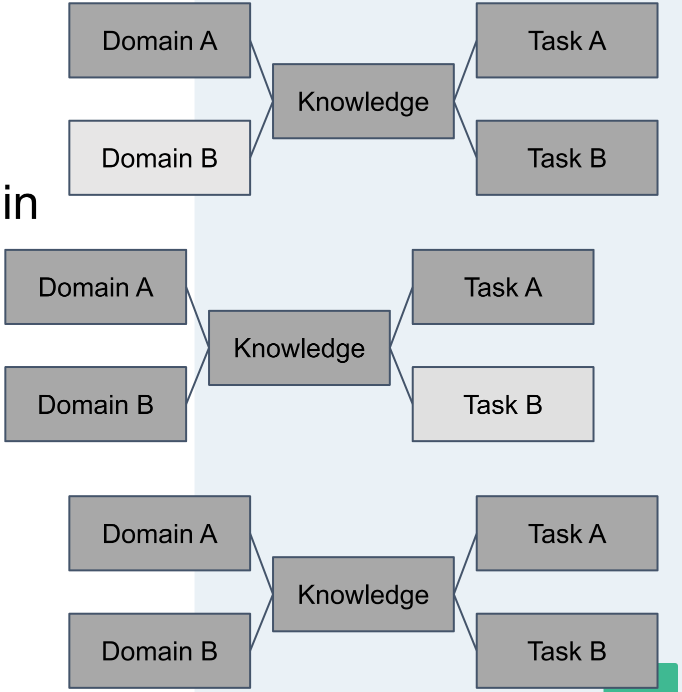
- E.g. ResNet, VGG19 as a backbone
- Learning the semantics/structure of the domain

Transductive transfer learning

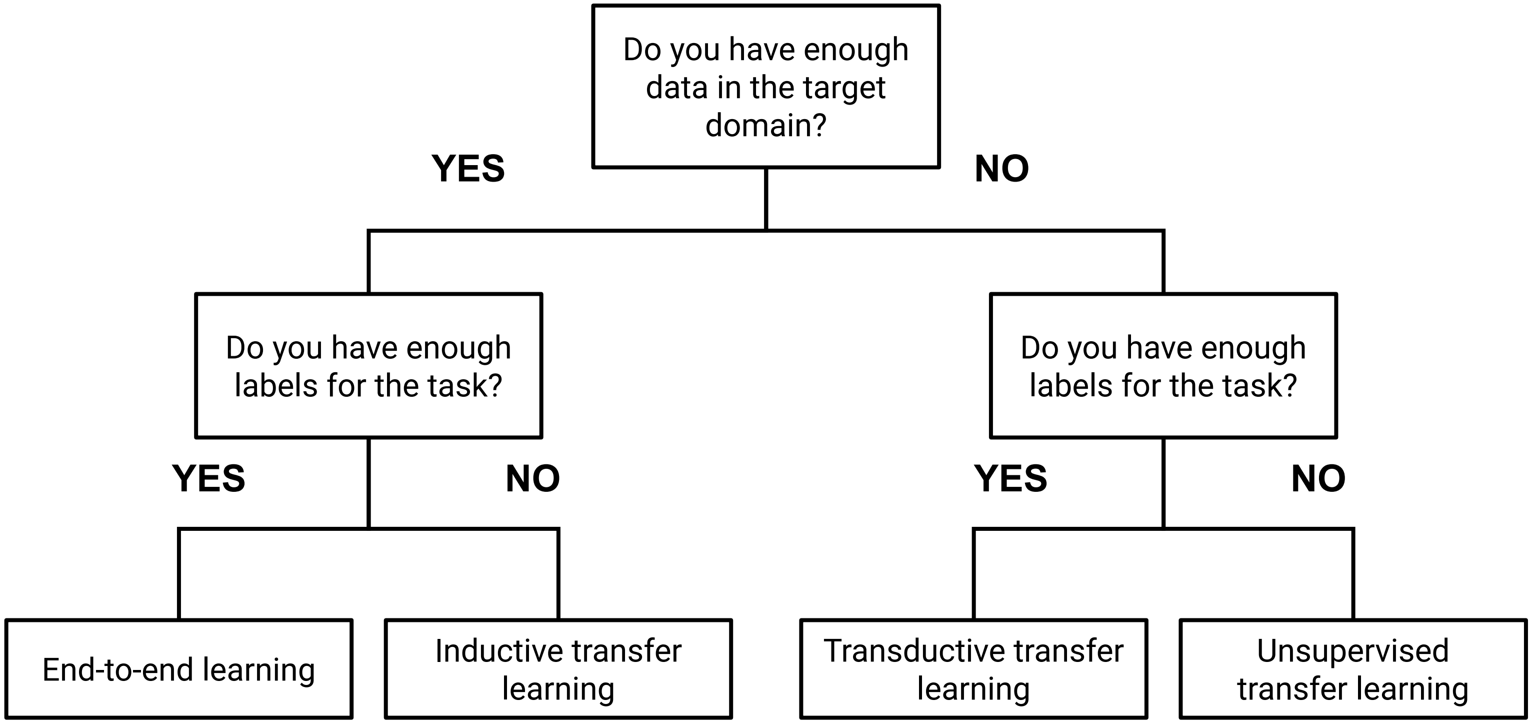
- Domain adaptation, covariance shift
- Learning the semantics/structure of the task

Unsupervised transfer learning

- Knowledge bootstrapping



# From duck tape to deep learning



# Thank you for your attention!

**Neural Machines UK**

hello@neuralmachines.co.uk
+44 77 8381 96 09
www.neuralmachines.co.uk