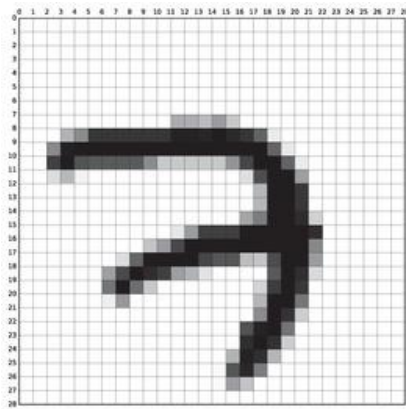


# Application of the Lottery Ticket Hypothesis in NLP and Early Pruning



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Intermission



(a) MNIST sample belonging to the digit '7'.



(b) 100 samples from the MNIST training set.

Source:

[https://www.mdpi.com/applsci/applsci-09-03169/article\\_deploy/html/images/applsci-09-03169-g001-550.jpg](https://www.mdpi.com/applsci/applsci-09-03169/article_deploy/html/images/applsci-09-03169-g001-550.jpg)



Source:

[https://www.bonaccorso.eu/wp-content/uploads/2016/07/28019400581\\_e1eb13ccc8\\_b.jpg](https://www.bonaccorso.eu/wp-content/uploads/2016/07/28019400581_e1eb13ccc8_b.jpg)

# Structure



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Introduction

Motivation

Related Work

Task Definition

Progress

Remaining Work

# Structure



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Introduction

Motivation

Related Work

Context of the thesis

Task Definition

Progress

Remaining Work

Content of the thesis

Int

## Initial Thoughts

Mot

- Many good reasons to initialize & train neural networks with many parameters

Rel

- Empirical evidence that many networks can be reduced after training while maintaining performance
  - aka. "Pruning"
  - Desirable due to bias towards small models (Ockham's razor)

Task

Pro

Rem

Int

## Initial Thoughts

Mot

- Good reasons to initialize & train neural networks with many parameters

Rel

- Empirical evidence that many networks can be reduced after training while maintaining performance
  - aka. "Pruning"
  - Desirable due to bias towards small models (Ockham's razor)

Task

Pro

Rem

- Main Question:
  - "How important are the pruned weights during training?"

Int

## Lottery Ticket Hypothesis

Mot

- Sheer number of subnetworks results in subnetworks with favorable initialization

Rel

Task

Pro

Rem

Int

## Lottery Ticket Hypothesis

Mot

- Sheer number of subnetworks results in subnetworks with favorable initialization

Rel

- Train a subnetwork with initial parameters
  - Similar performance ==> "lottery ticket"

Task

Pro

Rem

Int

## Lottery Ticket Hypothesis

Mot

- Sheer number of subnetworks results in subnetworks with favorable initialization

Rel

- Train a subnetwork with initial parameters
  - Similar performance ==> "lottery ticket"

Task

- Extraction of "lottery-ticket" after the full network is trained
  - Pruning weights based on magnitude finds a lottery ticket

Pro

Rem



# Motivation



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Int

Mot

Rel

Task

Pro

Rem

Int

Mot

Rel

## Time & Memory

- Speedup during execution just as regular pruning
  - But remarkable compression rate: up to ~50x
- Decrease in memory usage during execution
- Possible speedup during development
  - There might be a way to identify lottery tickets early

Task

Pro

Rem

Int

Mot

Rel

Task

Pro

Rem

## Time & Memory

- Speedup during execution just as regular pruning
  - But remarkable compression rate: up to ~50x
- Decrease in memory usage during execution
- Possible speedup during development
  - There might be a way to identify lottery tickets early

## Theory of Neural Networks

- "Lottery-tickets" contain weights necessary for training
- Identification of "lottery-tickets" might explain importance of weights

# Related Work



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Int

Mot

Rel

Task

Pro

Rem

Int

Mot

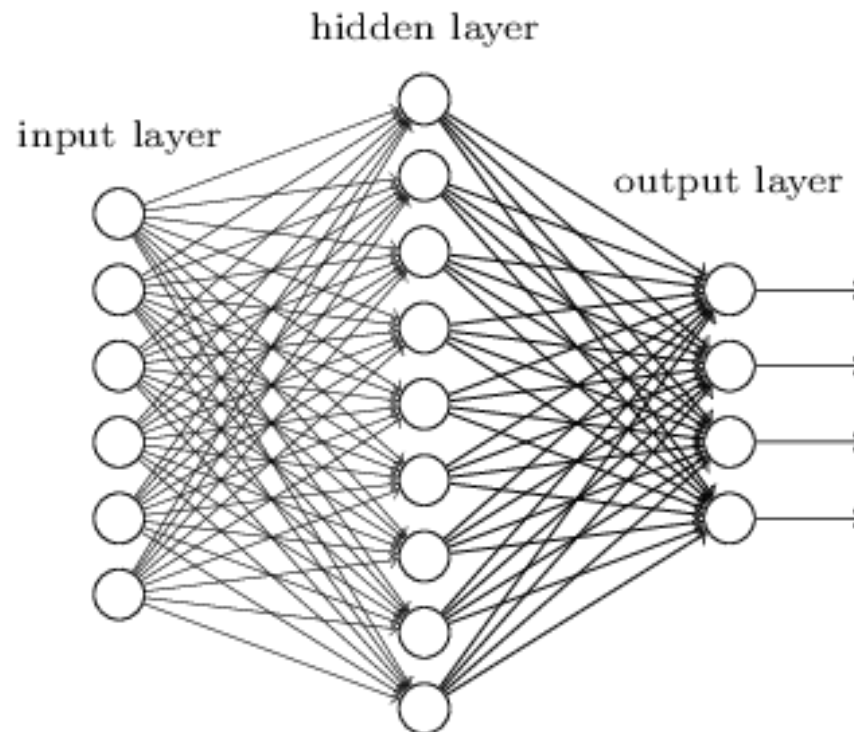
Rel

Task

Pro

Rem

## Fully Connected Neural Network



Source:  
[https://hackernoon.com/hn-images/1\\*Kdnux0Kw1yQ4D8dq\\_\\_mYCA.png](https://hackernoon.com/hn-images/1*Kdnux0Kw1yQ4D8dq__mYCA.png)

# Related Work – Background

Int

Mot

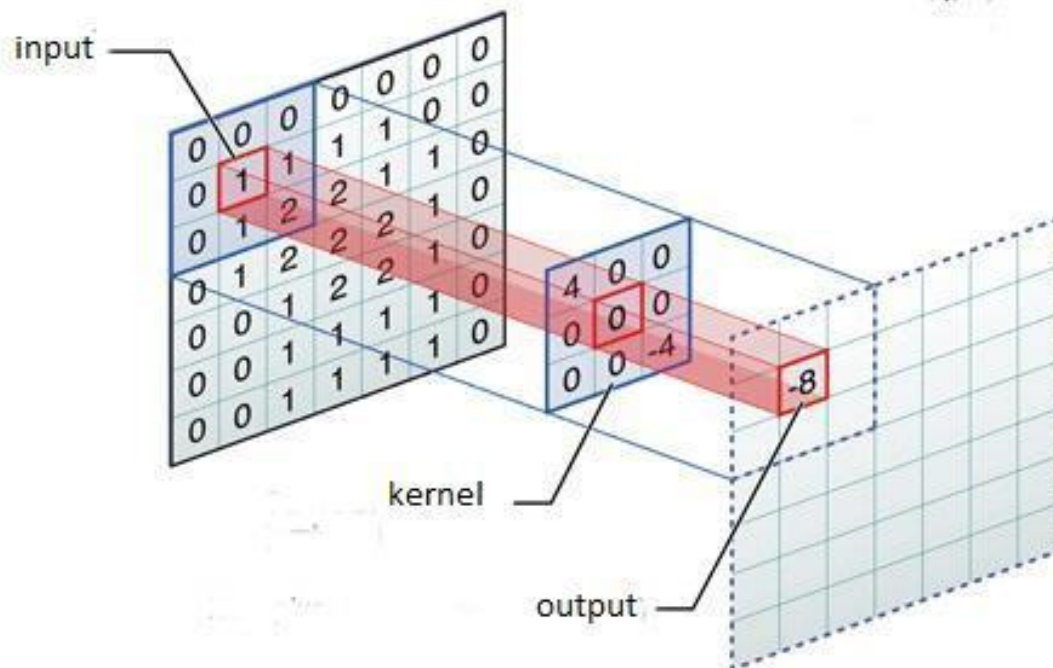
Rel

Task

Pro

Rem

## Convolution in Neural Networks



Source:

[https://miro.medium.com/proxy/0\\*dRD6PhKOnnClhz15.jpg](https://miro.medium.com/proxy/0*dRD6PhKOnnClhz15.jpg)

# Related Work – Background



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Int

Mot

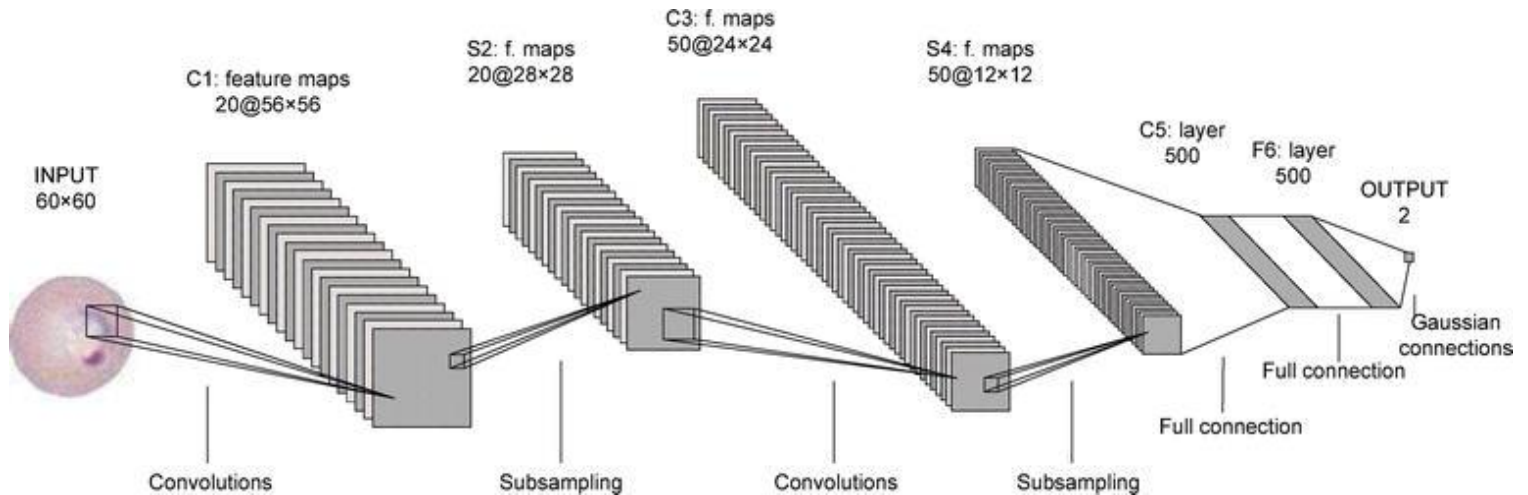
Rel

Task

Pro

Rem

## Convolutional Neural Network Architecture (Lenet-5)



Source:

<https://api.intechopen.com/media/chapter/58989/media/F4.png>

# Related Work – Background



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Int

Mot

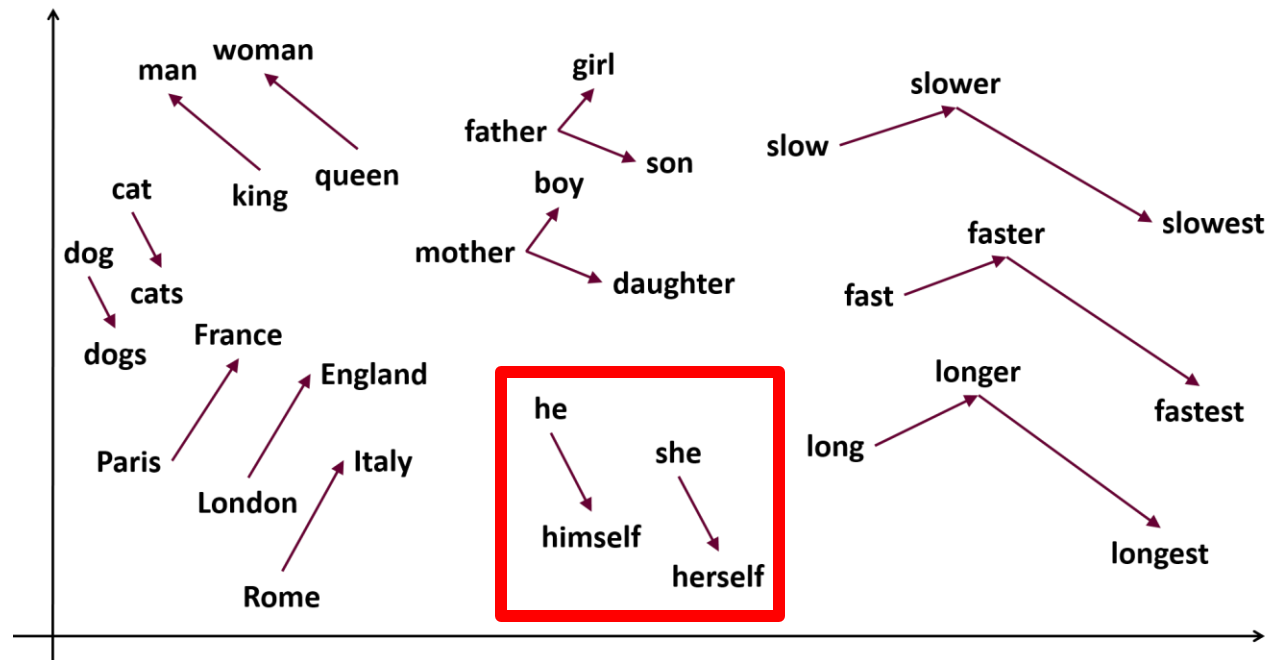
Rel

Task

Pro

Rem

## Language Models



Source:  
<https://samyzaf.com/ML/nlp/w ord2vec2.png>



# Related Work – CNN in NLP

Int

Mot

Rel

Task

Pro

Rem

## "Convolutional Neural Networks for Sentence Classification"

- 2014
- Task:
  - Varying Classifications
- Datasets:
  - Movie reviews
  - SST-1, SST-2
  - Subjectivity dataset
  - TREC question dataset
  - Customer reviews
  - MPQA

Int

Mot

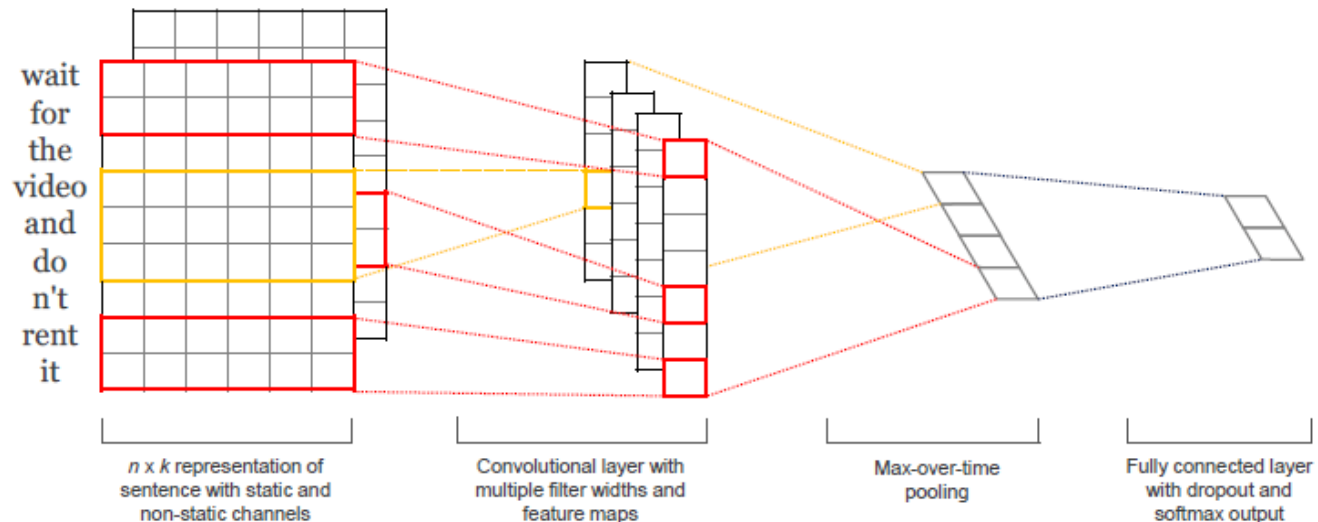
Rel

Task

Pro

Rem

## "Convolutional Neural Networks for Sentence Classification"



Source:  
"Convolutional Neural Networks for Sentence Classification"  
Figure 1

# Related Work - Pruning

Int

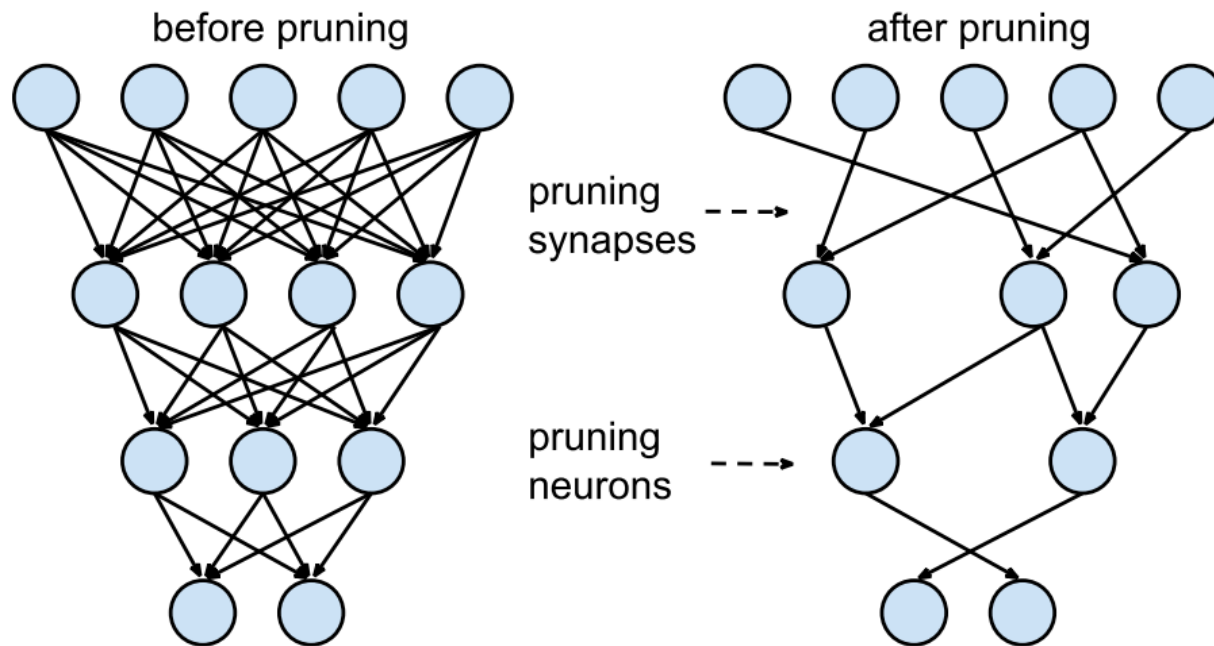
Mot

Rel

Task

Pro

Rem



Source:  
[https://www.mdpi.com/applsci/applsci-09-03169/article\\_deploy/html/images/applsci-09-03169-g001-550.jpg](https://www.mdpi.com/applsci/applsci-09-03169/article_deploy/html/images/applsci-09-03169-g001-550.jpg)

"Learning both Weights and Connections for Efficient Neural Networks"  
Figure.3

Int

Mot

Rel

Task

Pro

Rem

## "Learning both Weights and Connections for efficient Neural Networks"

- 2015
- Task:
  - Image Classification (ImageNet)
- Architectures:
  - LeNet (300-100-FC, 5-CNN)
  - AlexNet
  - VGG-16
- Compression:
  - 9x to 13x

Int

Mot

Rel

Task

Pro

Rem

## "ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression"

- 2017
- Task:
  - Image Classification (ImageNet)
- Architectures:
  - VGG-16
  - ResNet-50
- Compression:
  - Up to ~17x

Int

Mot

Rel

Task

Pro

Rem

## "The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks"

- 2019
- Task:
  - Image Classification (MNIST)
- Architectures:
  - Lenet-FCN (300-100-FCN)
  - Simple CNN (Conv-2, Conv-4, Conv-6)
  - VGG-19
  - ResNet-18
- Compression: ~20x to ~50x

# Related Work – Early Pruning

Int

Mot

Rel

Task

Pro

Rem

## "Really should we pruning after model be totally trained? Pruning based on a small amount of training"

- 2019
- Task:
  - Image Classification (MNIST, CIFAR-10)
- Architectures:
  - Unspecified CNN
  - VGG-19
- Compression --- Training Speed-Up:
  - ~10x --- 10x

Int

Mot

Rel

Task

Pro

Rem

## "Rethinking the Value of Network Pruning"

- 2018
- Observation:
  - Randomizing weights does not worsen a pruned network
- Weights are not essential to the quality of pruned network
- Pruning at its core is about finding suitable network architectures



Int

## "Network Architecture Search: A Survey"

- 2019
- **Categorization of NAS-Algorithms:**
  - Search Space:
    - Space of possible architectures
  - Search Strategy:
    - Space of possible architectures
  - Performance Estimation Strategy:
    - Space of possible architectures

Mot

Rel

Task

Pro

Rem

Int

Mot

Rel

Task

Pro

Rem

## "Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask"

- 2019
- Alteration of the Search Strategy (based on Magnitude):
  - large final (original strategy)
  - small final
  - large initial
  - small initial
  - large init & large final
  - small initial & small final
  - magnitude increase
  - movement
  - random (baseline strategy)

Int

Mot

Rel

Task

Pro

Rem

## "Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask"

- 2019
- Alteration of the Search Strategy (based on Magnitude):
  - large final (original strategy)
  - small final
  - large initial
  - small initial
  - large init & large final
  - small initial & small final
  - magnitude increase
  - movement
  - random (baseline strategy)

Int

Mot

Rel

Task

Pro

Rem

## "Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask"

- 2019
- Alteration of the Search Strategy (based on Magnitude):
  - large final (original strategy) → Large final & same sign
  - small final
  - large initial
  - small initial
  - large init & large final
  - small initial & small final
  - magnitude increase
  - movement
  - random (baseline strategy)

# Structure



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Introduction

Motivation

Related Work

Context of the thesis

Task Definition

Progress

Remaining Work

Content of the thesis

Int

Mot

Rel

Task

Pro

Rem

## Reproduction

- No source-code available
  - ➔ Produce own source-code
- Verify source-code by running experiments from the paper
  - Lenet-FCN
  - CNN-4
  - VGG-18

Int

Mot

Rel

Task

Pro

Rem

## Transfer to NLP

- Original context for the paper
  - Task: Image Classification
  - Dataset: "MNIST"
  - Model: Varying FCN and CNN
- Find comparable context in NLP
  - Task: Topic Classification
  - Dataset: "Reuters-21578"
  - Model: TBD
- Check if the Lottery-Ticket-Hypothesis holds

Int

Mot

Rel

Task

Pro

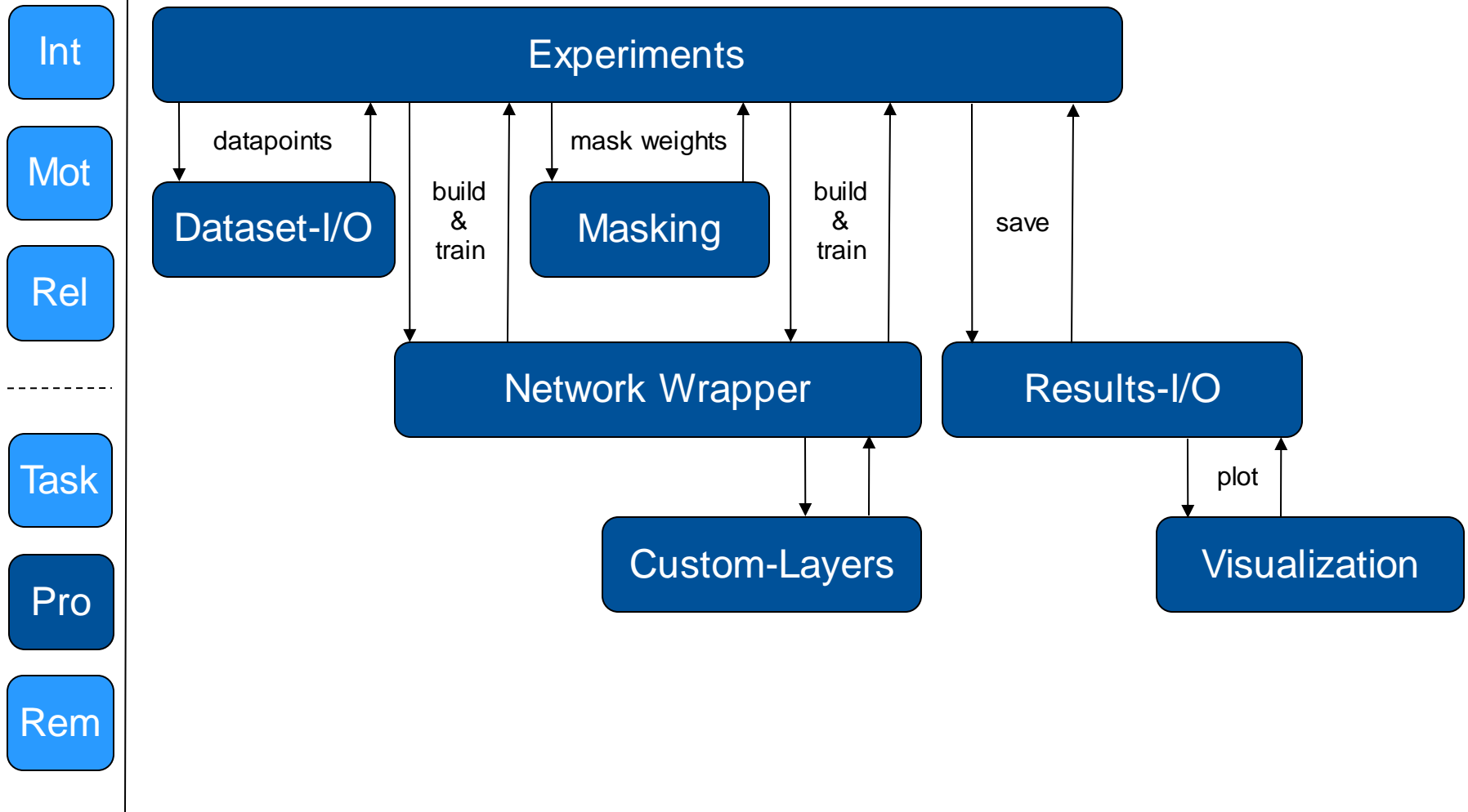
Rem

## Early Retrieval of Lottery Tickets

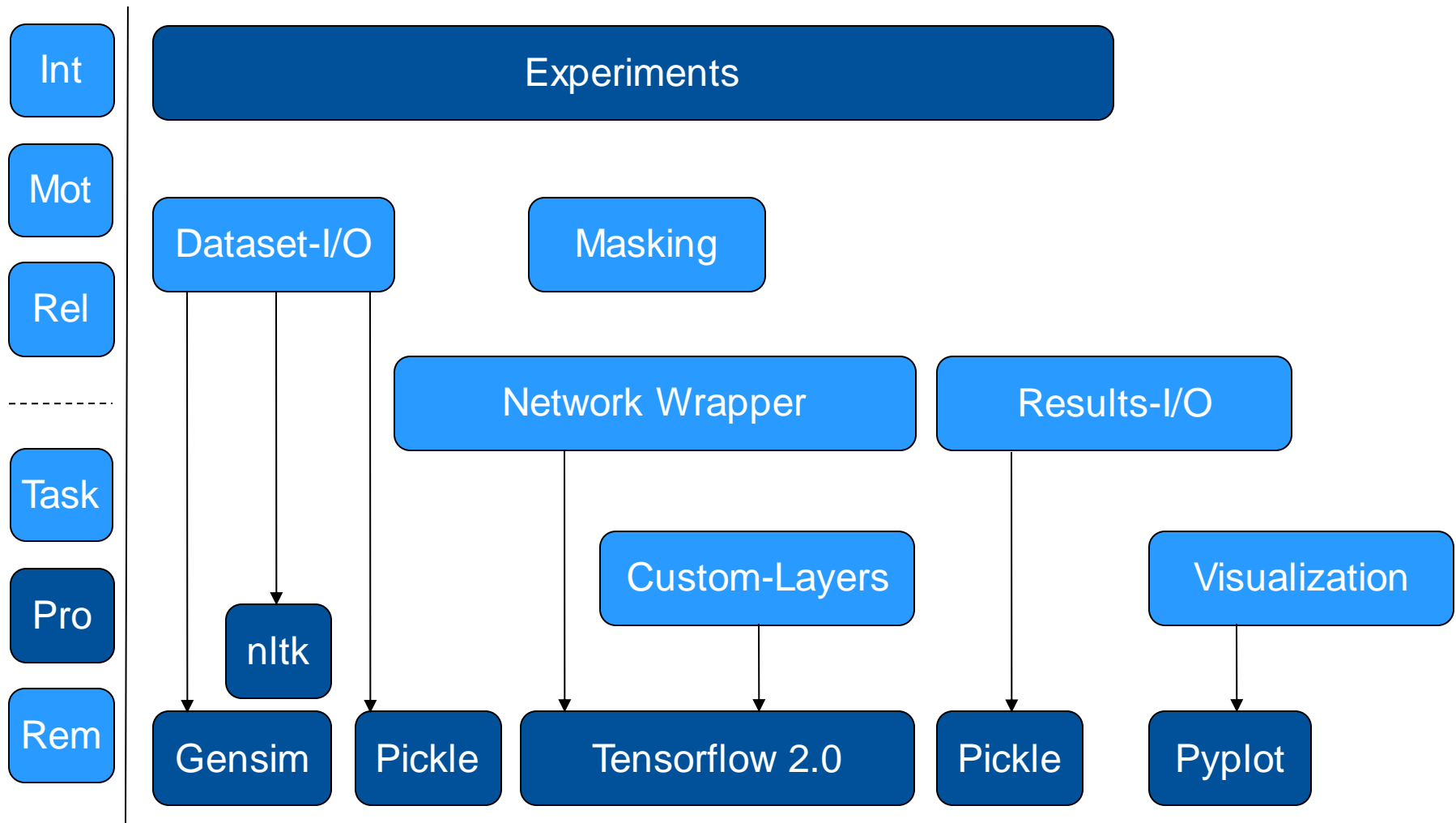
- Original method
  - Keep all weights with large final weights
  - Reset weights to original initial value
  - Retrain network
  - Repeat (Optional)
- Adaptation
  - "Select" weights earlier ~ develop early stopping criteria
  - Keep weights based on other metrics (Optional)



# Progress – Python-project



# Progress – Backend



Int

Mot

Rel

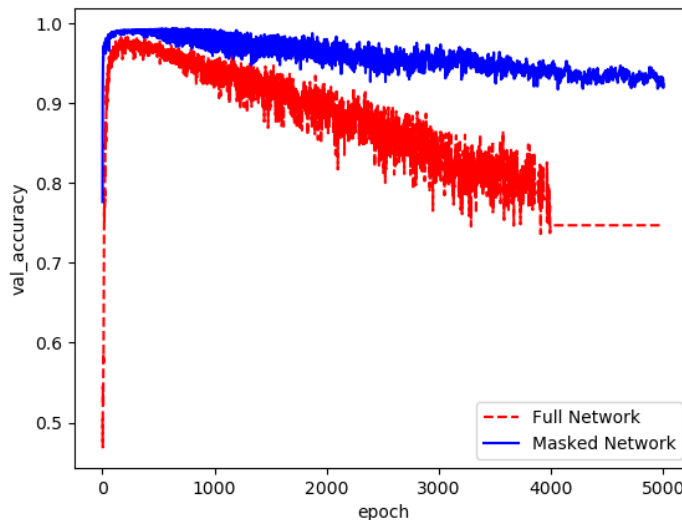
Task

Pro

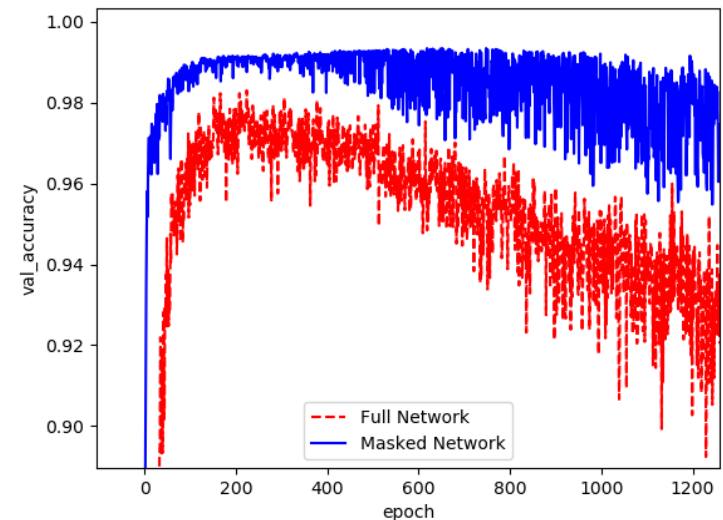
Rem

## Lenet-FCN-MNIST

- Validation-Accuracy
  - 20% pruned weights



Source:  
Produced by the author



Source:  
Produced by the author

# Progress – Experiments

Int

Mot

Rel

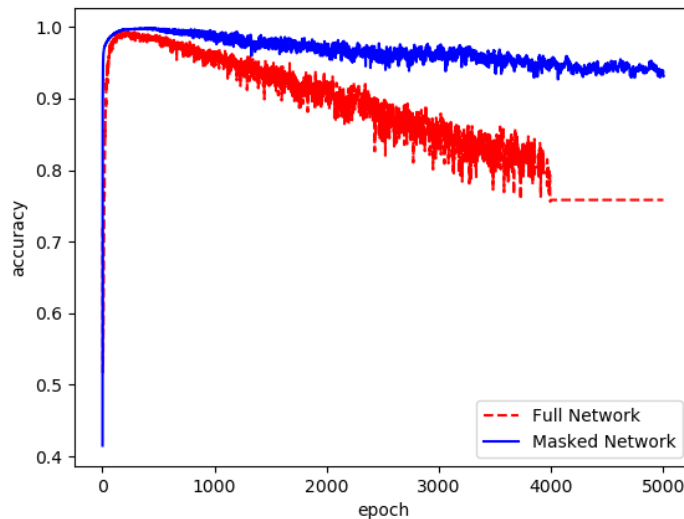
Task

Pro

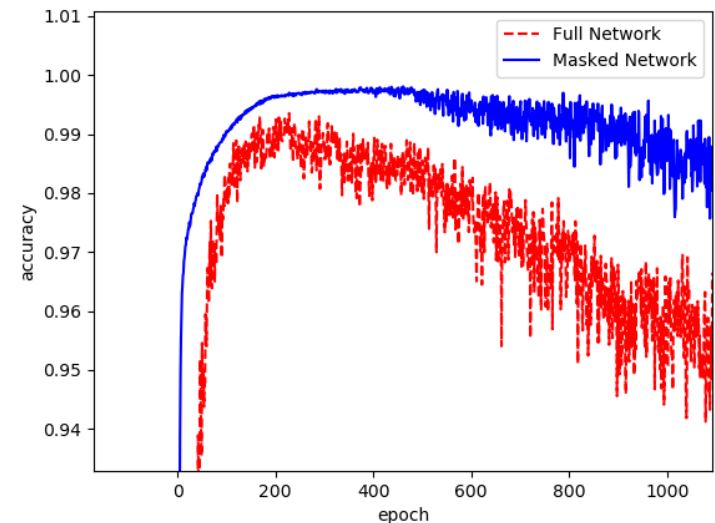
Rem

## Lenet-FCN-MNIST

- Training-Accuracy



Source:  
Produced by the author



Source:  
Produced by the author

# Progress – Background



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Int

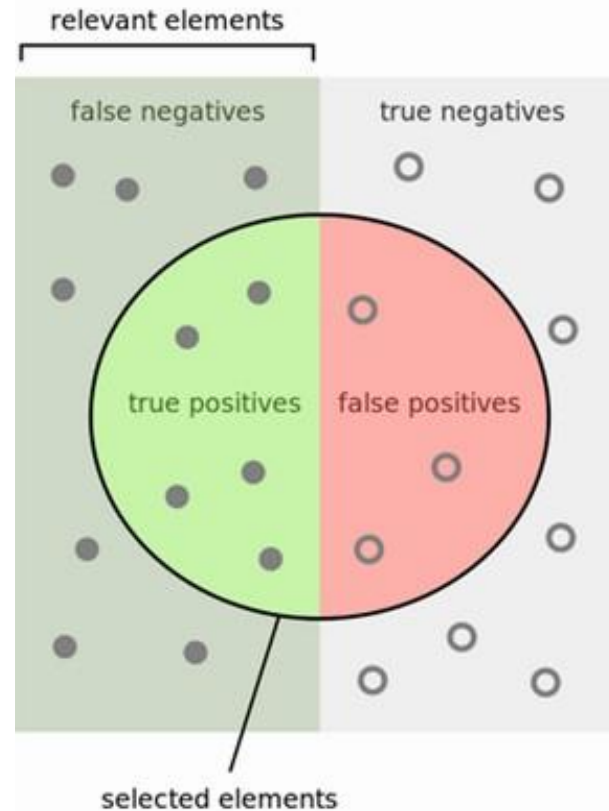
Mot

Rel

Task

Pro

Rem



How many selected  
items are relevant?

Precision =  $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$



How many relevant  
items are selected?

Recall =  $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$



Source:

<https://www.kdnuggets.com/images/precision-recall-relevant-selected.jpg>

Int

Mot

Rel

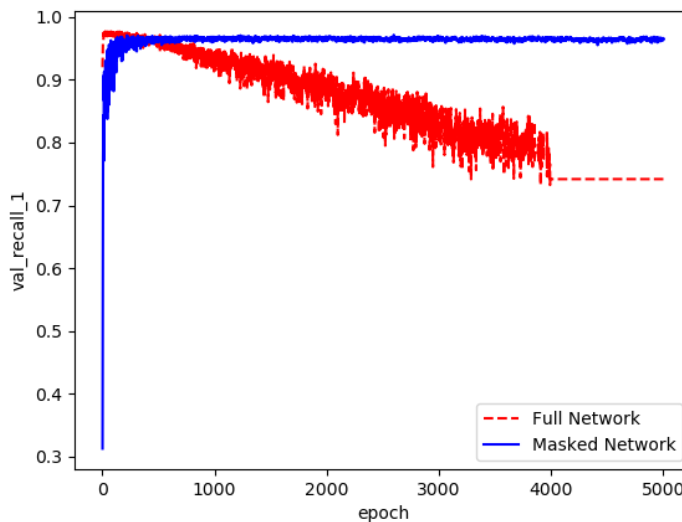
Task

Pro

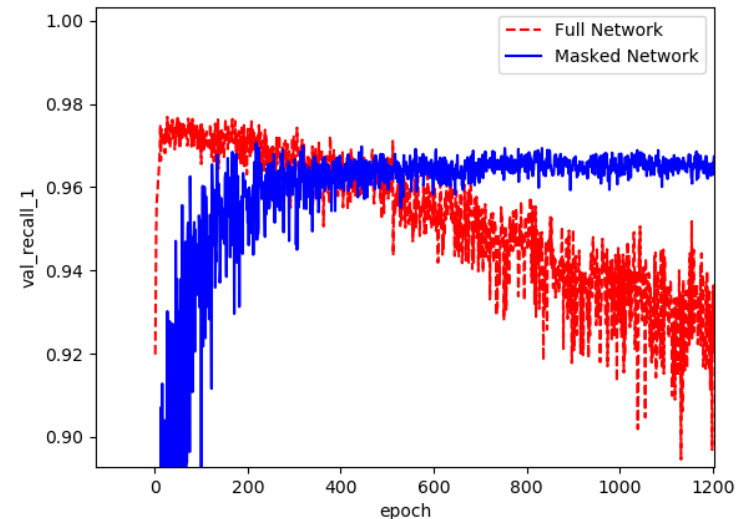
Rem

## Lenet-FCN-MNIST

- Validation-Recall



Source:  
Produced by the author



Source:  
Produced by the author

Int

Mot

Rel

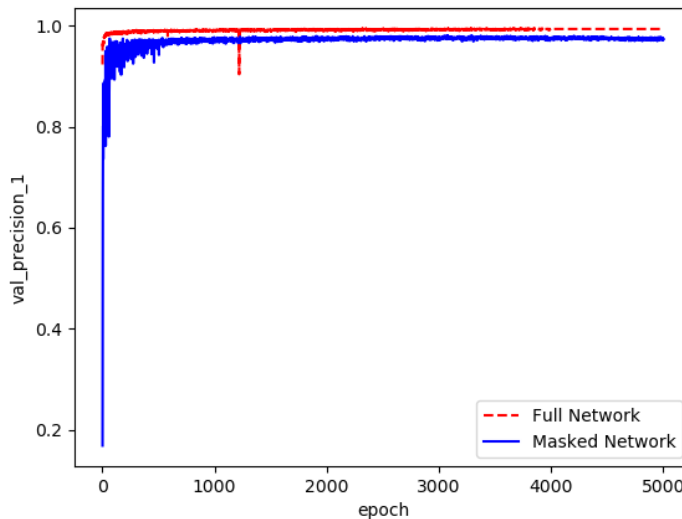
Task

Pro

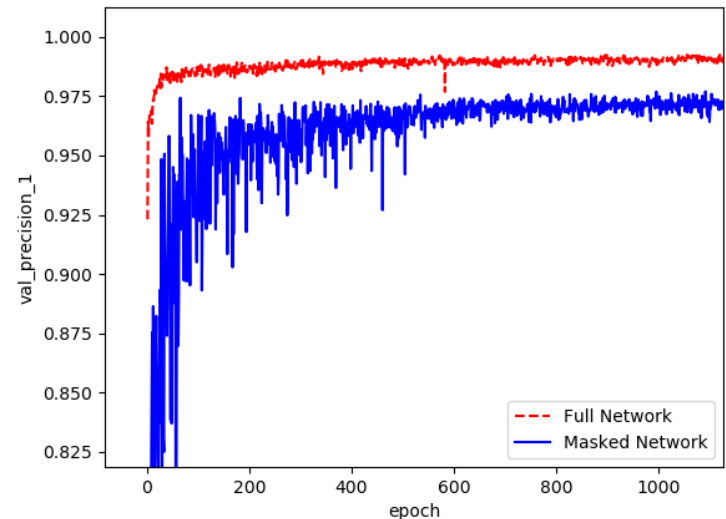
Rem

## Lenet-FCN-MNIST

- Validation-Precision



Source:  
Produced by the author



Source:  
Produced by the author

Int

## Remaining parts of the framework

- Custom Convolutional Layer
- Support for iterative Pruning

Mot

Rel

## More experiments

- MNIST / CNN-4
- MNIST / VGG-18
- Reuters / TBD
- MNIST / Lenet-FCN / Early Pruning

Task

Pro

Rem



---

# Thank you for your attention! Questions?

---



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

