# Master project Convolutional Neural Networks & Domain Decomposition

September 26th, 2023

TU Delft

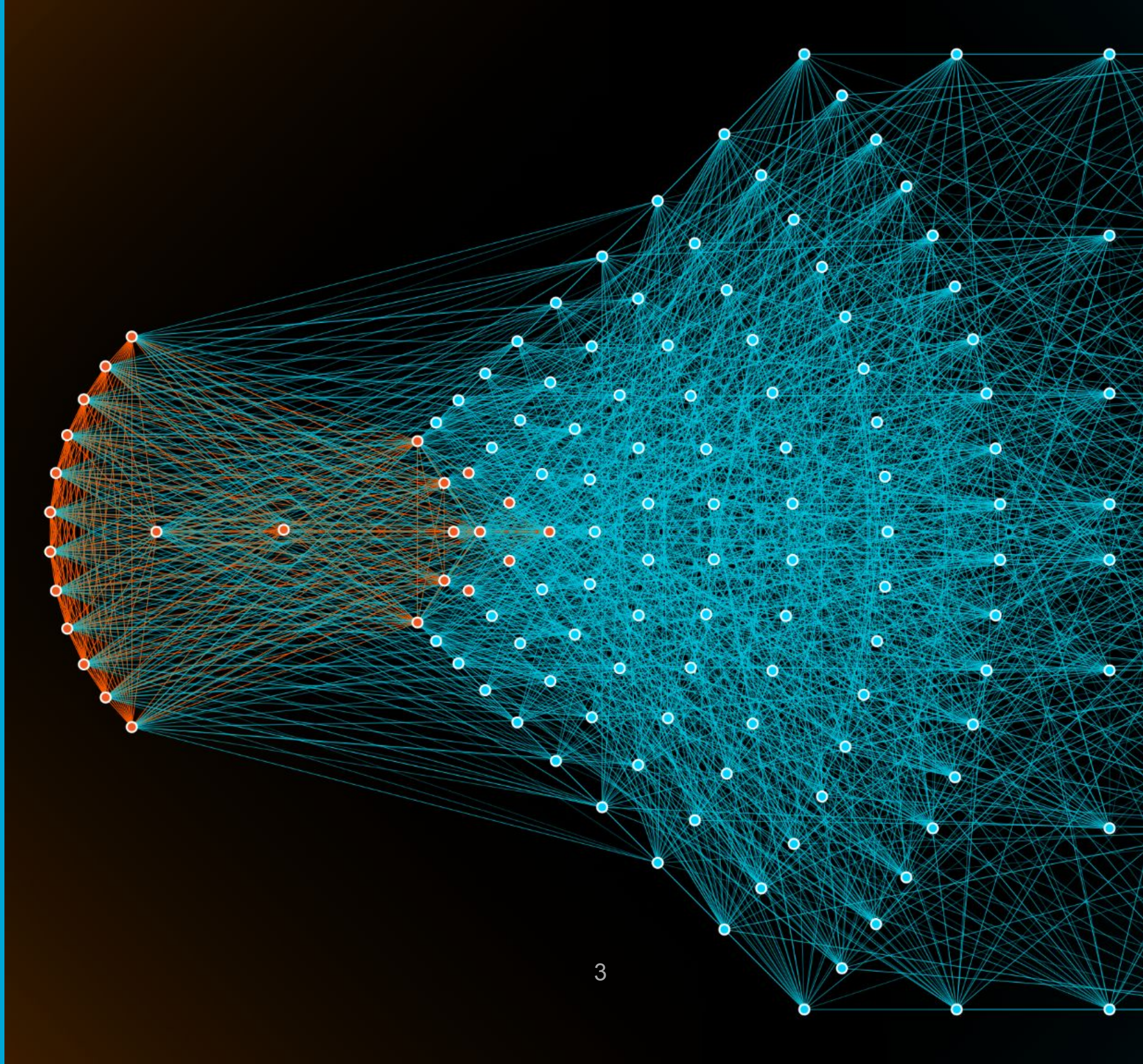
# Outline of this meeting

1. **Introduction**
2. **Existing research on Domain Decomposition-inspired (C)NNs**
3. **Research proposal**
   a. Proposed network architecture
   b. Some preliminary results
   c. Research purpose and sub-questions
4. **Questions and discussion**

TU Delft

# Introduction

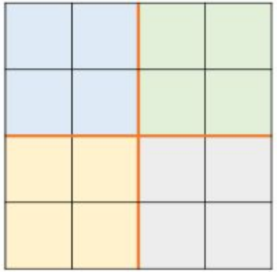
**Convolutional Neural Networks & Domain Decomposition**



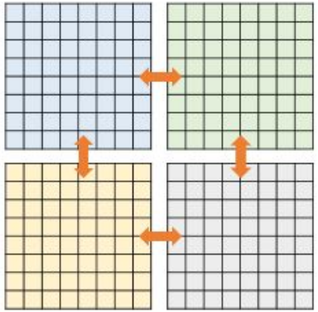
# Very brief recap on DDMs and CNNs (1)

**Domain Decomposition Methods (DDMs)**

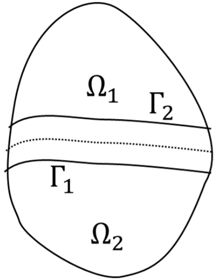
- Suitable for parallel computing
- Two classes:
  - *Overlapping DDMs:* sharing global information by overlapping boundaries
  - *Non-overlapping DDMs* : continuity at boundary is enforced by boundary conditions
- Multi-level DDMs: coarse problem coordinates the global solution (ensures continuity between subdomain solutions)



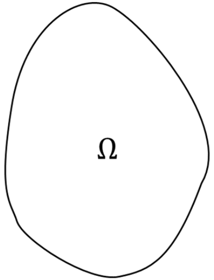
Coarse problem



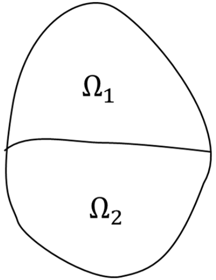
Fine problem




Overlapping decomposition




Full domain




Non-overlapping decomposition

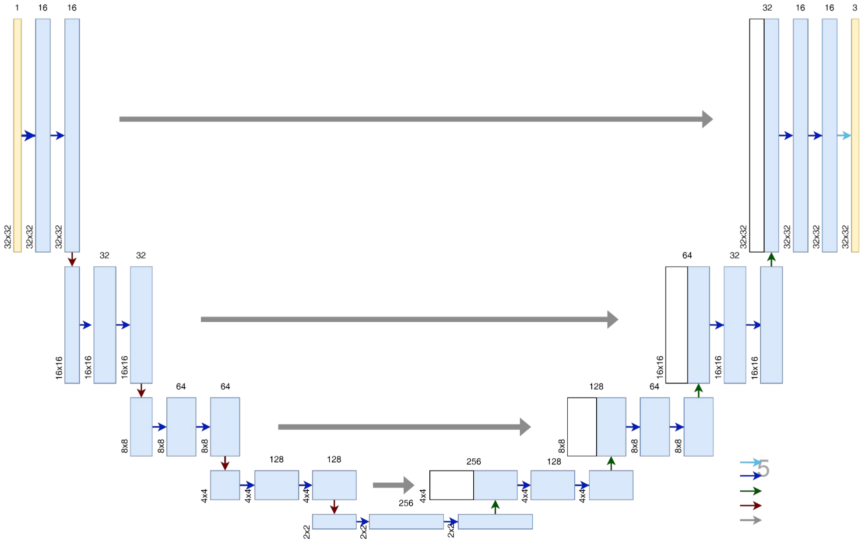
# Very brief recap on DDMs and CNNs (2)

## Convolutional Neural Networks (CNNs)

- Feedforward Neural Network that uses filters (kernel) optimization
- Hierarchical architecture: multiple layers, contraction and expansive paths
- Known for their ability to recognize patterns and extract features → classification, segmentation tasks

**Example: U-Net, a network for biomedical segmentation**

- Contraction path
  - Reduces amount of spatial information, increases feature information
- Expansive path
  - Combines feature and spatial information using information from *skip connections* and feature maps



# Challenges and opportunities in combining DDMs and CNNs

## Opportunities

1. **Parallel processing**. DDMs inherently encourage parallel processing, which is also a desired characteristic for CNNs. CNN data (images, voxel images) is intuitively easy to decompose into subdomains (sub-images)

2. **Scalability**. Certain DDMs support scalability, a valuable trait when dealing with large resolution image datasets.

3. **Memory efficiency.** DDMs partition the domain and the computations among multiple devices, leading to efficient memory management. This can als be advantageous for distributing a CNN among multiple devices.

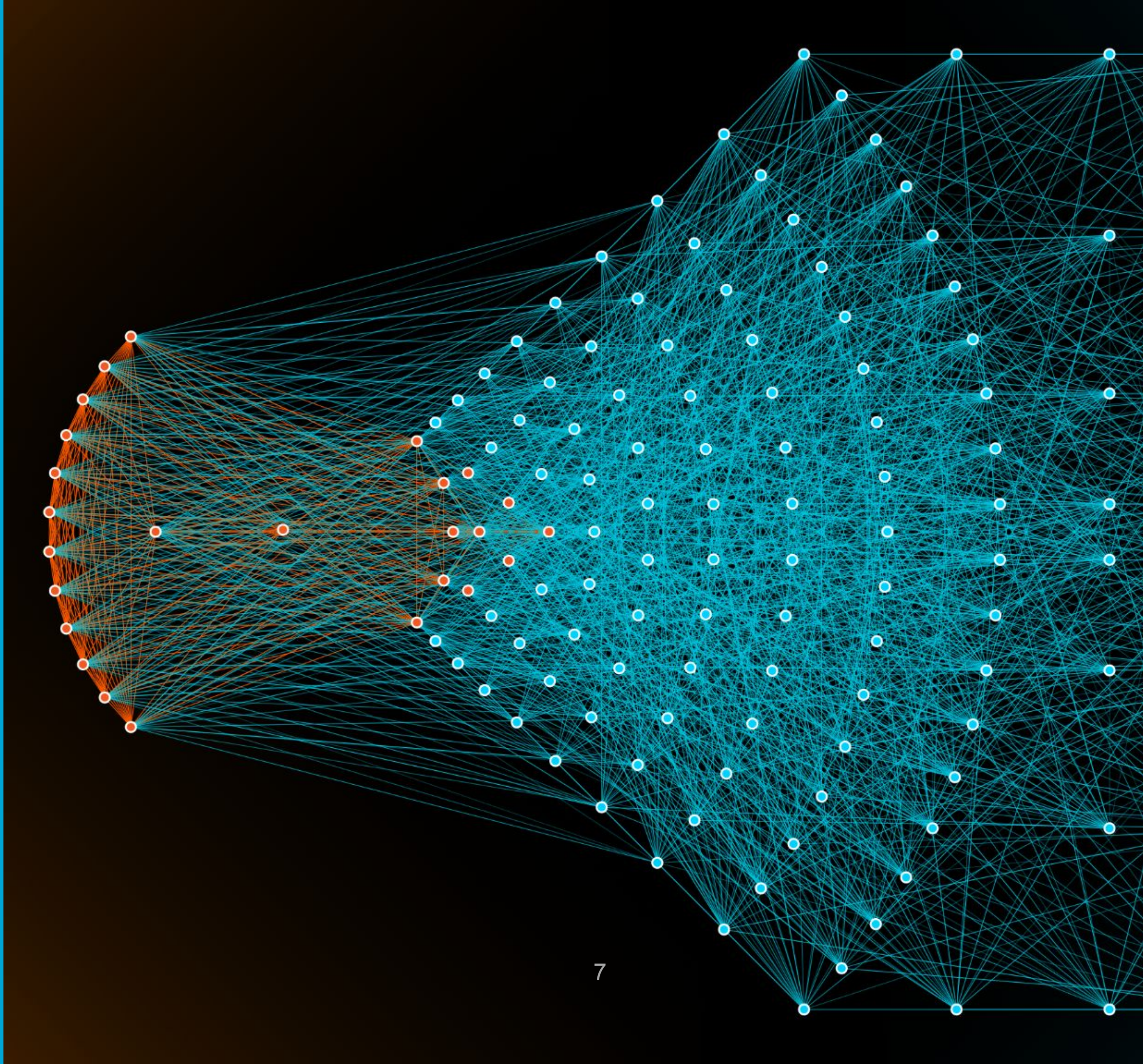
## Challenges

1. **Communication between subdomains** can easily lead to a communication overhead during training/inference.

2. **Fundamental application difference**. DDMs are used to approximate PDE solutions, whereas CNNs are mostly used for ML tasks not directly related to PDEs.

# Existing Research

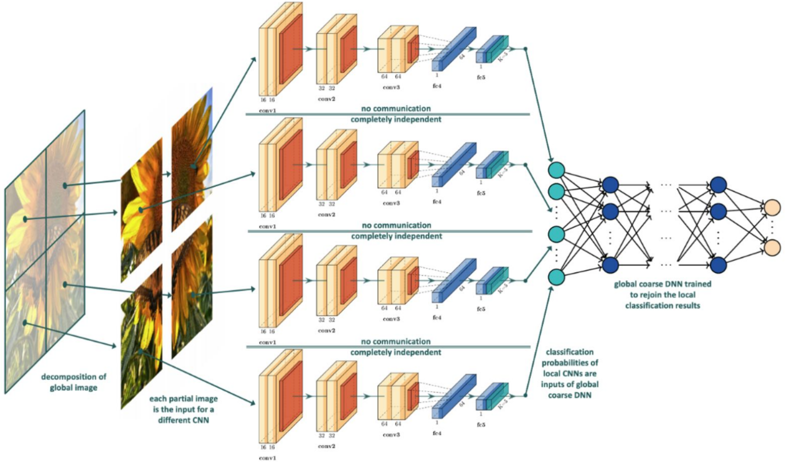
**Convolutional Neural Networks & Domain Decomposition**



# What is already done? (1)

**A DD-based CNN-DNN Architecture (Klawonn et al., 2023)**

- Image decomposed into sub-images
- Each sub-image processed by separate CNN sub-network that produces a local prediction
    - Allows for fully parallel training (!)
- Local predictions are combined in a Fully Connected Network (FCN) to obtain one global prediction.

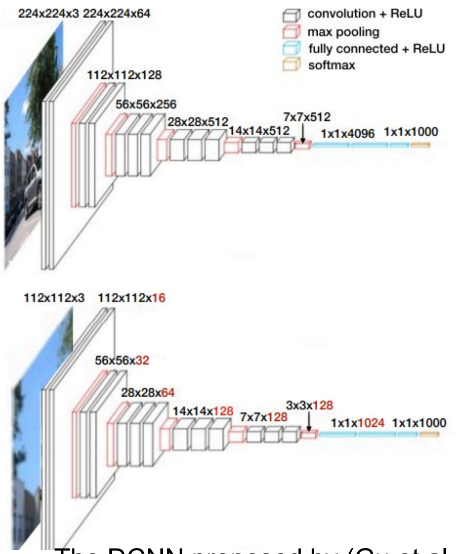



Architecture proposed by Klawonn et al. (Klawonn et al., 2023)

# What is already done? (2)

**Decomposition and composition of DCNNs (Gu et al., 2022)**

- Image is partitioned into N sub-images
- Large CNN is partitioned along the channel dimension into N sub-networks
  - Can be trained separately
- Sub-networks are combined into one large CNN, fine-tuning is applied then.




The DCNN proposed by (Gu et al., 2022)

# Research gap

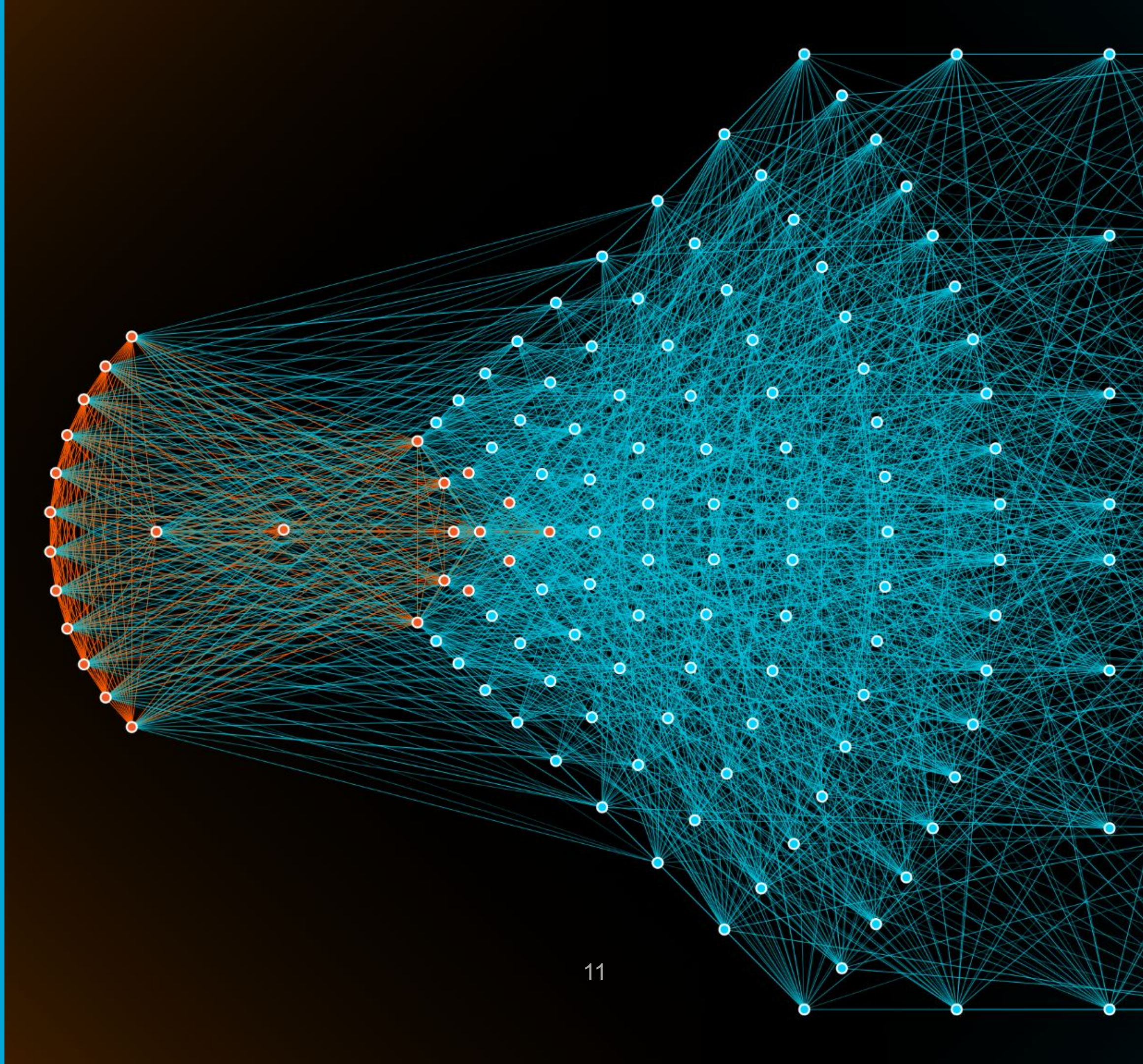
**Some unanswered questions:**

- How to transfer information between different subdomains when the information is far apart (more than a few pixels)?
- How to construct a DDM-inspired convolutional network suitable for image **segmentation** tasks?
- Can we exchange information from latent space (intermediate feature maps) between subnetworks instead of spatial information?
    - Feature maps: lower spatial dimension, “high-level” information

# Research Proposal

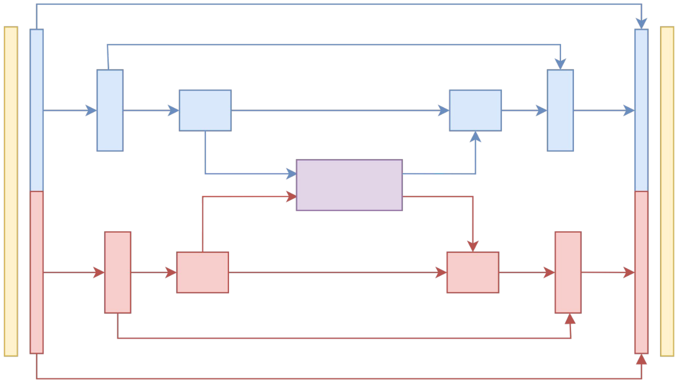
**Convolutional Neural Networks & Domain Decomposition**

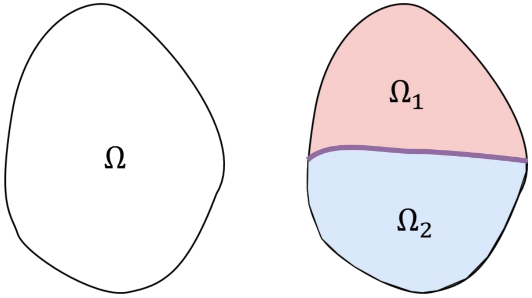
TU Delft



# Proposed network architecture (1)

- We propose a **segmentation network**, based on the U-Net architecture
- Each input image is partitioned into two (or more) sub-images
- How is this network based on DDMs?
  - Subdomains $\Omega_1$ en $\Omega_2$ $\leftrightarrow$ image subdomains 1 and 2 (can be overlapping or non-overlapping)
  - Solvers on both subdomains $\leftrightarrow$ network operating on both subimages (can be the same solver/network)
  - Communication of boundary conditions on $\Gamma$ via coarse problem $\leftrightarrow$ FC network connecting both domains

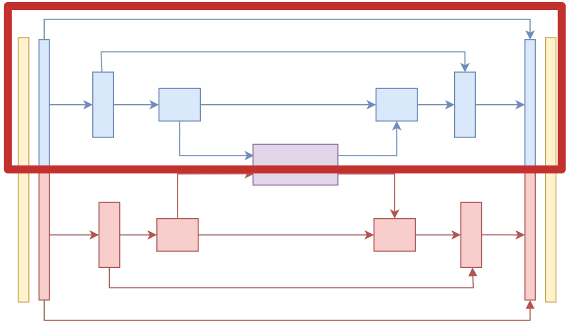




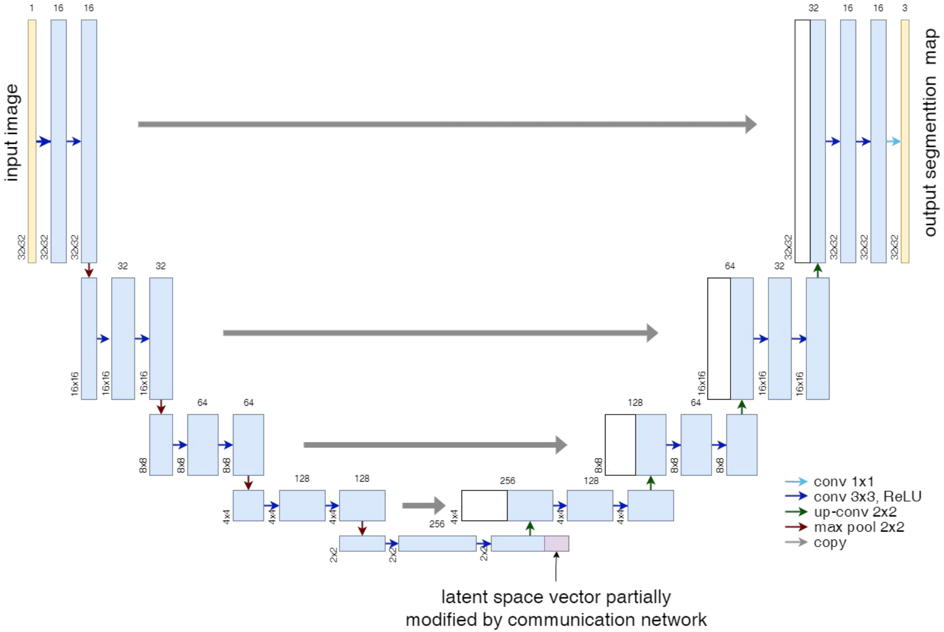

Full domain

Non-overlapping decomposition

# Proposed network architecture (2)

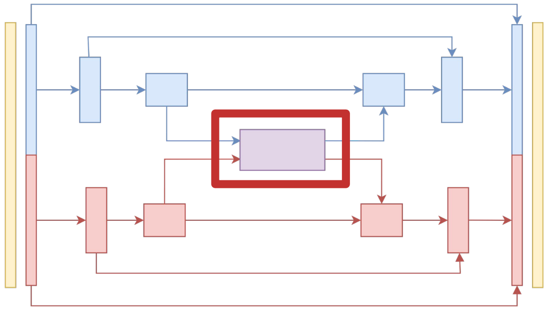


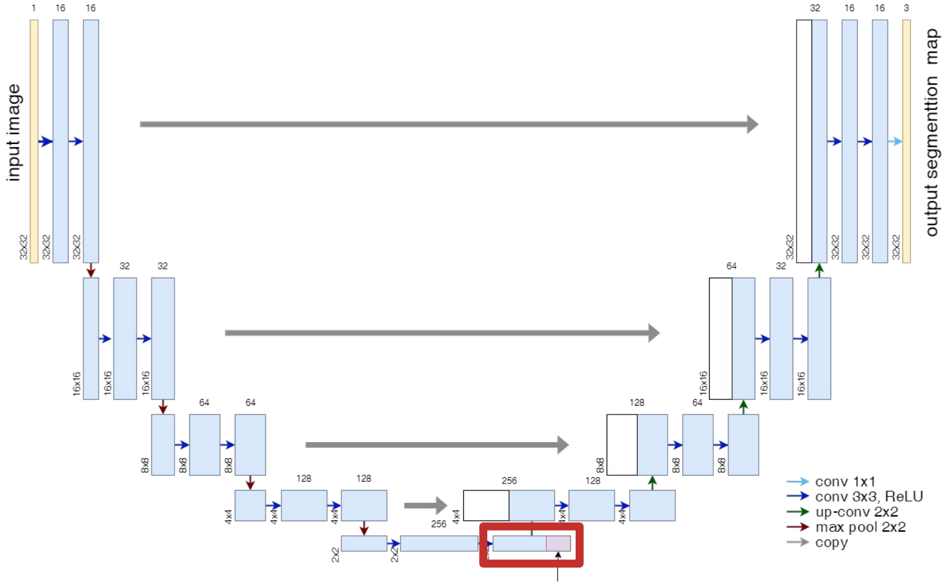
- **Sub-network**: based on the architecture of U-Net
  - Exactly the same architecture as U-Net, except for the latent space feature maps (modified by communication network)



# Proposed network architecture (3)

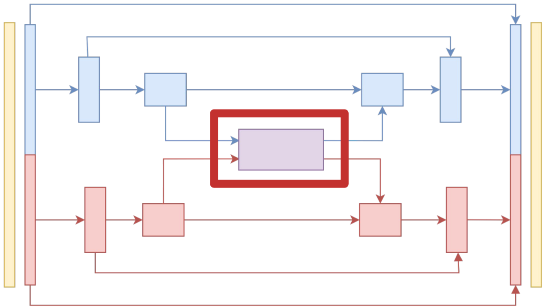
- **Communication network:** exchanges information between two sub-domains by modifying a part of the latent feature maps



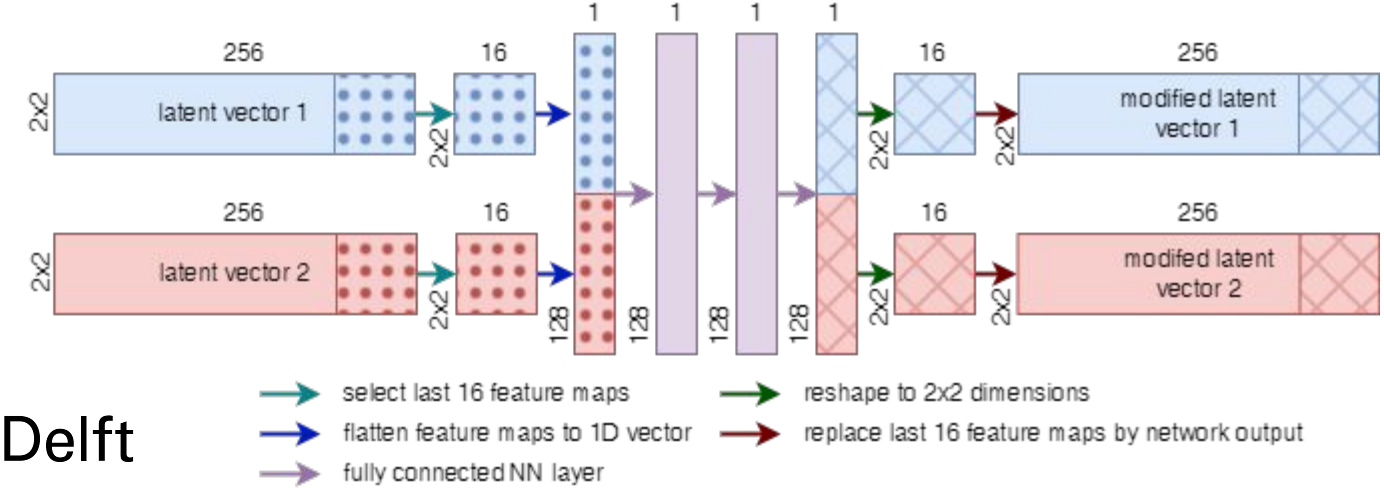



latent space vector partially
modified by communication network

# Proposed network architecture (4)



- **Communication network:** FCN that exchanges information between two sub-domains by modifying a part of the latent feature maps.
    - *Inputs*: latent feature maps for two sub-domains
    - *Outputs:* modified latent feature maps for two-subdomains

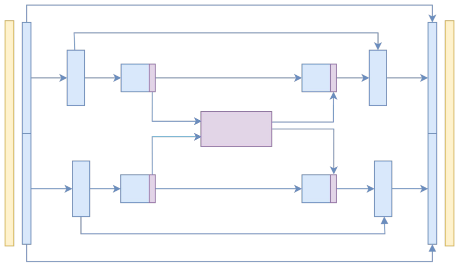



TU Delft

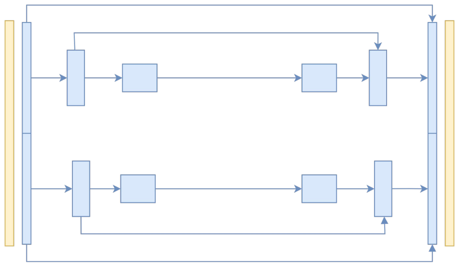
# Preliminary results (1)

**Central questions**:

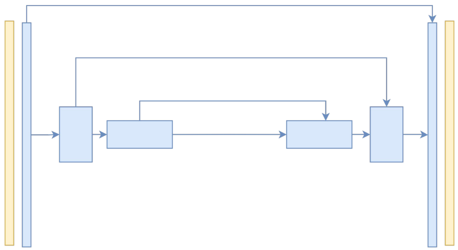
- Does the proposed network architecture succeed in transferring global information?
- How does it compare to a architecture with two subdomains without communication?
- How does it compare to a baseline U-Net model?



*Proposed network architecture (with communication network)*



*Proposed network architecture (without communication subnetwork)*

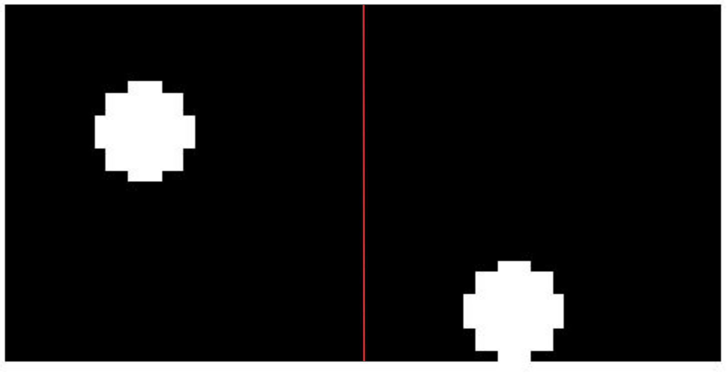


*Baseline U-Net model*

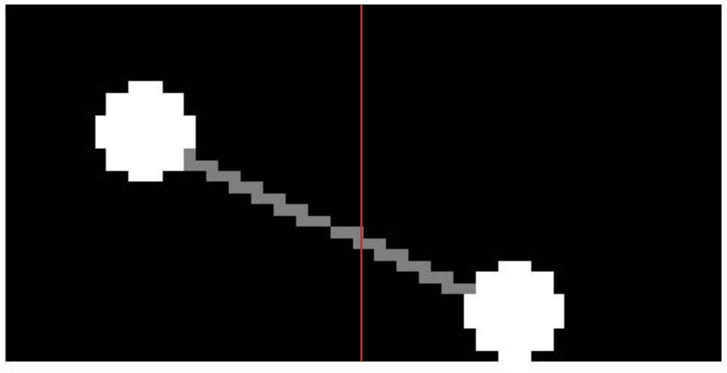
# Preliminary results (2)

**Toy problem:**

- Designed to require: global communication for successful performance
- **Input**: Two randomly located circles (one on each subdomain).
- **Task**: predict the line connecting the two circles
- **Metric**: accuracy of the prediction of the line segment (% correctly predicted pixels)



*(a) Input image*

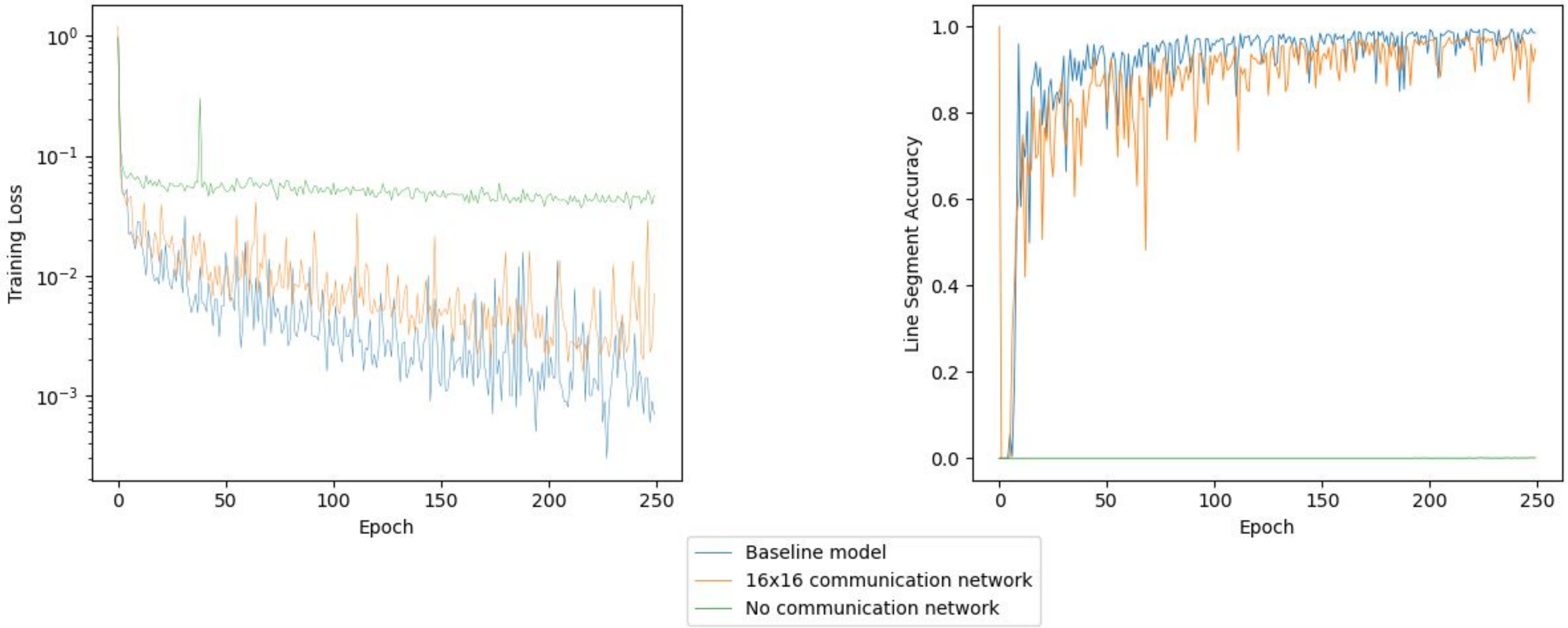


*(b) Desired output*

Image (a) and corresponding mask (b) pair example from the dataset used for training and evaluation

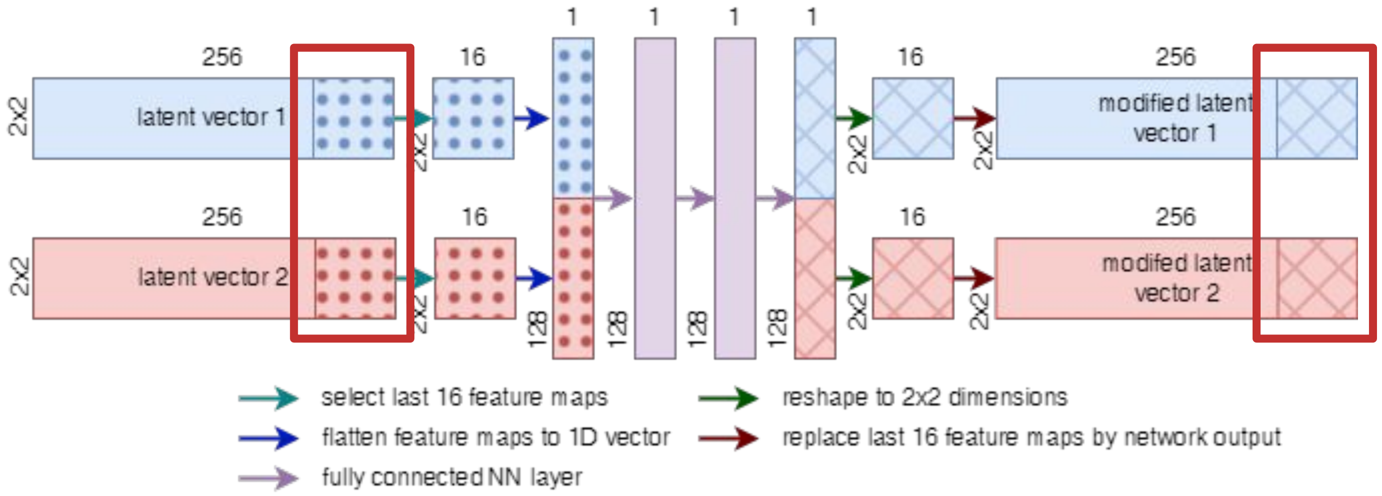
TU Delft

# Preliminary results (3)



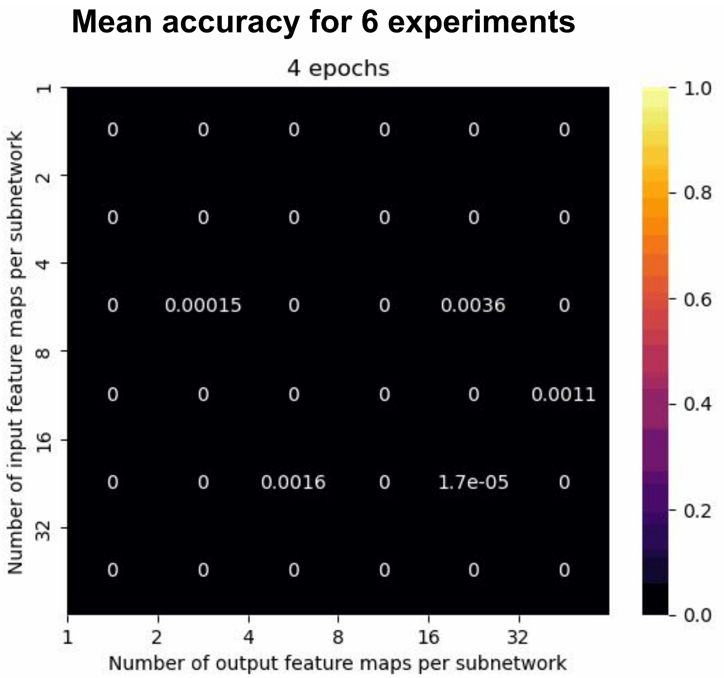

*Training Loss and Line Segment Accuracy for the toy problem for three different segmentation networks.*

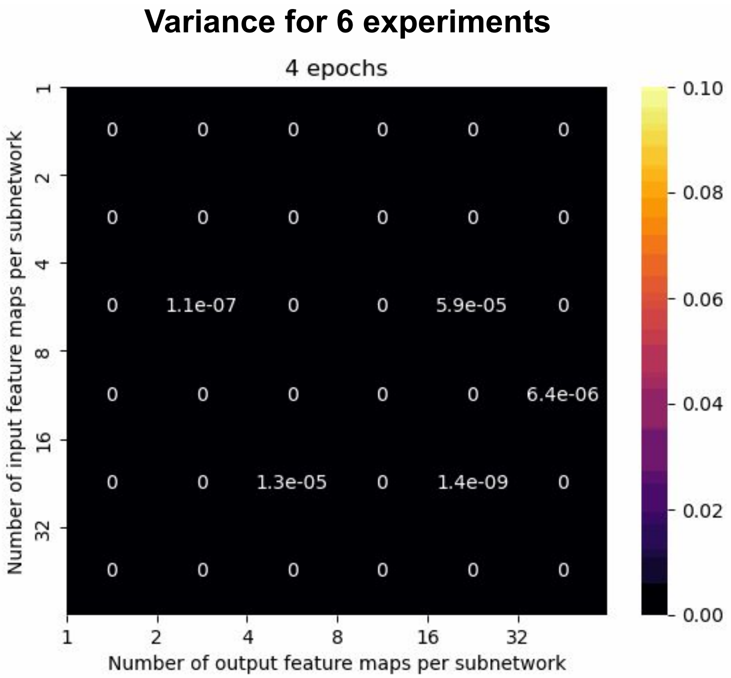
# Influence of the size of the communication network




TU Delft

# Mean and variance of accuracy for different input + output sizes of the communication network




TU Delft

# Research Proposal

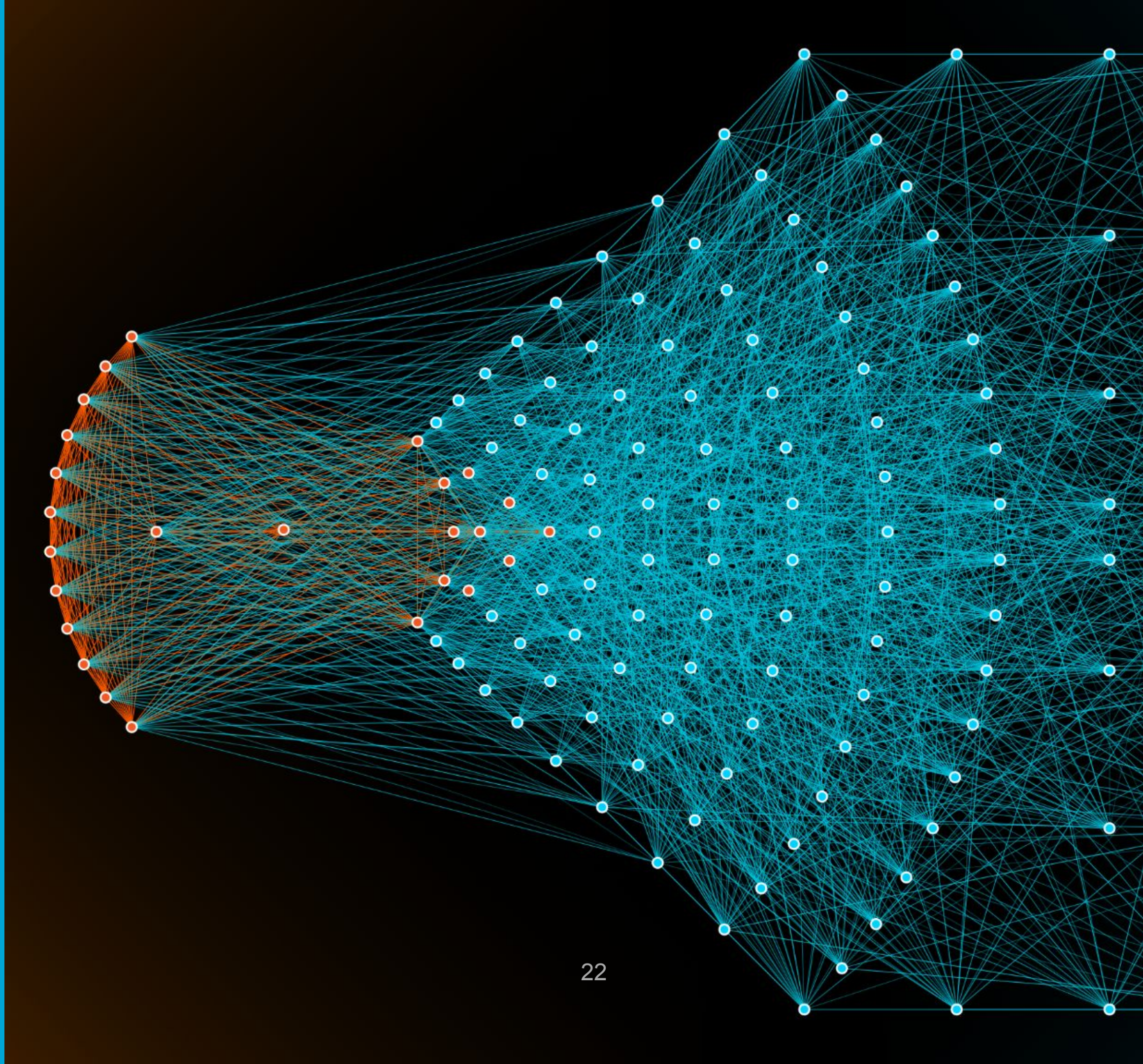
**Research purpose:** *To develop and investigate a Domain Decomposition-inspired CNN architecture, consisting of CNN sub-networks interconnected by a Feature Map Communication Network*

**Sub-questions**

- *What should the size of the communication network be for accurate predictions?*
- *How does the amount of necessary communication change for different types of data sets and segmentation tasks?*
- *How can we extend the proposed network to more subdomains? Do we need a larger communication network then? How does this scale with problem size and complexity?*
- *How do the accuracy, computational performance, and memory requirements of this model compare to baseline methods such as the U-Net model?*
- *How can the proposed architecture be used to speed up parallel training and evaluation and/or improve accuracy for segmentation tasks?*

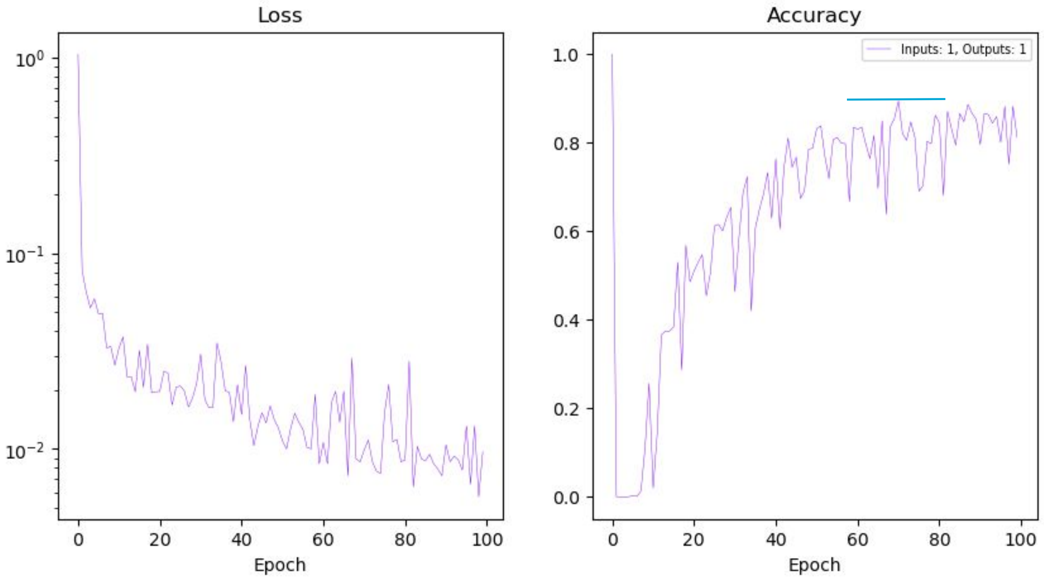
# Questions and Discussion

**Convolutional Neural Networks & Domain Decomposition**



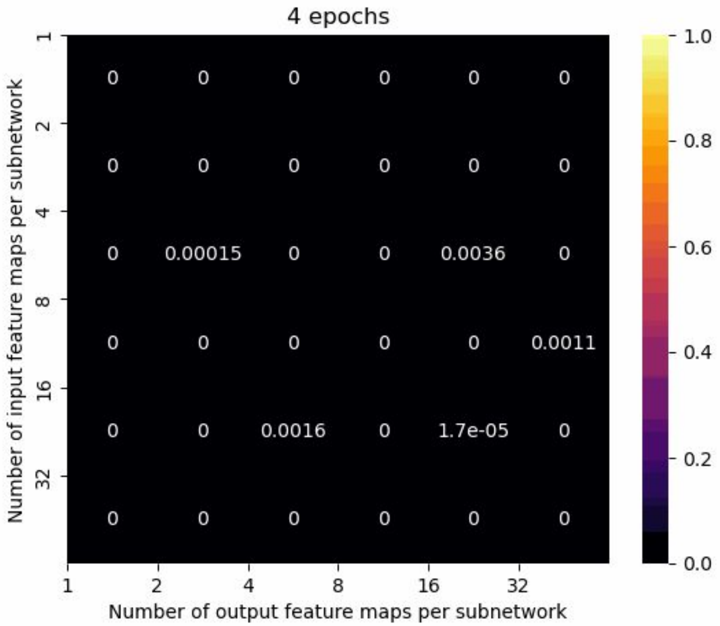
# Mean and variance of accuracy for repeated experiments (1)

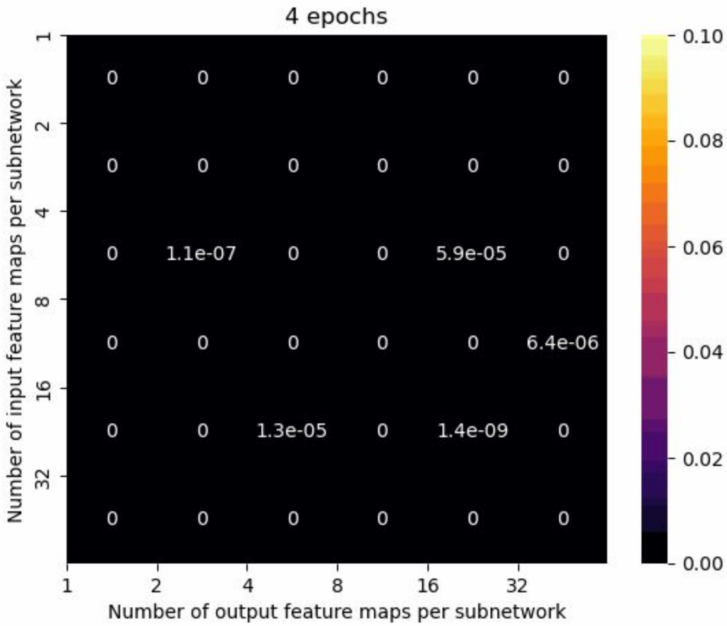
- Training hyperparameters:
  - Batch size: fixed at 16
  - Learning rate: fixed at 0.0005
  - Epochs: limited at **100**
- Model parameters:
  - Input and output sizes of the communication network are both in [1,2,4,8,16,32]
- Repeated the experiment **6** times
  - Weights initialized following the same procedure, however with another seed value
- In total: **216** networks are trained
- Metric: maximal validation accuracy of the line segments between epoch 1 and epoch 100




TU Delft

# Mean and variance for repeated experiments (2)

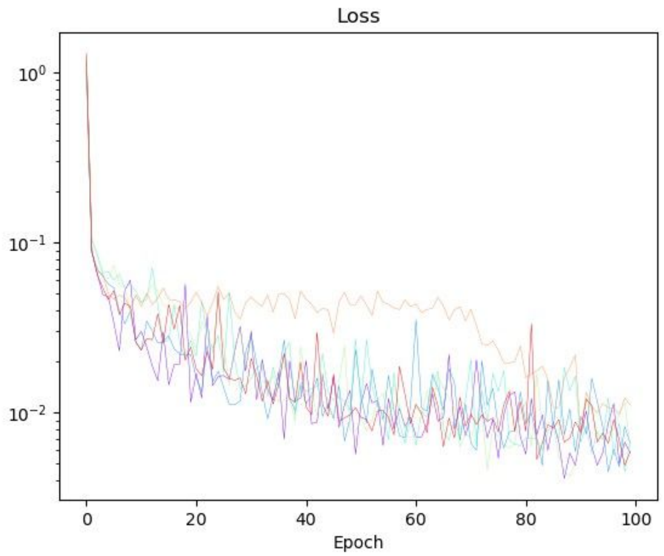


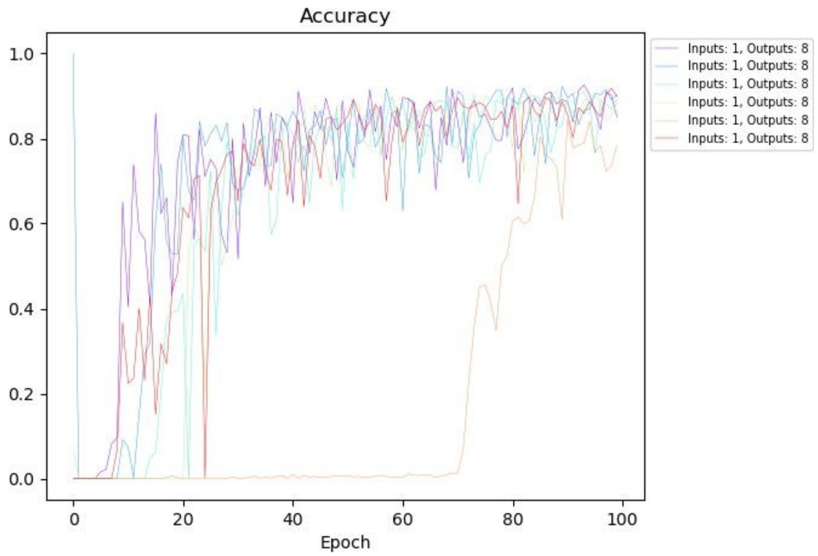




TU Delft

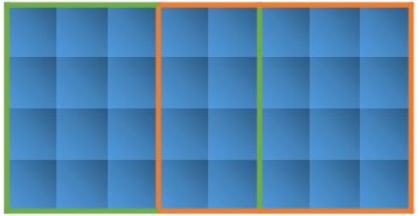
# Mean and variance of accuracy for repeated experiments (3)

- **Outliers**
  - 1 input, 2 outputs (high variance, until 100 epochs)
  - 1 input, 8 outputs (high variance, until 70 epochs)

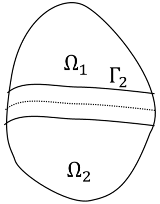


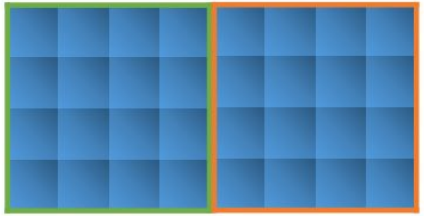
# Sub-domains for DDMs and CNNs



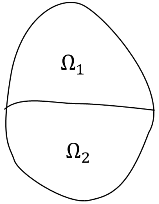
overlapping
subdomains



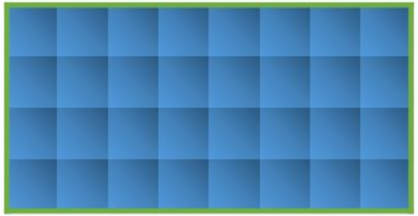

Overlapping
decomposition



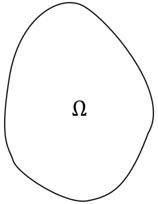
non-overlapping
subdomains




Non-overlapping
decomposition



classical U-Net




Full domain

# Sources


- Axel Klawonn, Martin Lanser, and Janine Weber. A domain decomposition-based CNN-DNN architecture for model parallel training applied to image recognition problems. *arXiv preprint arXiv:2302.06564, 2023.*

- Linyan Gu, Wei Zhang, Jia Liu, and Xiao-Chuan Cai. Decomposition and composition of deep convolutional neural networks and training acceleration via subnetwork transfer learning. *Electronic Transactions on Numerical Analysis*, 56:157– 186, 2022