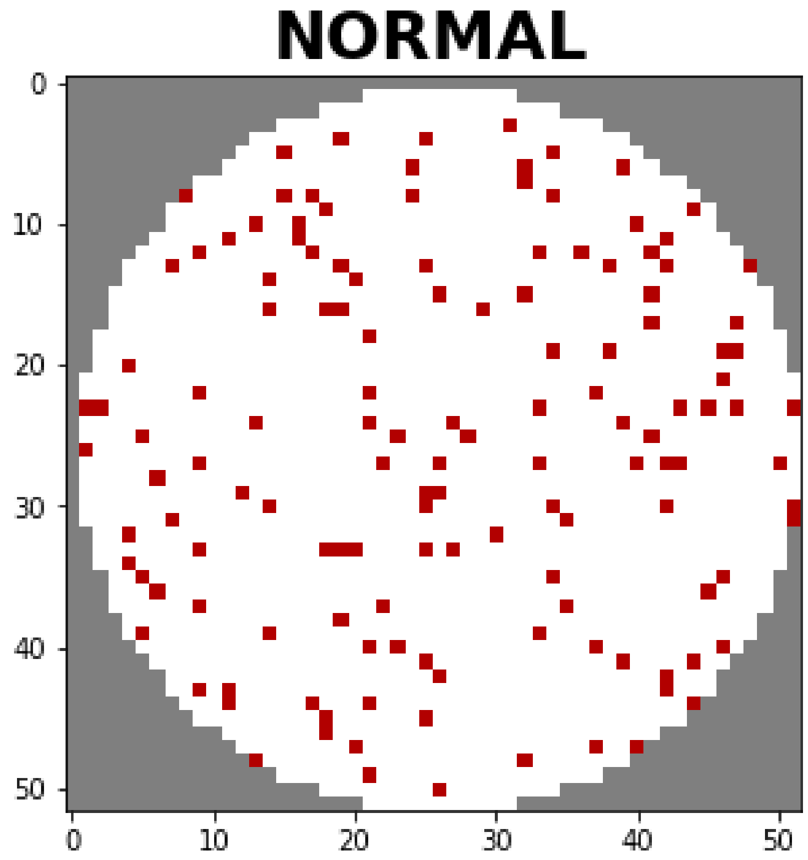
Intro

- Silicon wafers that comprise an array of chips are susceptible to defects.
- While some defects are expected (yield < 100%), systematic patterns of defective chips in a wafer are indicative of problems in the process line.
- The 'pattern' of defective chips in a wafer is indicative of the nature/location of the problem.
- Typically wafer maps are reviewed manually and dispositioned into error types.
- This project looks at training a conv-net to identify and classify defects in semiconductor wafers

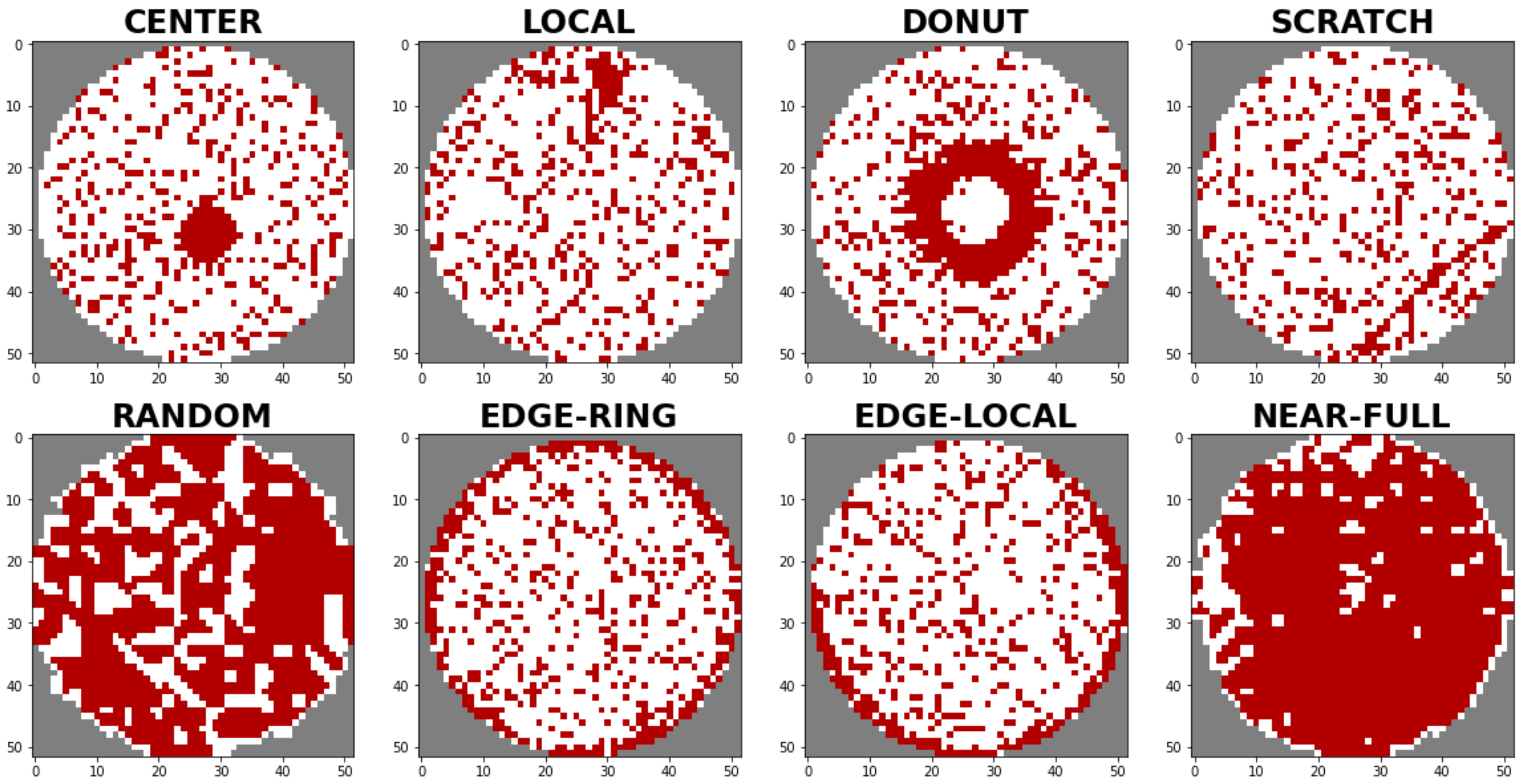


- Defective die
- Functional die
- Off-wafer

- Wang *et al*, *IEEE Transactions on Semiconductor Manufacturing*, 2020

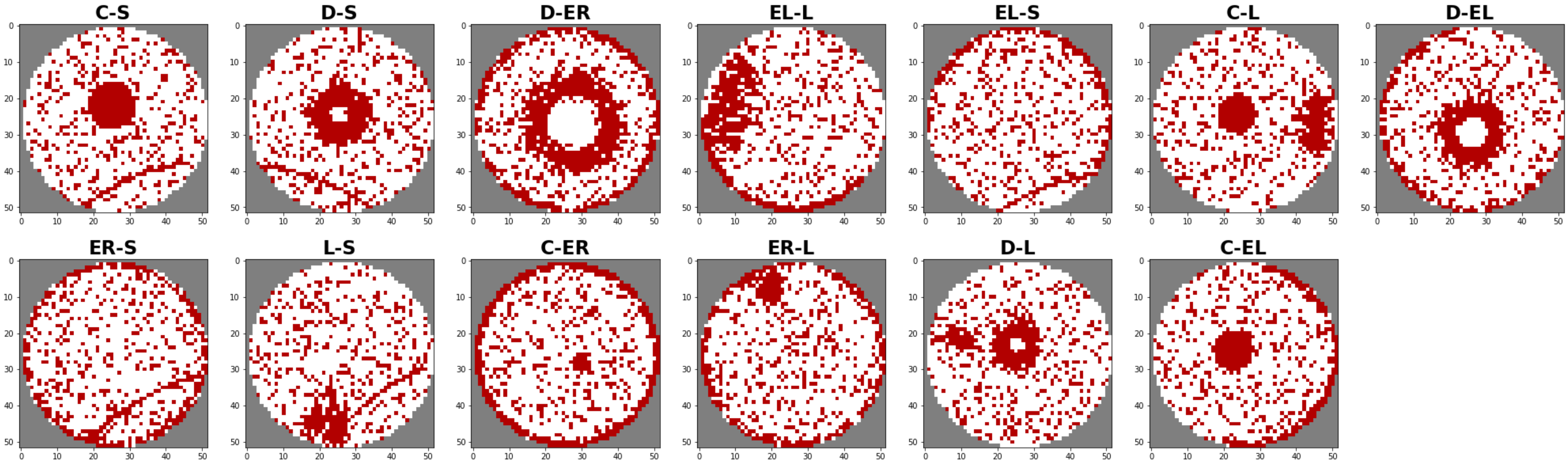
Dataset – Defect Types

- Public dataset from fab in China
- Each wafer map is a 54x54 array of values identifying each location as a functional die, defective die or off-wafer. Each label is a one-hot encoded vector indicating the presence (or absence) of 8 different defect types.
- There are wafers with 0, 1, 2, 3 and 4 defect types.



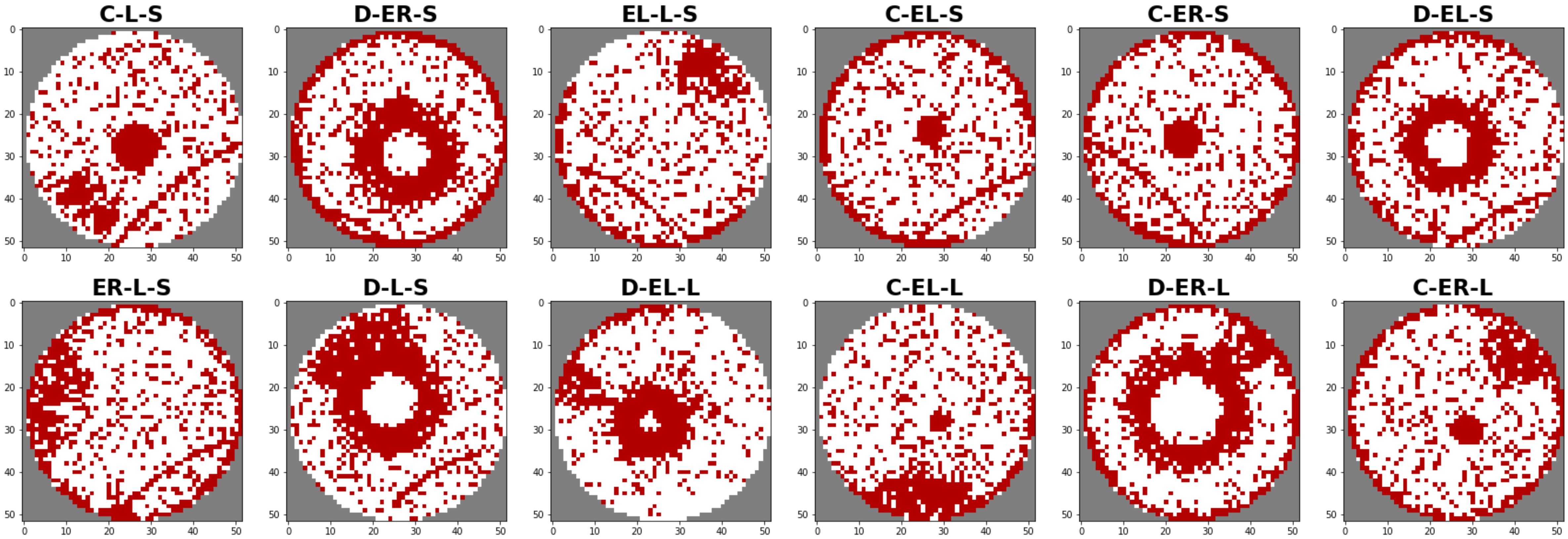
Dataset – 2 Defects Types in a Wafer

- Public dataset from fab in China
- Each wafer map is a 54x54 array of values identifying each location as a functional die, defective die or off-wafer. Each label is a one-hot encoded vector indicating the presence (or absence) of 8 different defect types.
- There are wafers with 0, 1, 2, 3 and 4 defect types.



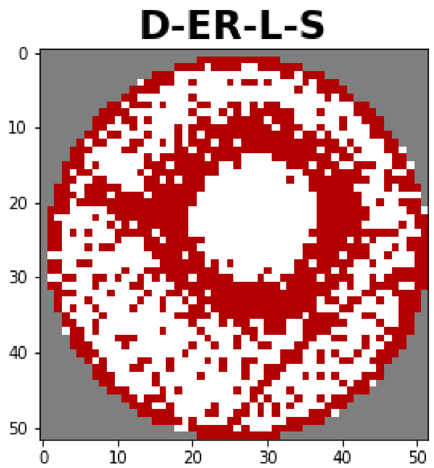
Dataset – 3 Defects Types in a Wafer

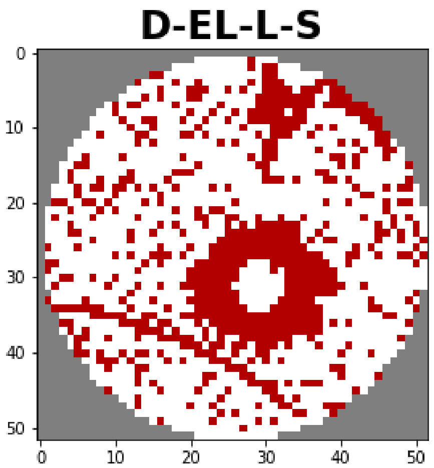
- Public dataset from fab in China
- Each wafer map is a 54x54 array of values identifying each location as a functional die, defective die or off-wafer. Each label is a one-hot encoded vector indicating the presence (or absence) of 8 different defect types.
- There are wafers with 0, 1, 2, 3 and 4 defect types.

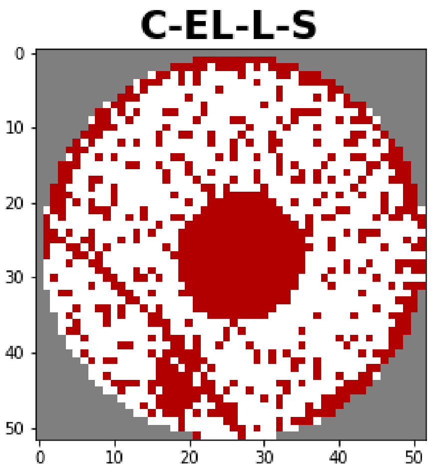


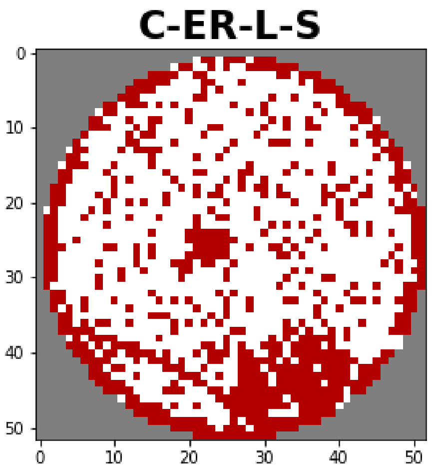
Dataset – 4 Defects Types in a Wafer

- Public dataset from fab in China
- Each wafer map is a 54x54 array of values identifying each location as a functional die, defective die or off-wafer. Each label is a one-hot encoded vector indicating the presence (or absence) of 8 different defect types.
- There are wafers with 0, 1, 2, 3 and 4 defect types.

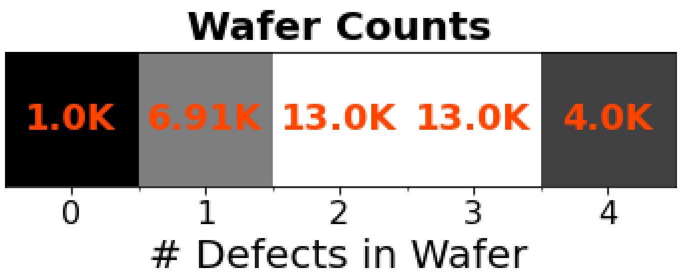


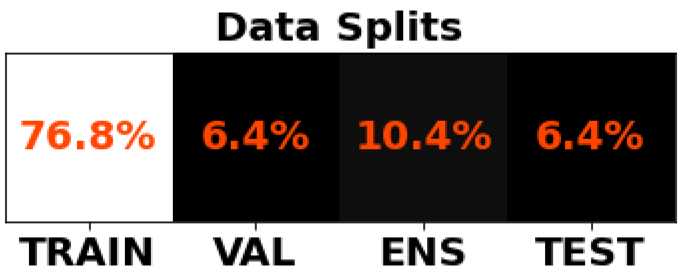


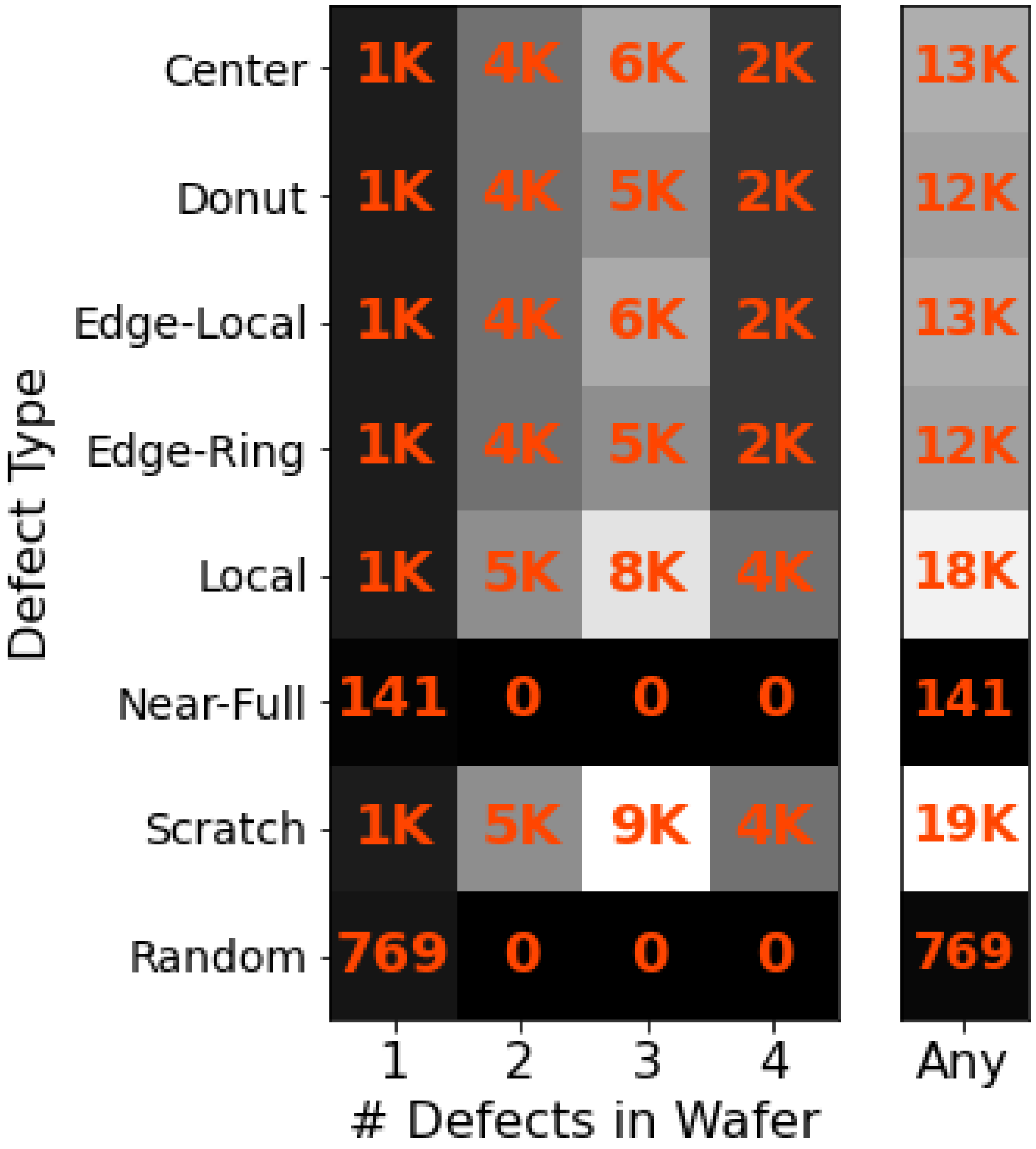




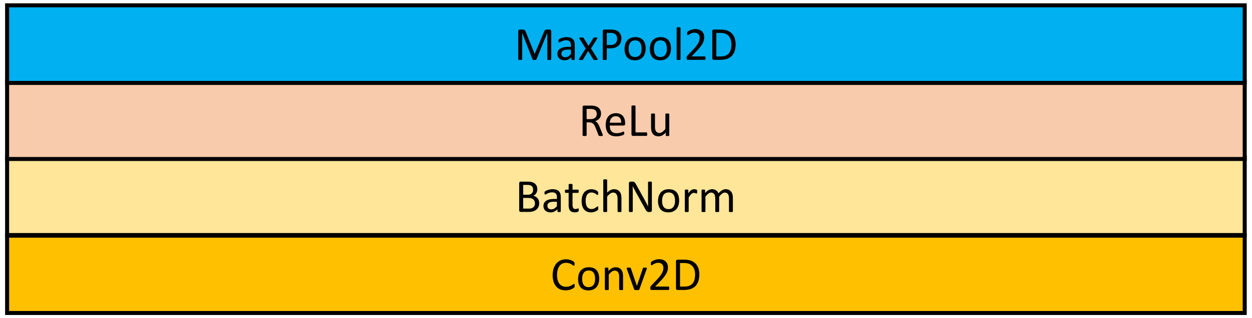
Dataset – Counts and Splits





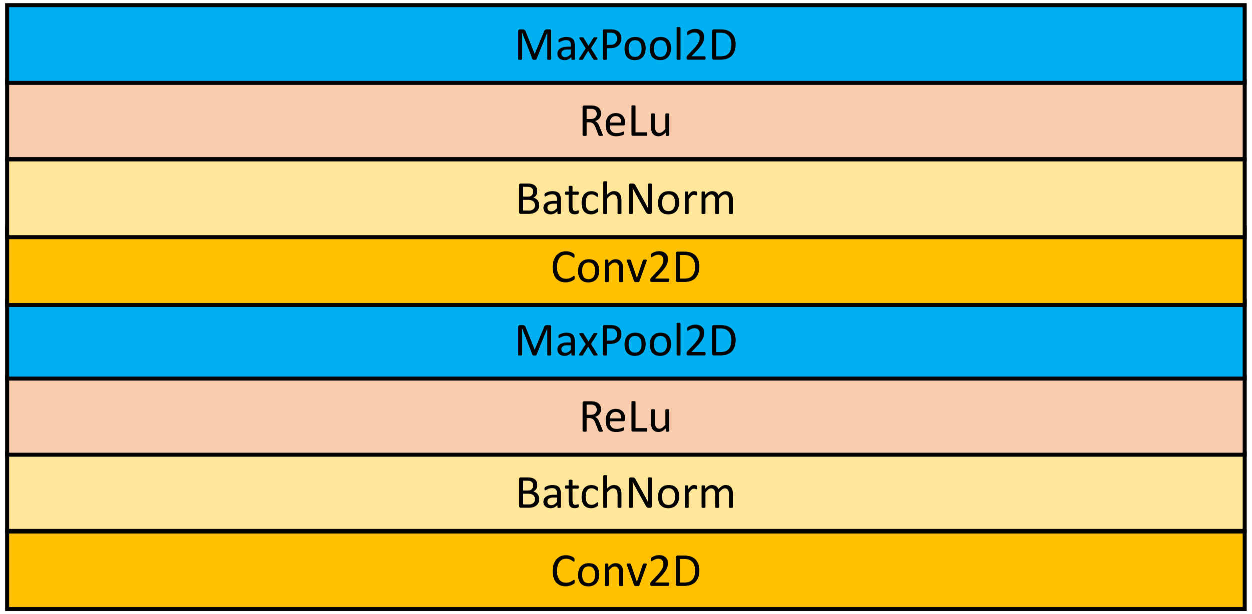


Model Architecture



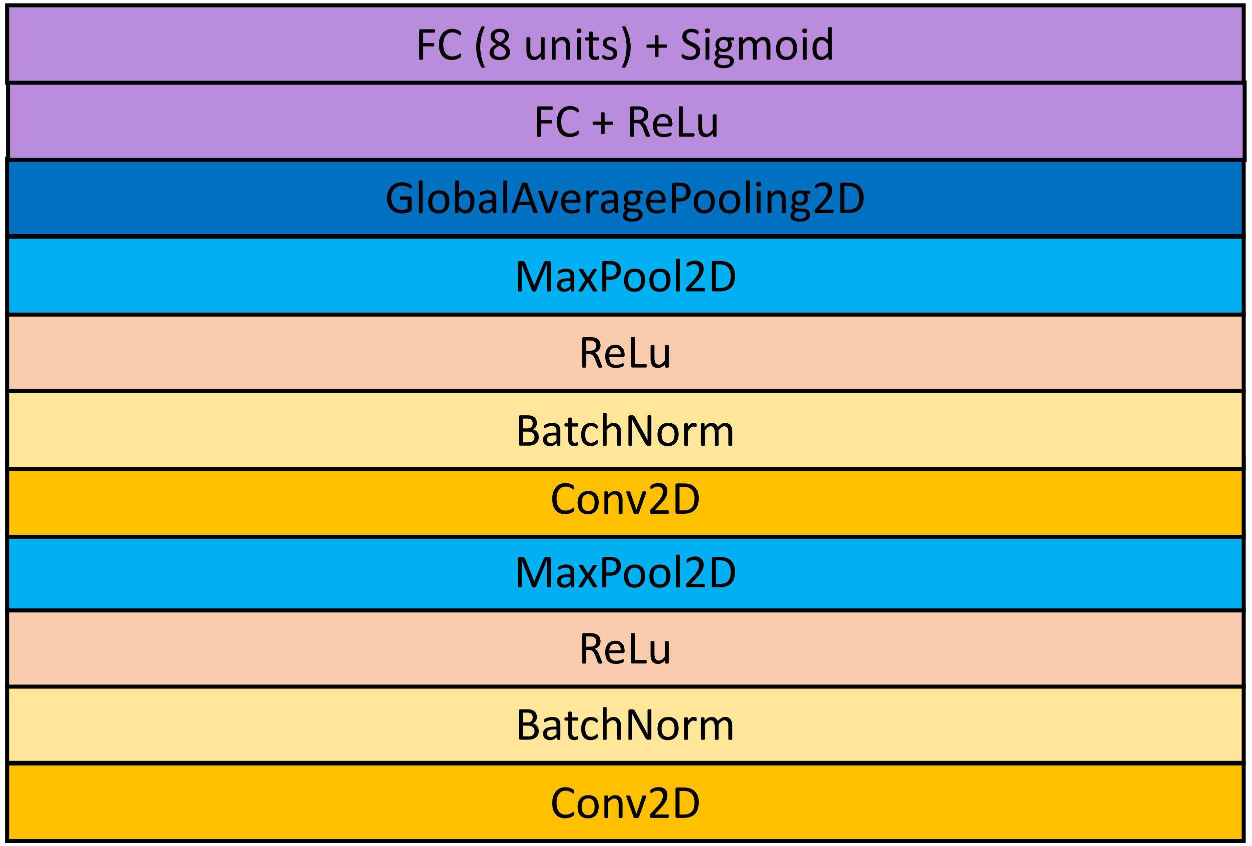
- Ioffe & Szegedy (2015) *arXiv: 1502.03167v3*
- Li et al (2018) *arXiv: 1801.05134v1*

Model Architecture



- Ioffe & Szegedy (2015) *arXiv: 1502.03167v3*
- Li et al (2018) *arXiv: 1801.05134v1*

Model Architecture



- Ioffe & Szegedy (2015) *arXiv: 1502.03167v3*
- Li et al (2018) *arXiv: 1801.05134v1*

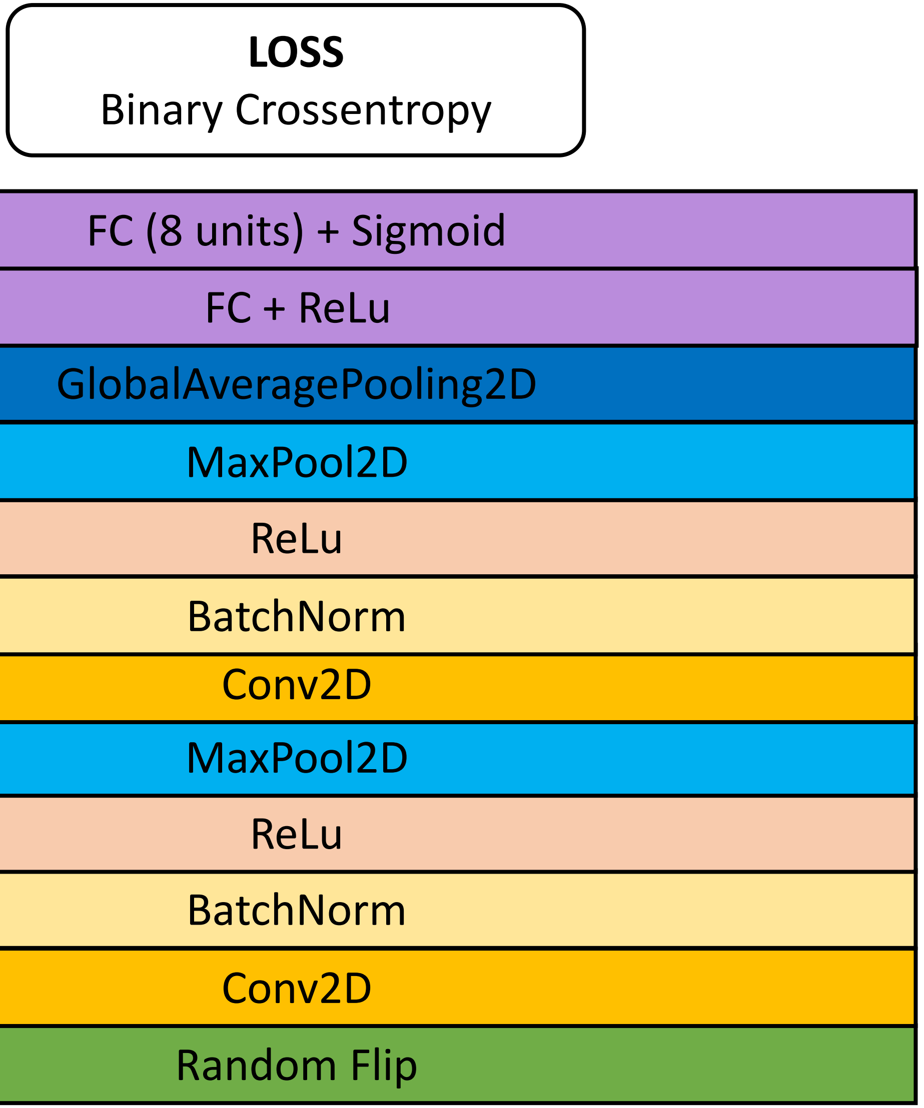
Model Architecture

LOSS
Binary Crossentropy

FC (8 units) + Sigmoid
FC + ReLu
GlobalAveragePooling2D
MaxPool2D
ReLu
BatchNorm
Conv2D
MaxPool2D
ReLu
BatchNorm
Conv2D

- Ioffe & Szegedy (2015) *arXiv: 1502.03167v3*
- Li et al (2018) *arXiv: 1801.05134v1*

Model Architecture



- Ioffe & Szegedy (2015) *arXiv: 1502.03167v3*
- Li et al (2018) *arXiv: 1801.05134v1*

Hyperparameter Exploration

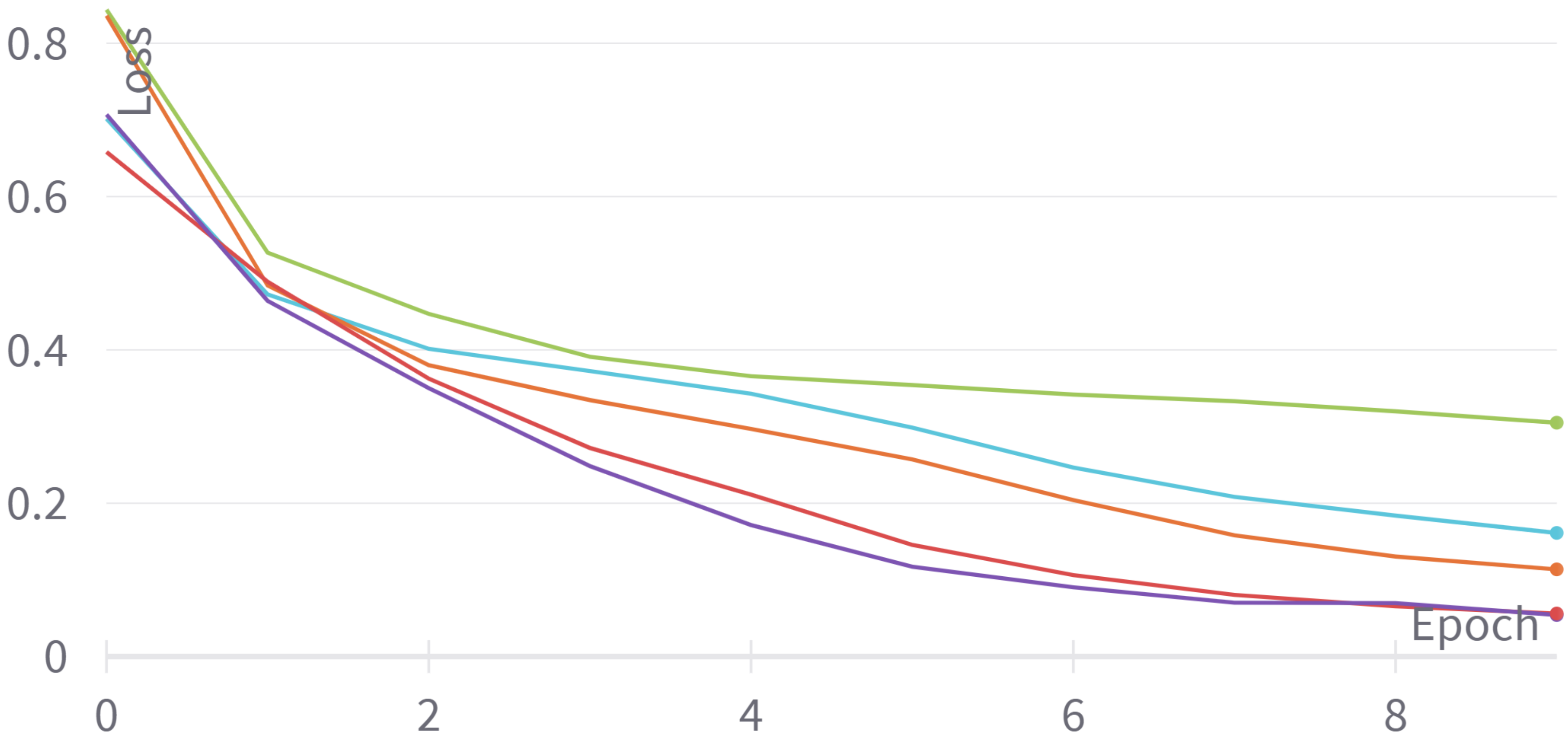
Hyperparameter Values Explored

C1_filterSize	3, 5, 7
C1_numFilters	16, 32, 64
C1_stride	1, 2
P1_filterSize	2
P1_stride	2
C2_filterSize	3, 5
C2_numFilters	64, 128
C2_stride	1, 2
P2_filterSize	2
P2_stride	2
FC1_numUnits	32, 64
FC2_numUnits	8
Total Combinations	288

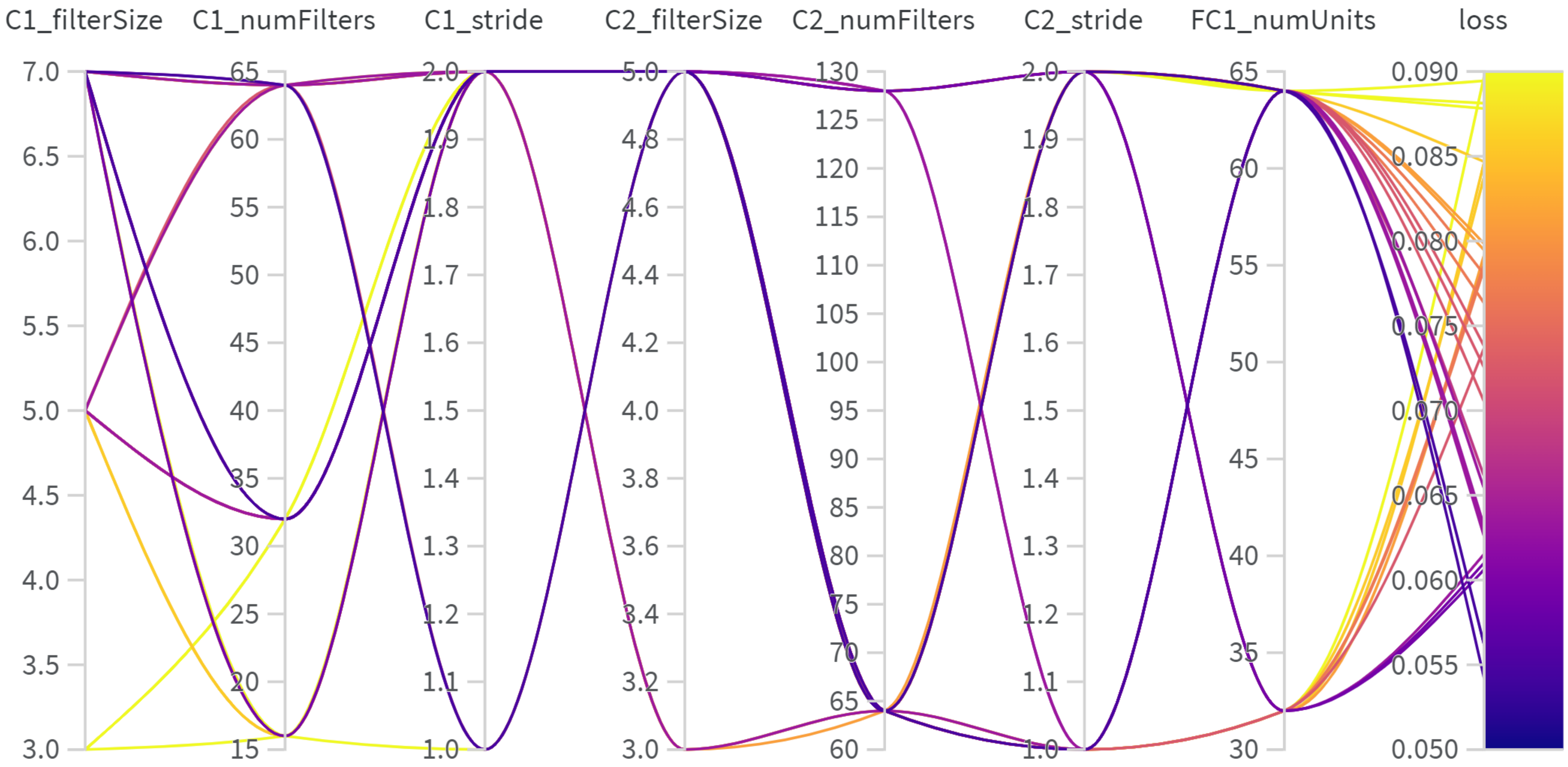
Training Parameters

Optimizer	Adam (parm vals)
Batch Size	512
Learning Rate	0.01
Epochs	10

Hyperparameter Exploration: Selected Training Curves



Hyperparameter Exploration: Loss v.s. HP values



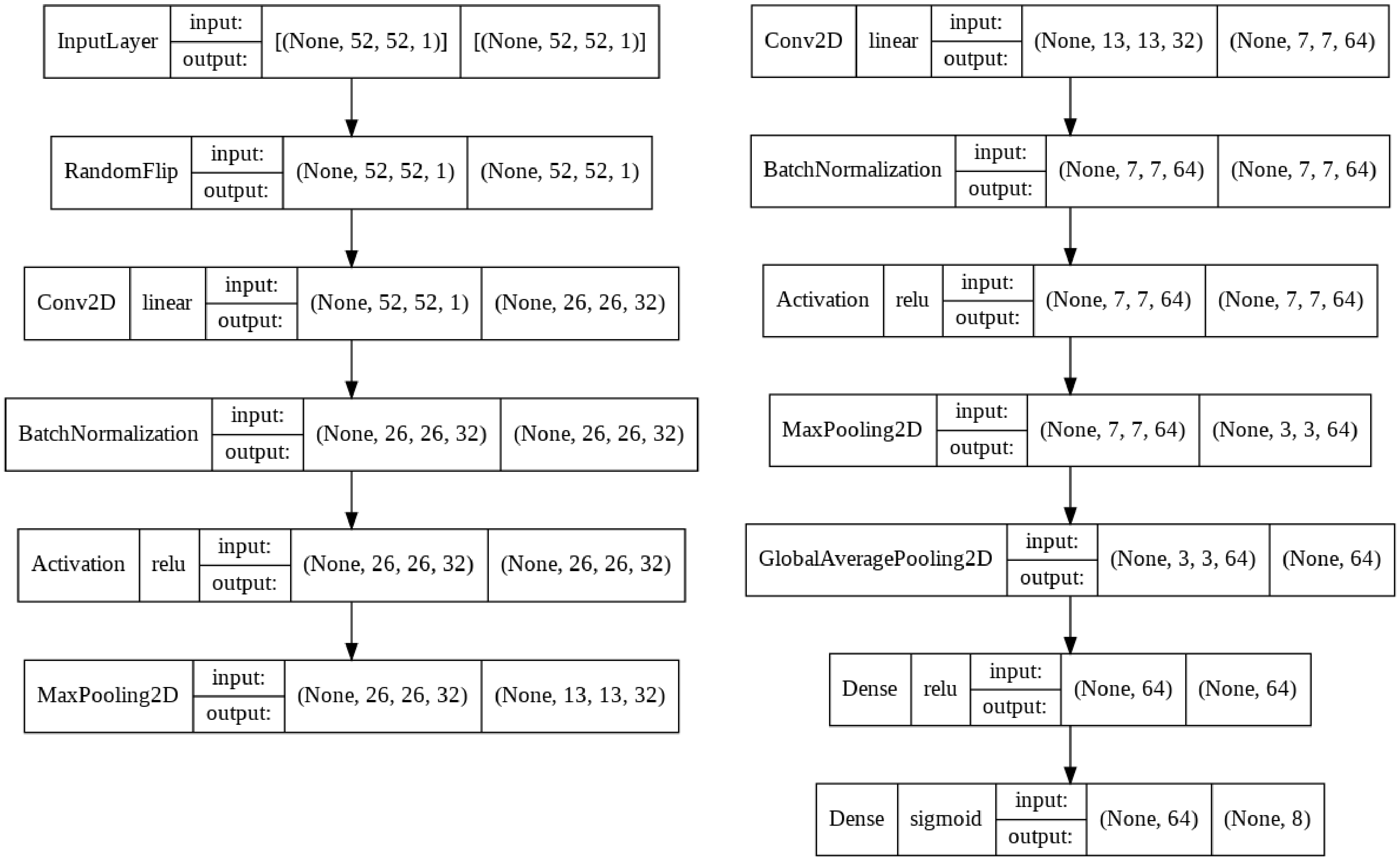
Hyperparameter Exploration: Selected HP Values

The 2 Lowest Loss Hyperparameter Combinations

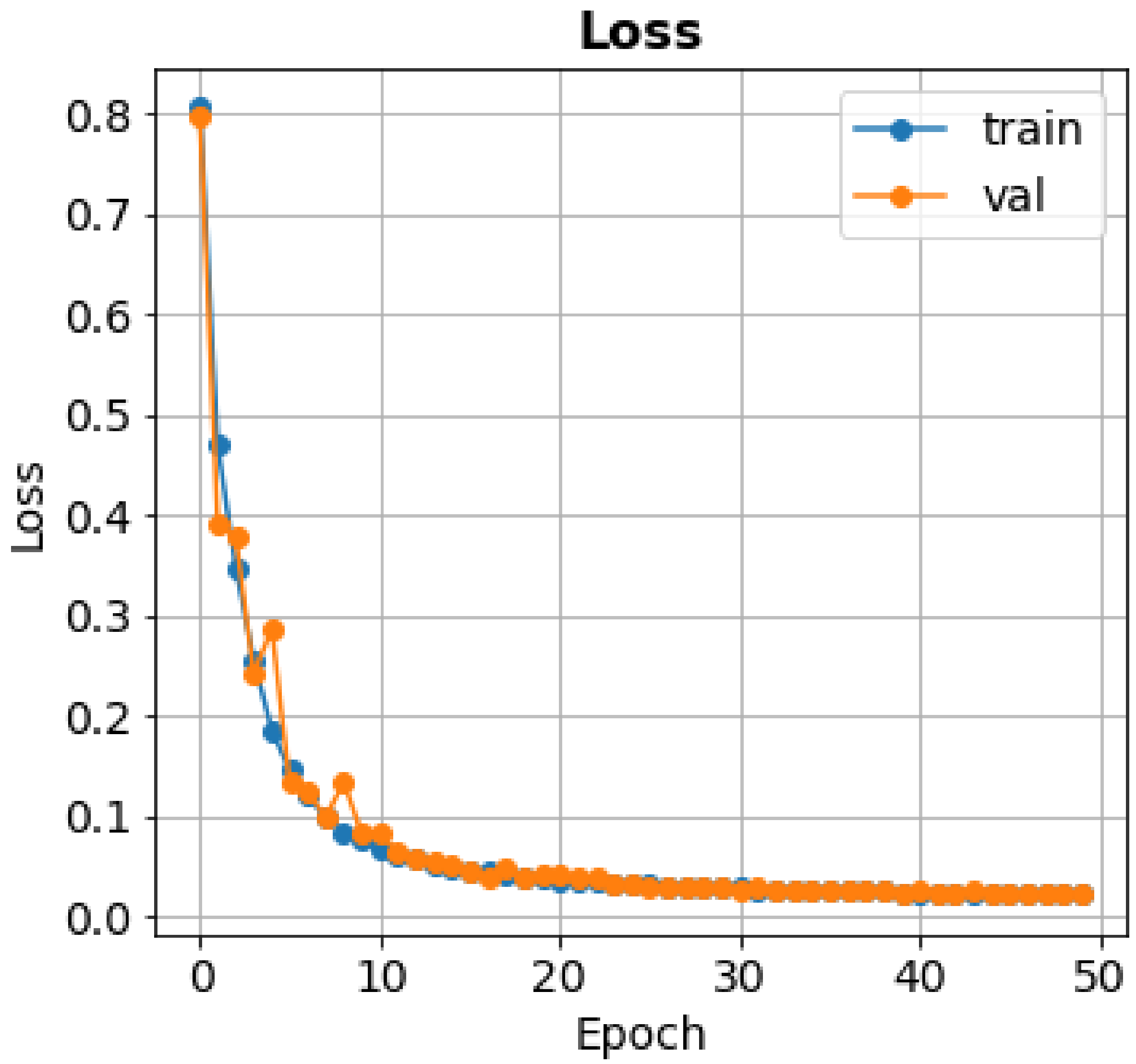
Loss	0.0541	0.0561
C1_filterSize	7	7
C1_numFilters	64	32
C1_stride	2	2
P1_filterSize	2	2
P1_stride	2	2
C2_filterSize	5	5
C2_numFilters	64	64
C2_stride	2	2
P2_filterSize	2	2
P2_stride	2	2
FC1_numUnits	64	64
FC2_numUnits	8	8
Total Parameters	110,472	57,460

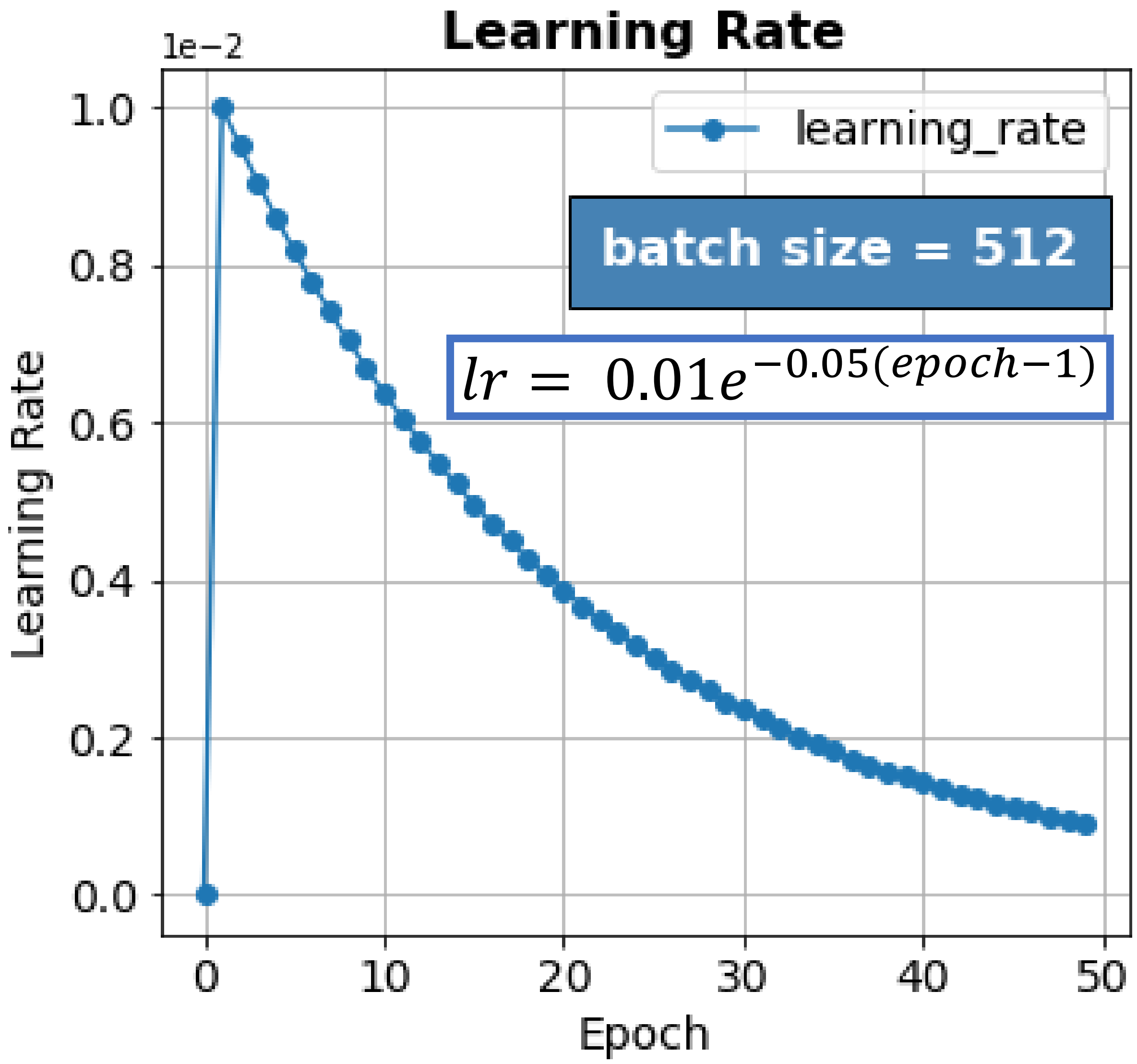
Selected This One

Final Model Shapes



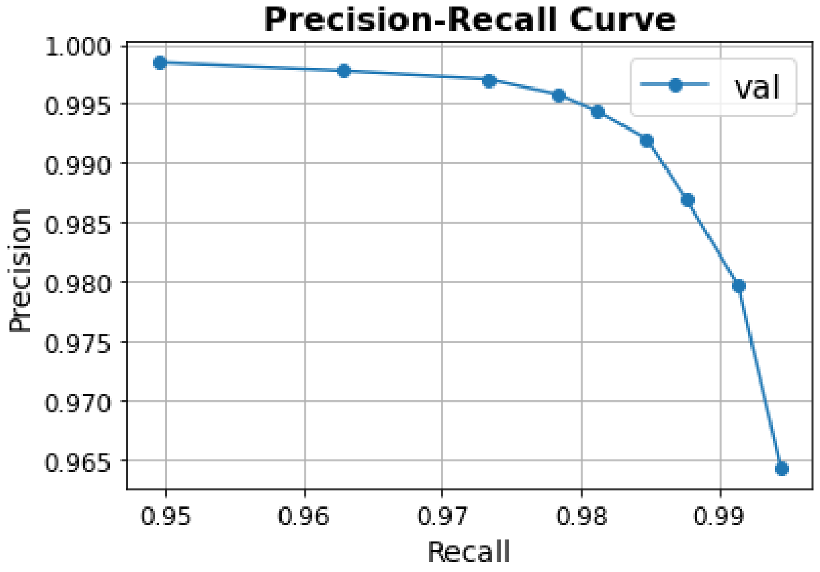
Full Training





Performance

	TRAIN	VAL
ACCURACY	99.3%	99.3%
RECALL	98.3%	98.1%
PRECISION	99.4%	99.4%
F1	98.9%	98.8%



VAL RECALL

Defect Type	1	2	3	4	Any
Any	98.6	98.8	97.8	97.6	
Center	100	100	99.2	100	99.6
Donut	100	100	99.7	100	99.9
Edge-Local	93.8	98.8	95.7	93.8	96.2
Edge-Ring	100	99.6	99.4	99.2	99.5
Local	96.9	97.5	97.5	97.7	97.5
Near-Full	100	NA	NA	NA	100
Scratch	100	97.5	96.6	96.1	96.9
Random	100	NA	NA	NA	100

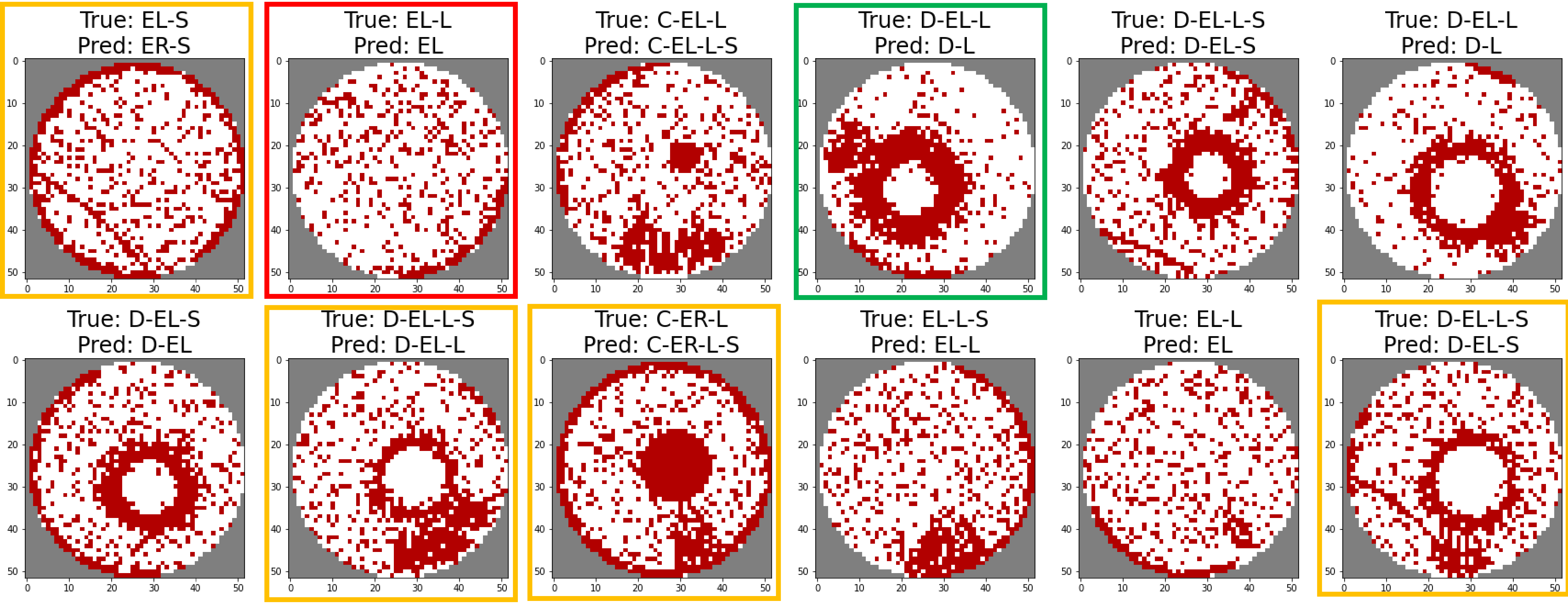
\# Defects in Wafer

VAL PRECISION

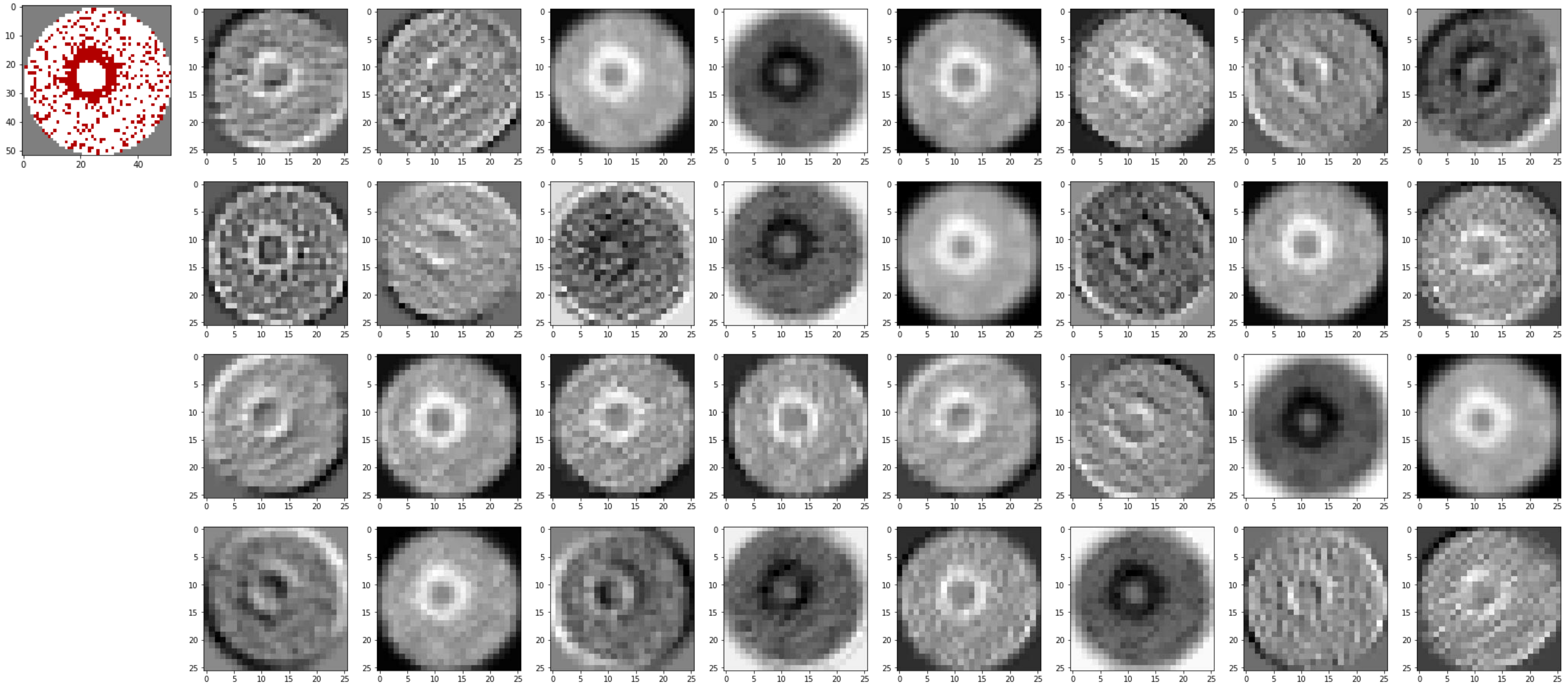
Defect Type	1	2	3	4	Any
Any	98.9	99.4	99.5	99.7	
Center	100	100	100	100	100
Donut	100	100	100	100	100
Edge-Local	98.4	99.6	99.4	100	99.5
Edge-Ring	98.5	99.6	98.5	97.7	98.7
Local	98.4	100	99.8	100	99.8
Near-Full	90.0	NA	NA	NA	90.0
Scratch	98.5	97.5	99.1	100	98.8
Random	100	NA	NA	NA	100

\# Defects in Wafer

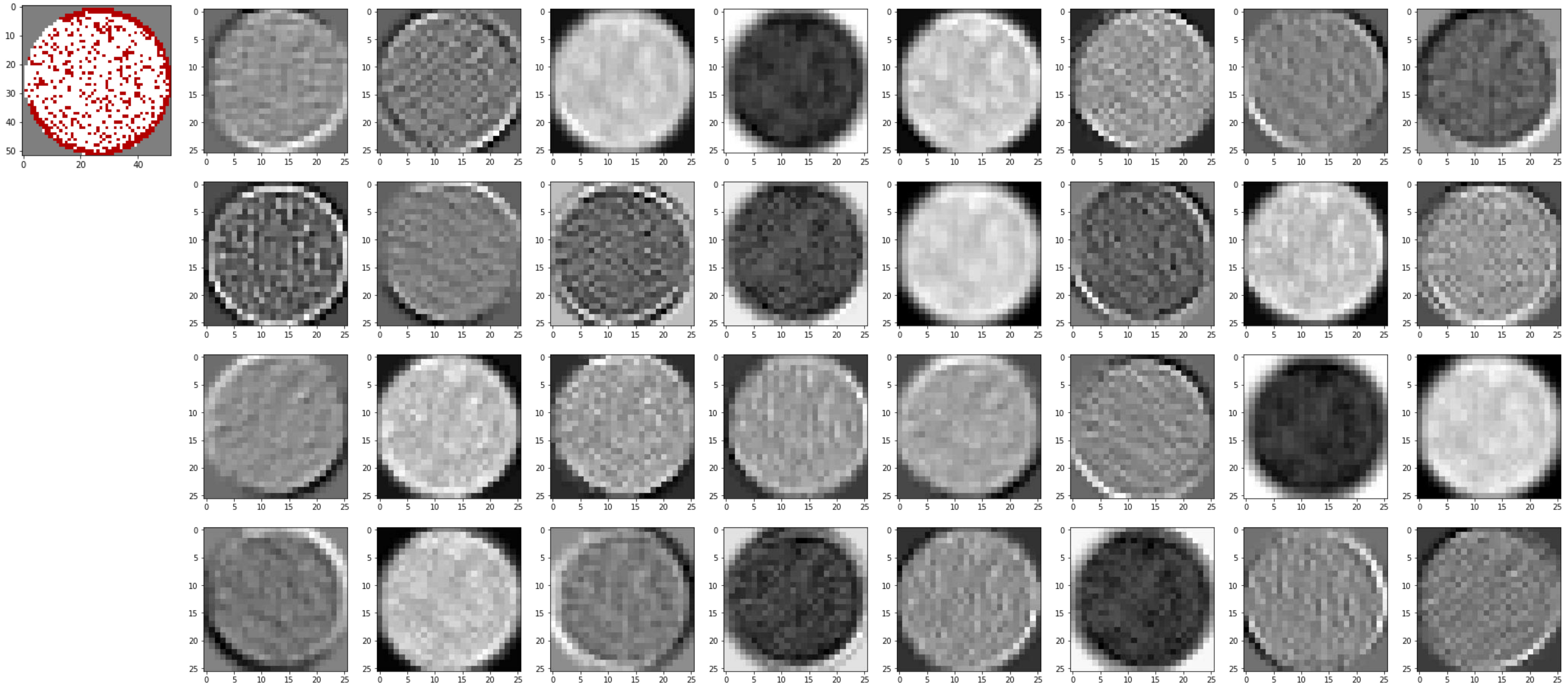
Examples of Mis-labeled Wafers



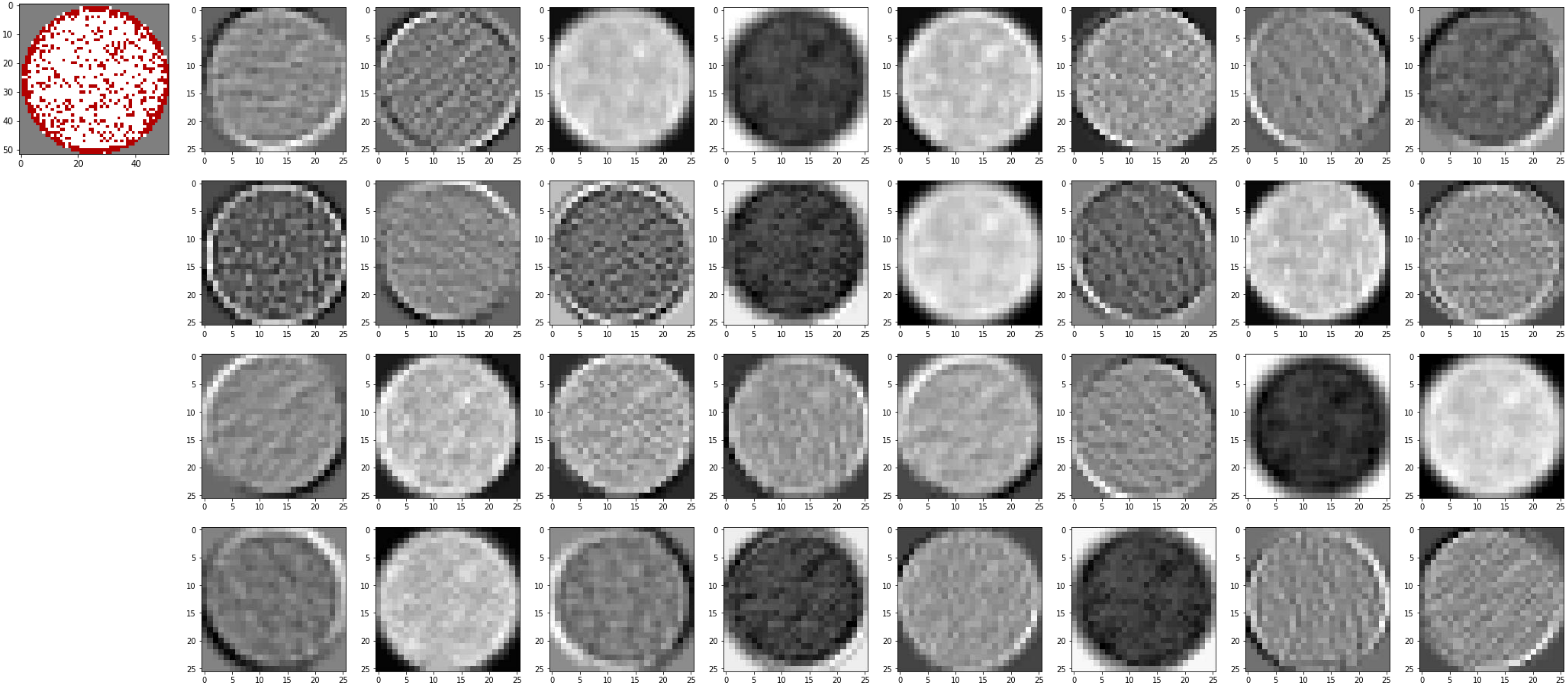
Feature Maps from the 1st Conv Layer: DONUT



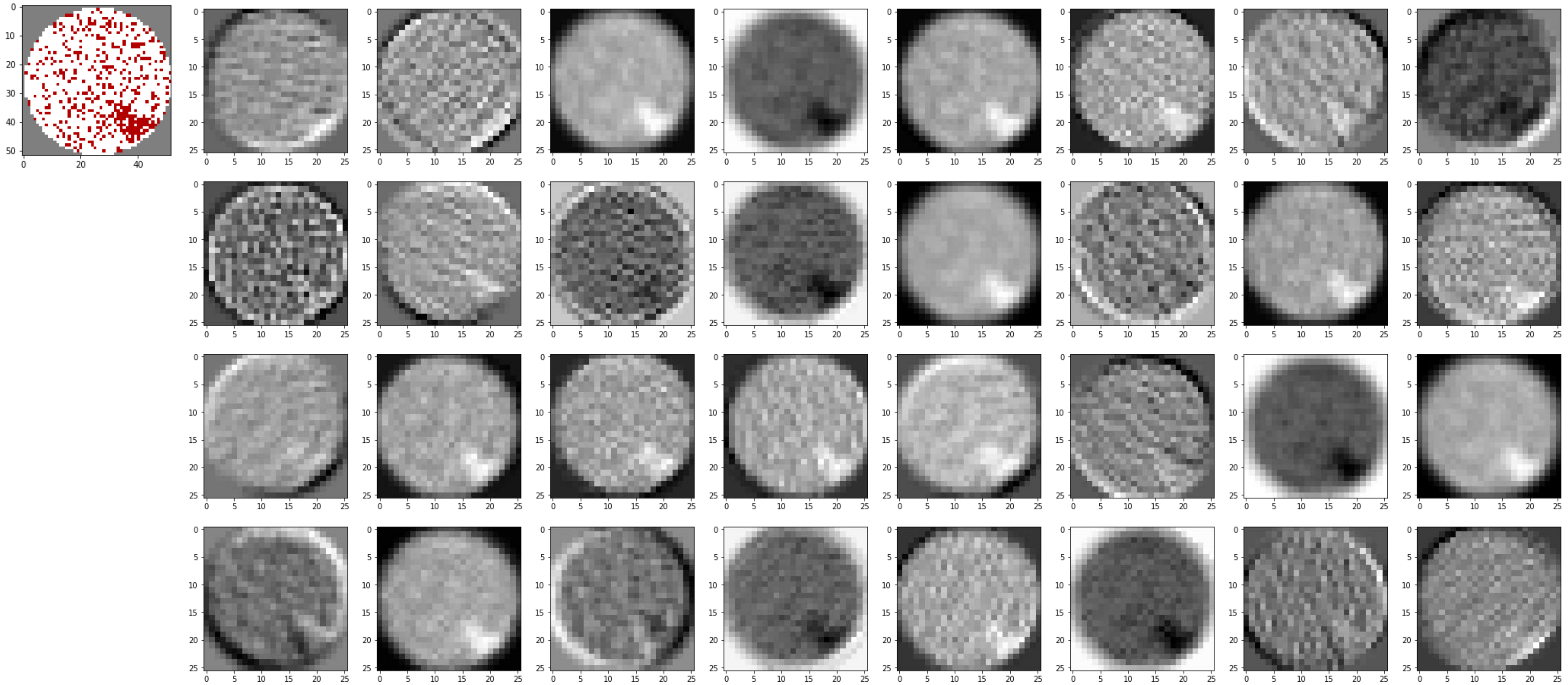
Feature Maps from the 1st Con Layer: EDGE-LOCAL



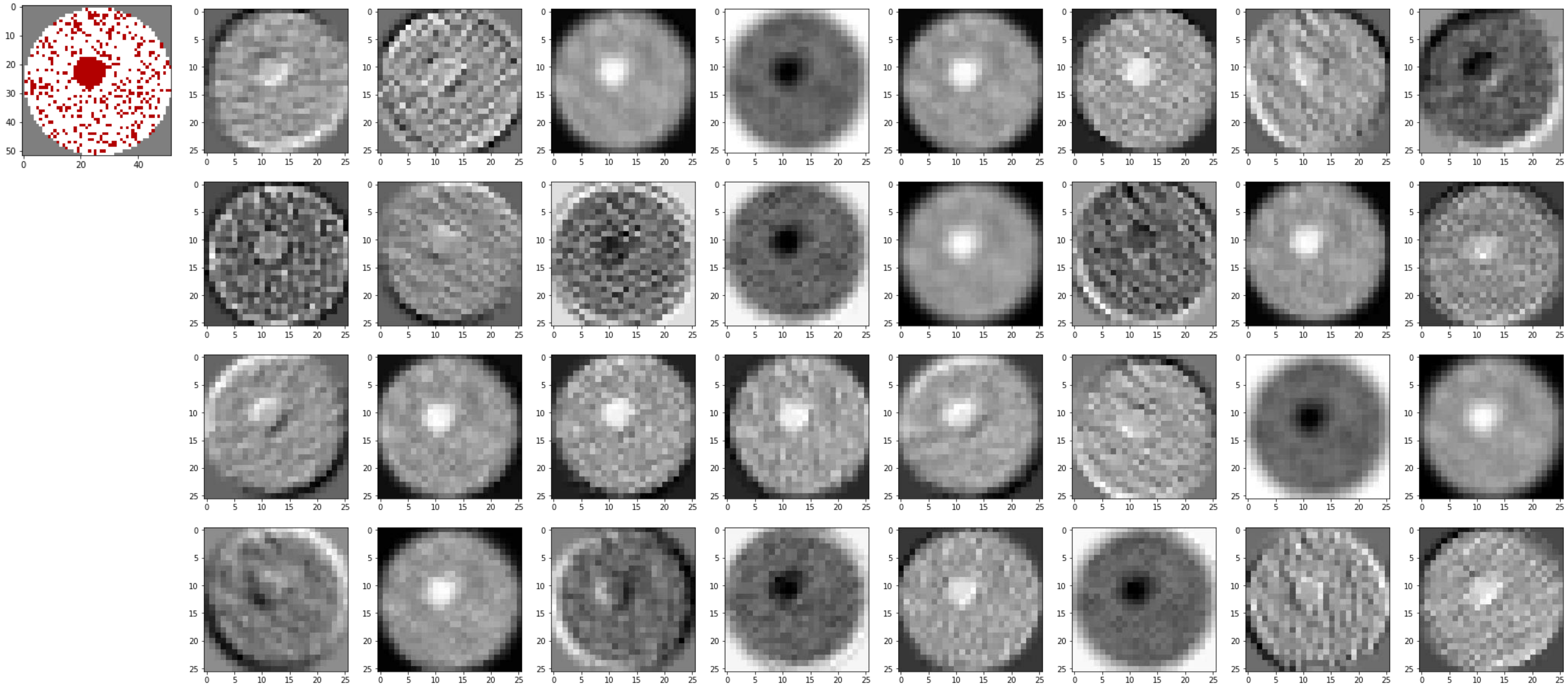
Feature Maps from the 1st Con Layer: EDGE-RING



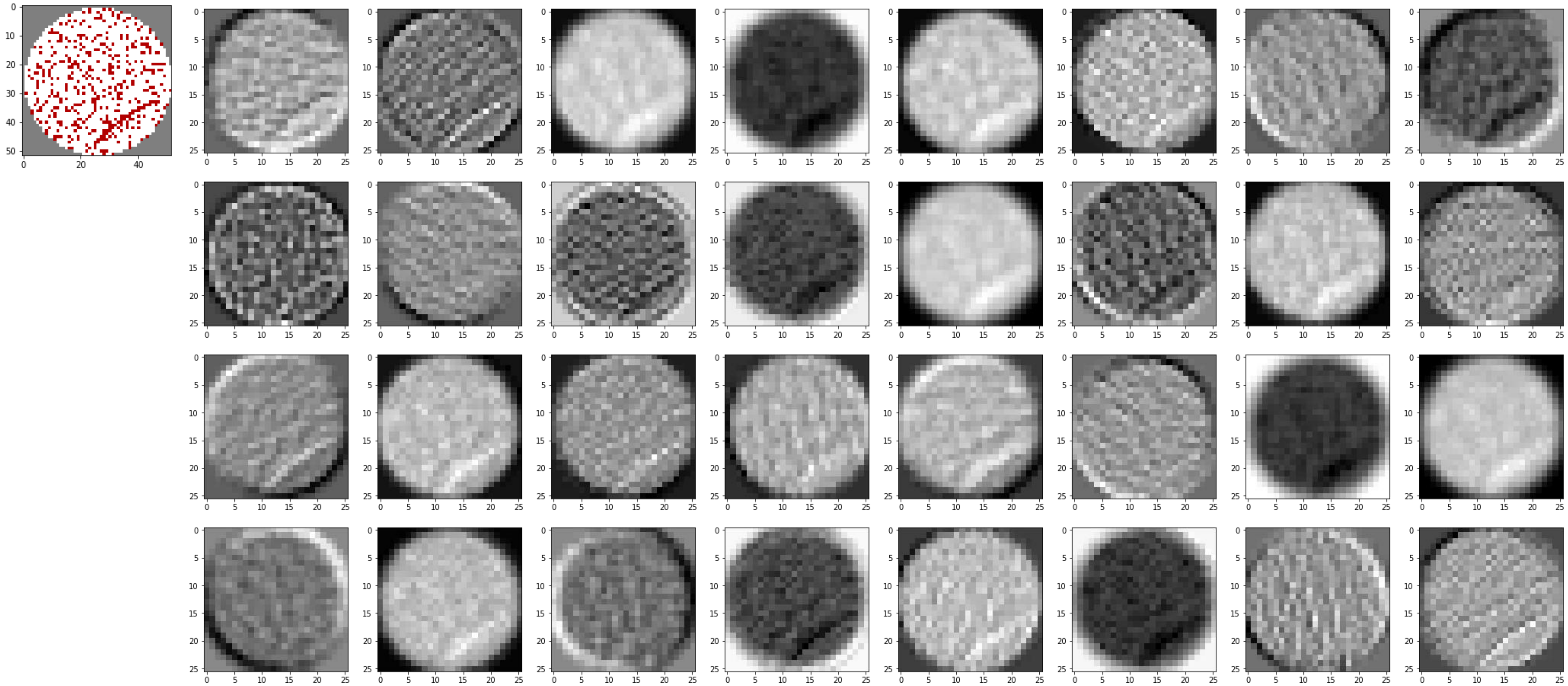
Feature Maps from the 1st Con Layer: LOCAL



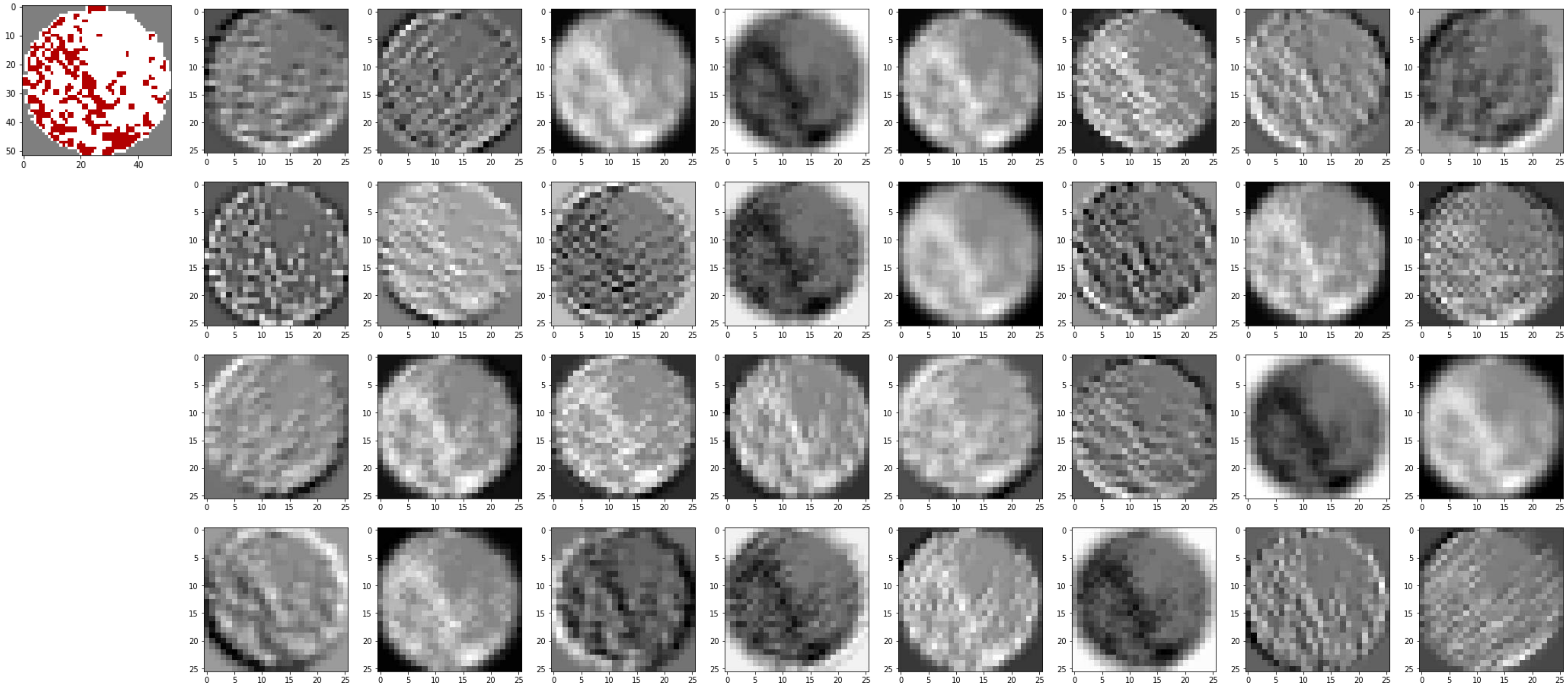
Feature Maps from the 1st Con Layer: CENTER



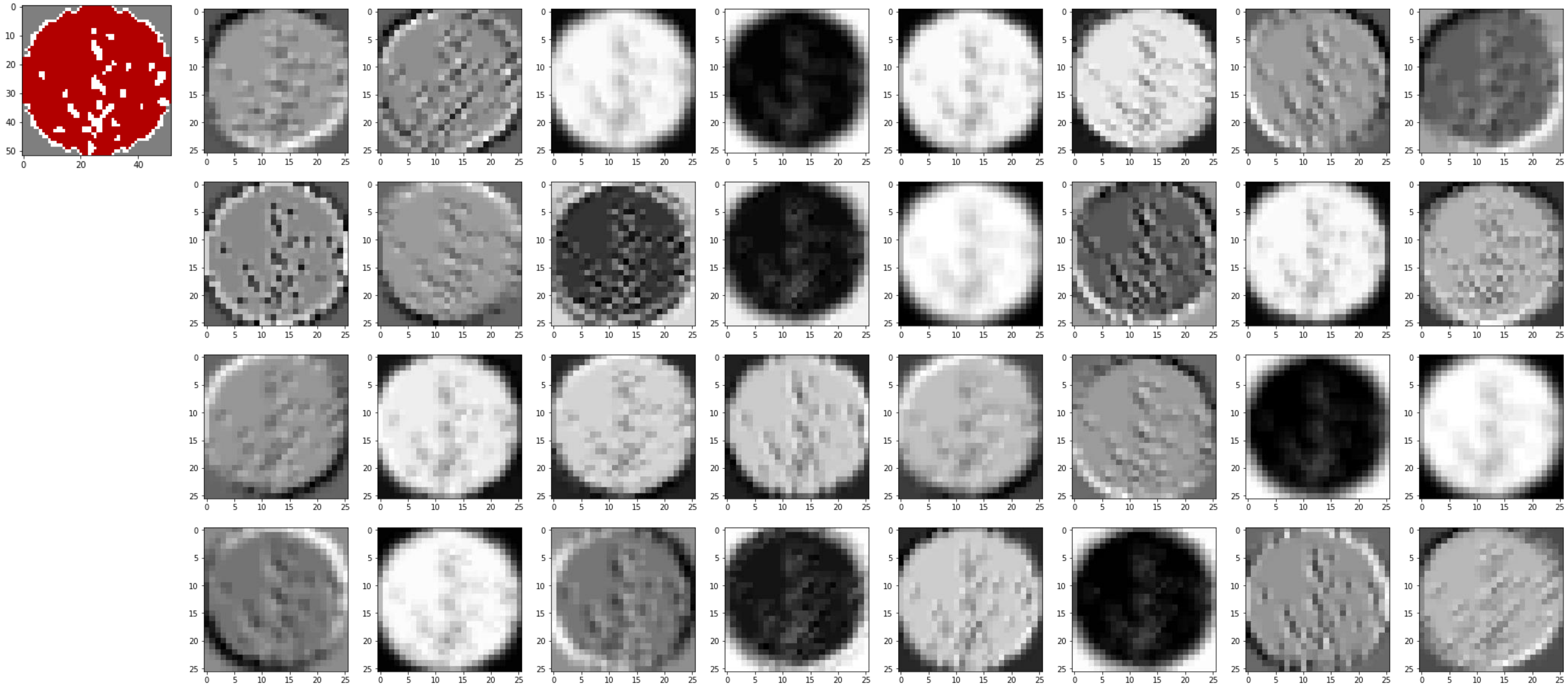
Feature Maps from the 1st Con Layer: SCRATCH



Feature Maps from the 1st Con Layer: RANDOM

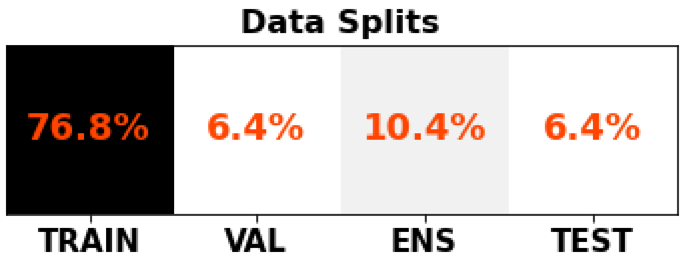


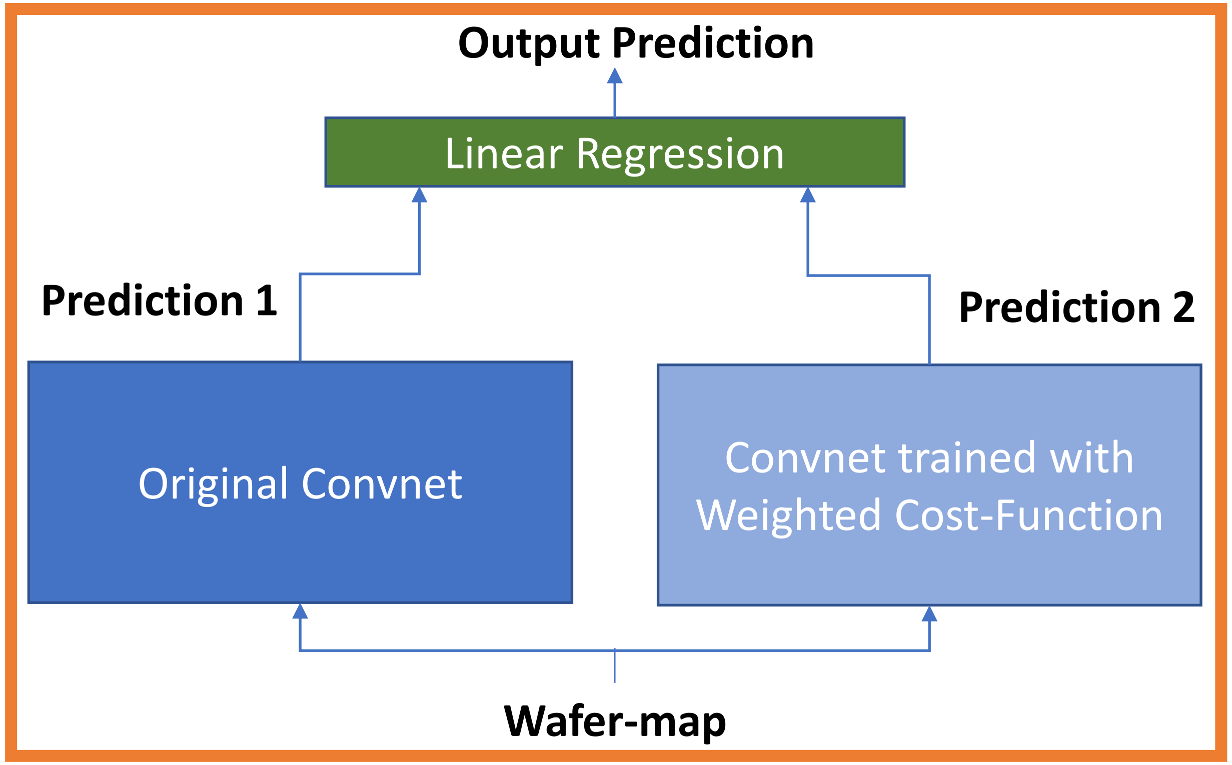
Feature Maps from the 1st Con Layer: NEAR-FULL



Weighted Cost + Ensemble Model

- Train a second model in which wafers the first model mis-predicts are weighted higher in the loss.
- This weighting is applied such that:
  - $Total\ weight\ of\ inaccurately\ predicted\ wafers = \alpha \times Total\ weight\ of\ accuractely\ predicted\ wafers$





$$weight = \left(\mathbb{1}_{pred=true} + \mathbb{1}_{pred\neq true} \frac{\alpha \sum \mathbb{1}_{pred=true}}{\sum \mathbb{1}_{pred\neq true}} \right) \frac{(1+\alpha)\sum \mathbb{1}_{pred=true}}{\sum \mathbb{1}_{pred=true} + \sum \mathbb{1}_{pred\neq true}}$$