

**num\_iteration**

**train accuracy**

2000/

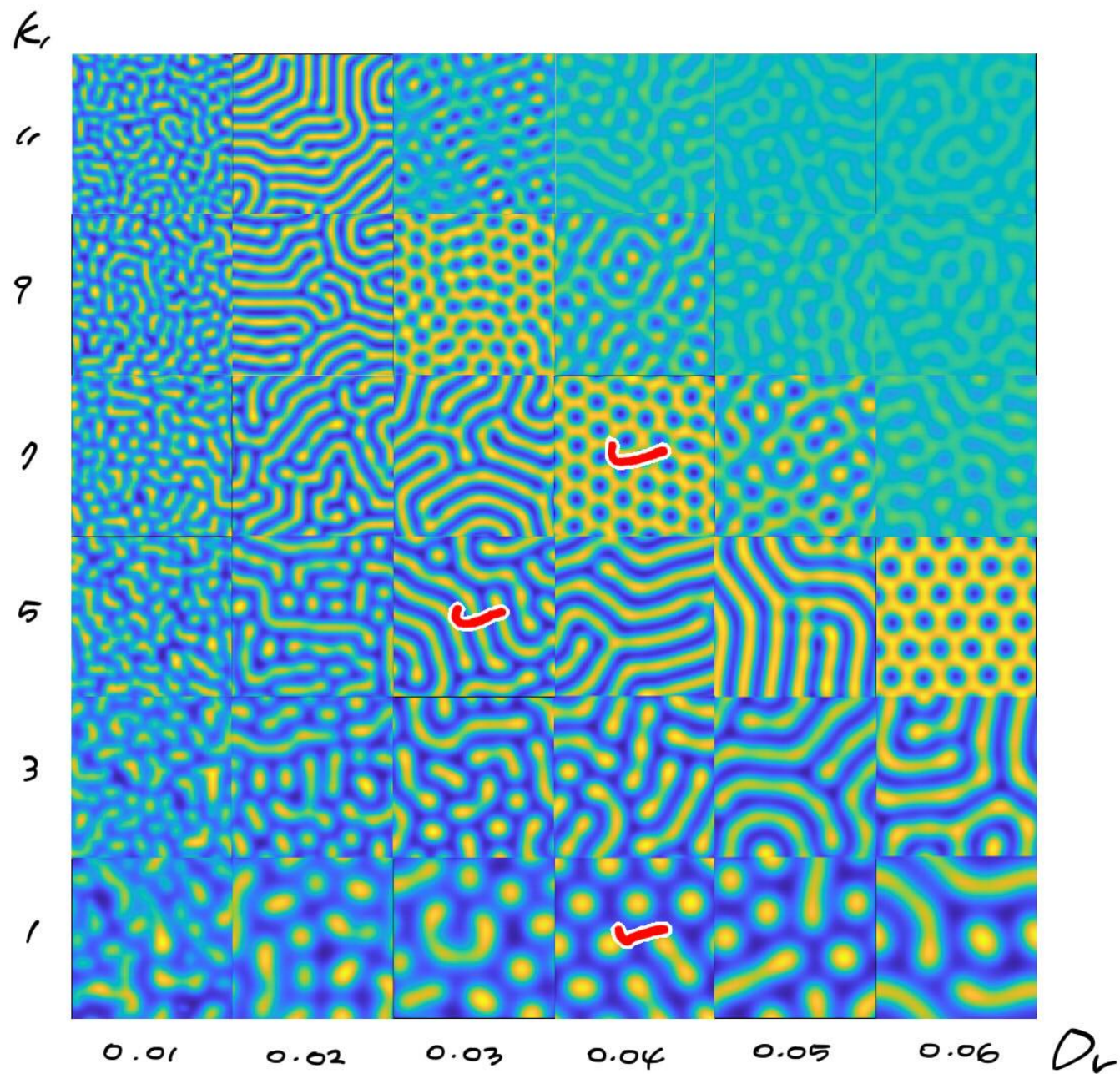
1.0

0.08/

0.34

**learning\_rate**

**test accuracy**



Dissimilar	train :100 all : 144		train : 250 all : 360		train : 500 all : 720		train :1000 all : 1440	
1) Cnn	100 /	1.0 1.0	100 /	1.0 1.0	100 /	1.0 1.0	100 /	1.0 1.0
2) 1-layer with GD	2000/ 0.005/	0.99 0.43	2000/ 0.005/	0.78 0.36	2000/ 0.004/	0.67 0.36	2000/ 0.004/	0.60 0.43
3) 1layer with Adam	2000/ 0.1/	1.0 0.47	2000/ 0.1/	1.0 0.40	2000/ 0.03/	0.79 0.36	2000/ 0.01/	0.71 0.43
4) Derivative layer with GD	2000/ 10/	1.0 0.36	2000/ 10/	1.0 0.37	2000/ 10/	1.0 0.40	2000/ 10/	0.96 0.34
5) Derivative layer with Adam	2000/ 1/	1.0 0.31	2000/ 0.9/	1.0 0.37	2000/ 0.9/	1.0 0.36	2000/ 0.9/	1.0 0.34
6) 2weight with GD	2000/ 0.005/	1.0 0.43	2000/ 0.005/	0.79 0.43	2000/ 0.004/	0.68 0.42	2000/ 0.004/	0.58 0.43
7) 2weight with Adam	2000/ 0.07/	1.0 0.38	2000/ 0.06/	1.0 0.41	2000/ 0.06/	1.0 0.44	2000/ 0.06/	0.96 0.40
8) 2weight & $u-u^3$ with GD	2000/ 0.001/	1.0 0.77	2000/ 0.001/	0.96 0.82	2000/ 0.001/	0.95 0.90	2000/ 0.001/	0.94 0.89
9) 2weight & $u-u^3$ with Adam	600/ 0.04/	1.0 0.88	600/ 0.04/	1.0 0.90	600/ 0.04/	1.0 0.90	600/ 0.04/	1.0 0.90

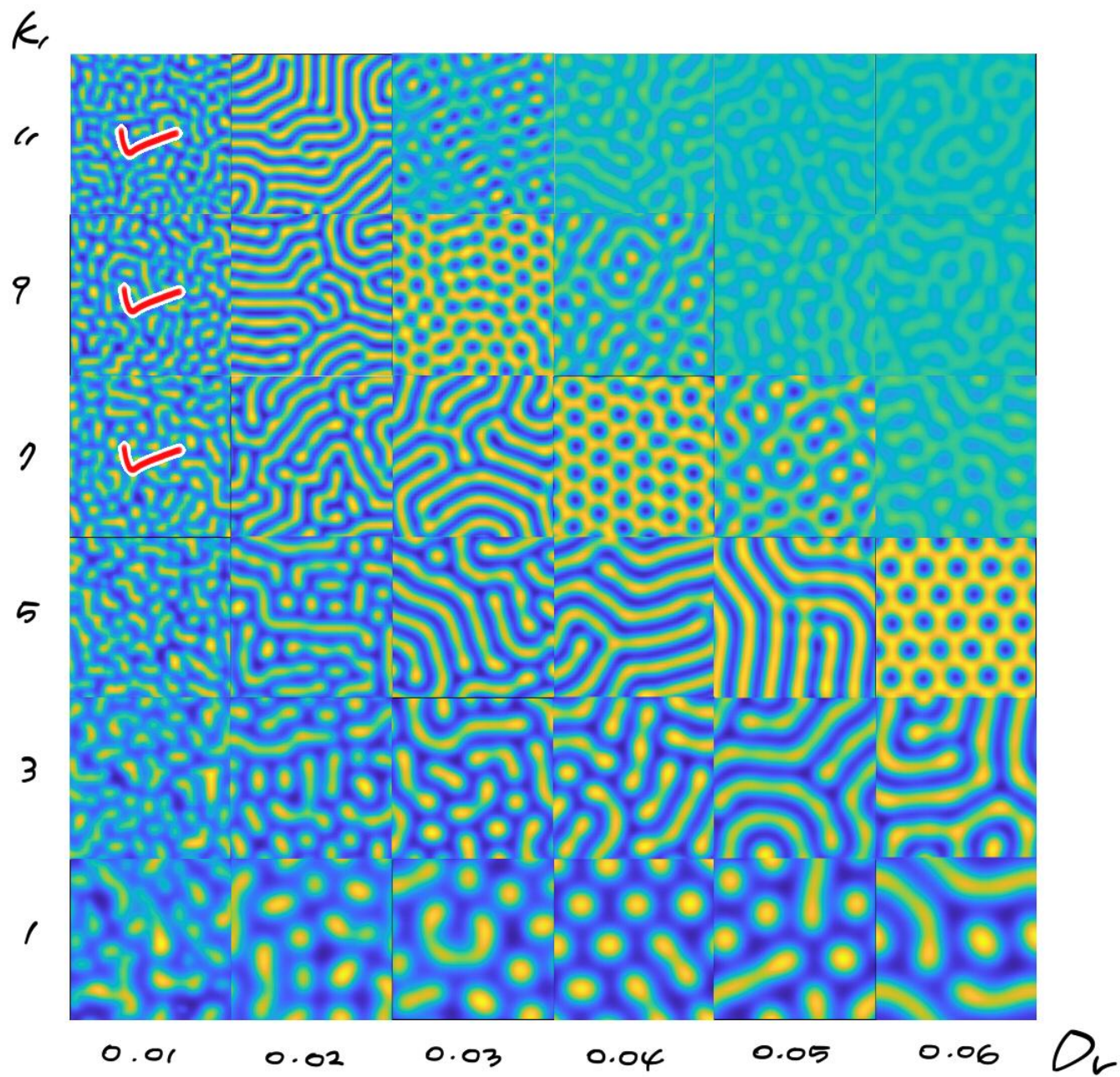
10) $u-u^3$ with GD	2000/ 0.001/	1.0 0.77	2000/ 0.001/	96 0.82	2000/ 0.001/	95 0.90	2000/ 0.001/	94 0.89
11) $u-u^3$ with Adam	600/ 0.04/	1.0 0.88	600/ 0.04/	1.0 0.90	600/ 0.04/	1.0 0.90	600/ 0.04/	1.0 0.90

*# Forward Propagation*

$Z = \text{np.dot}(W1, \text{gradient\_vec}(X)) + \text{np.dot}(W2, X - X^{**3}) + b$

$A = \text{softmax}(Z)$     *# compute activation*





Similar	train :100 all : 144		train : 250 all : 360		train : 500 all : 720	
1) Cnn	100 /	1.0 0.54	100 /	1.0 0.65	100 /	1.0 0.65
2) 1-layer with GD	2000/ 0.005/	1.0 0.22	2000/ 0.005/	0.96 0.34	2000/ 0.005/	0.86 0.33
3) 1layer with Adam	2000/ 0.08/	1.0 0.34	2000/ 0.08/	1.0 0.36	2000/ 0.08/	1.0 0.31
4) Derivative layer with GD	2000/ 10/	1.0 0.36	2000/ 10/	1.0 0.32	2000/ 10/	1.0 0.37
5) Derivative layer with Adam	2000/ 1/	1.0 0.34	2000/ 1/	1.0 0.32	2000/ 1/	1.0 0.42
6) 2weight with GD	2000/ 0.005/	1.0 0.40	2000/ 0.005/	0.97 0.33	2000/ 0.005/	0.86 0.36
7) 2weight with Adam	2000/ 0.07/	1.0 0.38	2000/ 0.07/	1.0 0.37	2000/ 0.07/	1.0 0.46
8) 2weight & $u-u^3$ with GD	2000/ 0.001/	1.0 0.36	2000/ 0.001/	0.97 0.24	2000/ 0.001/	0.82 0.32
9) 2weight & $u-u^3$ with Adam	2000/ 0.04/	1.0 0.31	2000/ 0.04/	1.0 0.37	2000/ 0.04/	1.0 0.37