

Supplementary Information

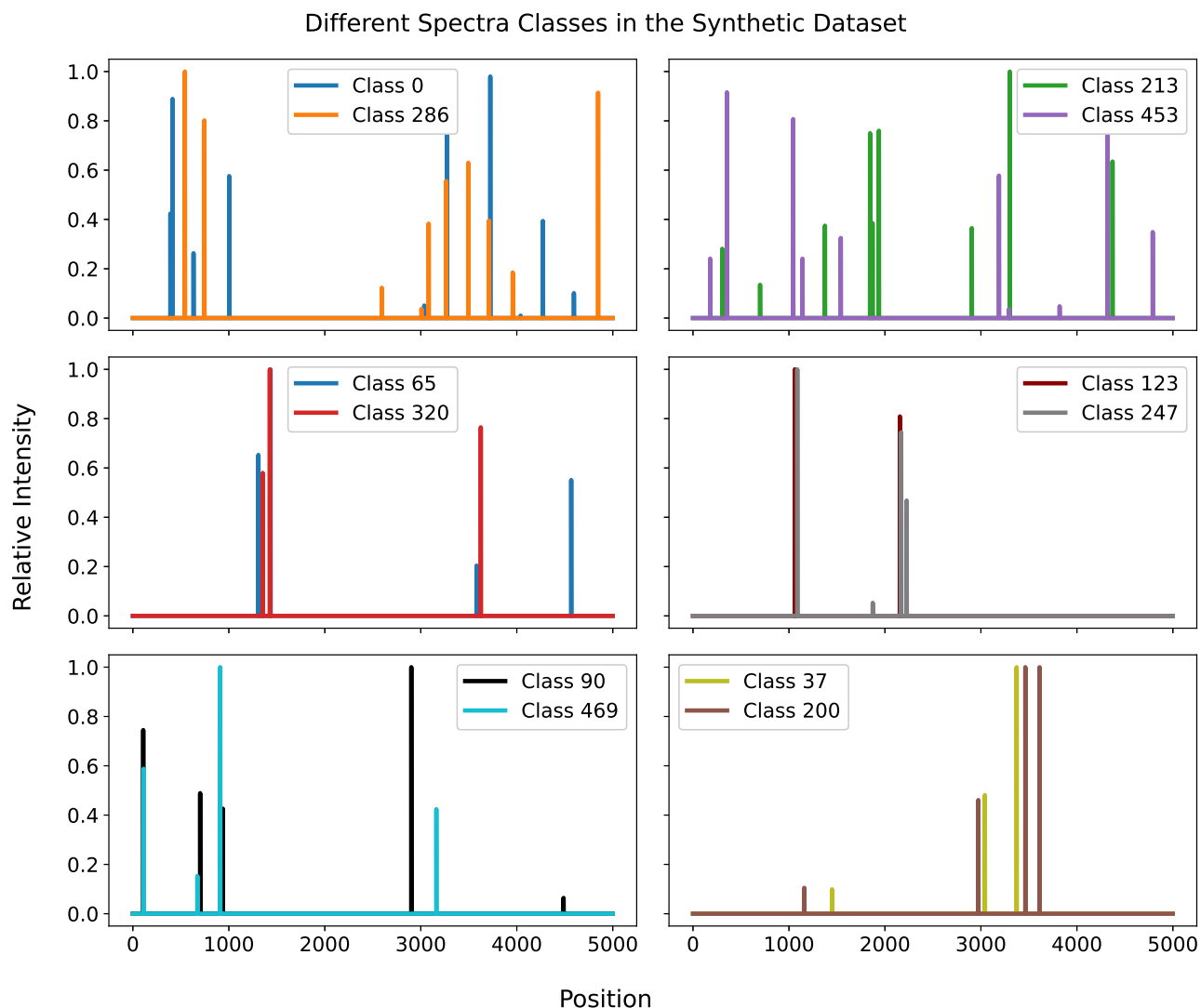
Validating neural networks for spectroscopic classification on a universal synthetic dataset

Jan Schuetzke¹, Nathan J. Szymanski^{2,3}, Markus Reischl¹

¹Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, 76131, Germany

²Department of Materials Science & Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

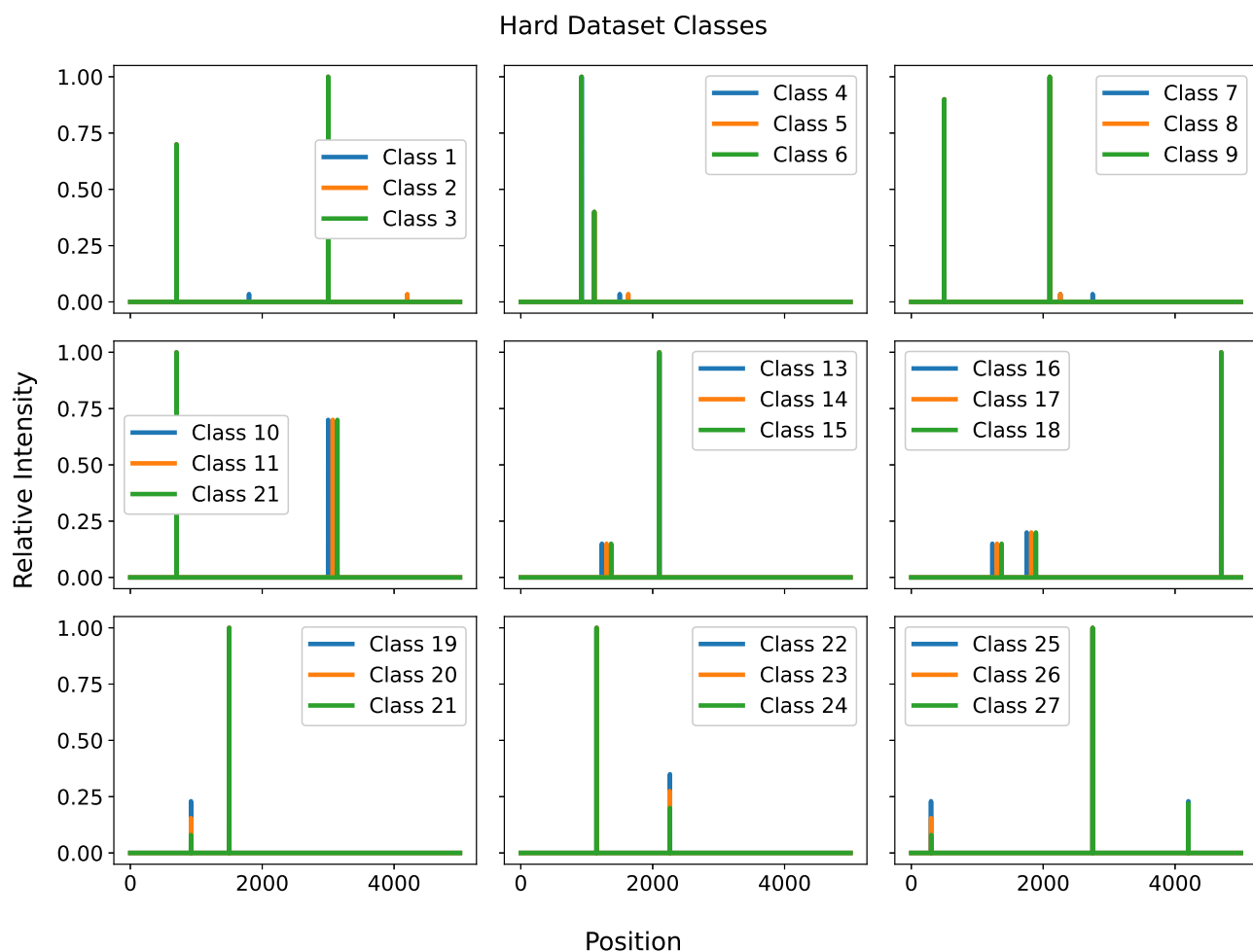
³Department of Materials Science & Engineering, UC Berkeley, Berkeley, CA, 94720, USA



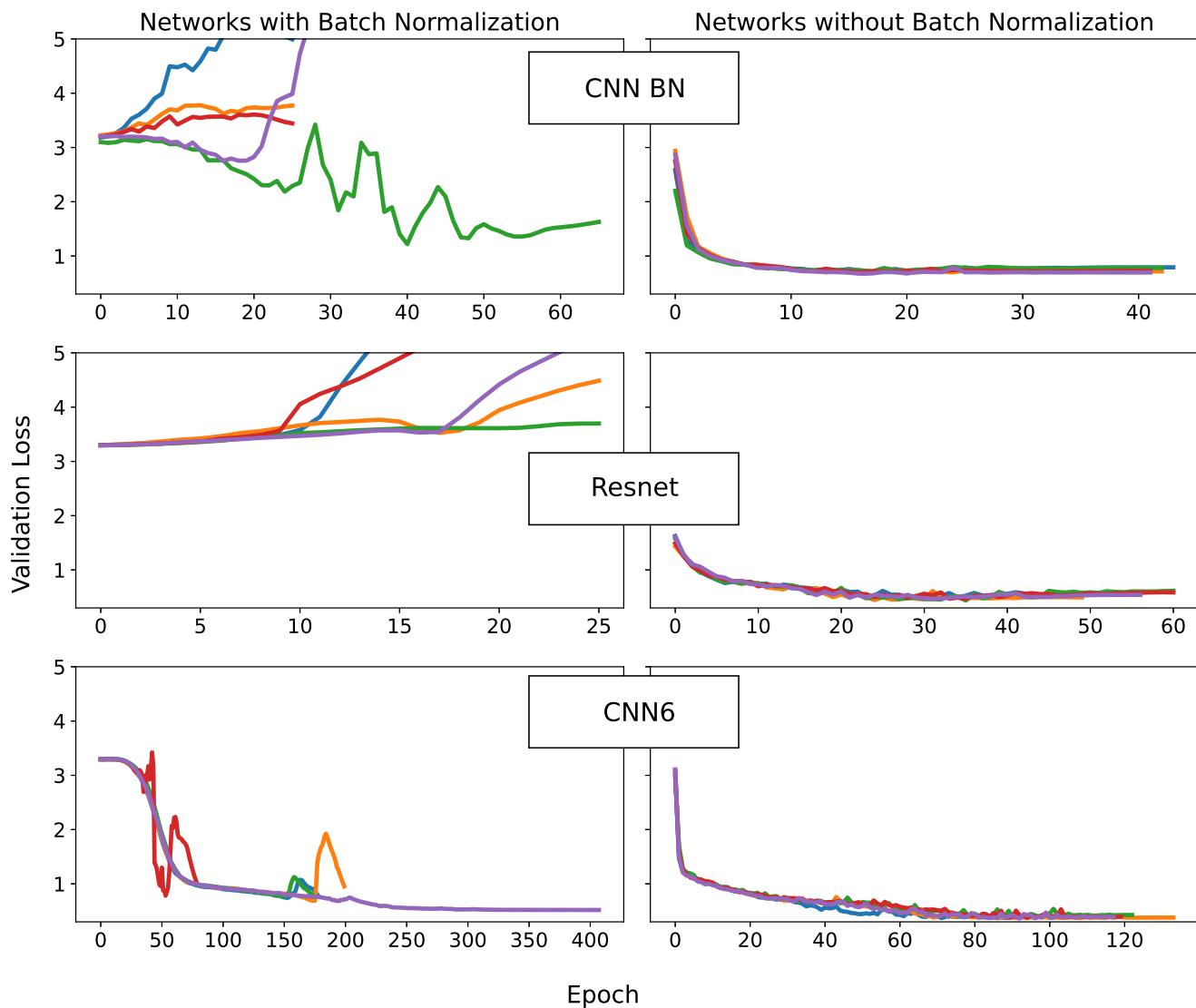
Supplementary Figure 1. Selected classes of synthetic spectra from our 500-class dataset. We have grouped similar spectra and present 6 pairs of visually similar patterns. The first row displays patterns with a high number of peaks, some of which are aligned in positions and relative heights. The second row illustrates spectra with peaks that are close to others but still important for the classification. To effectively separate such pairs, it is crucial to avoid suppressing the minor peak in the convolutional layers. The third row shows spectra where either some positions or the relative heights are mostly identical.

Supplementary Table 1. Full details on the accuracy scores on the general benchmark for each model and seed (initialization).

Model Name	Seed	Training Set		Validation Set		Test Set	
		Accuracy	Misclassifications	Accuracy	Misclassifications	Accuracy	Misclassifications
CNN2	0	99.740%	65	99.100%	45	99.244%	34
CNN2	1	99.608%	98	99.040%	48	99.133%	39
CNN2	2	99.820%	45	99.340%	33	99.311%	31
CNN2	3	99.788%	53	99.020%	49	98.889%	50
CNN2	4	99.600%	100	98.940%	53	99.200%	36
CNN3	0	99.660%	85	99.600%	20	99.244%	34
CNN3	1	99.864%	34	99.380%	31	99.400%	27
CNN3	2	99.888%	28	99.500%	25	99.311%	31
CNN3	3	99.848%	38	99.520%	24	99.444%	25
CNN3	4	99.656%	86	99.560%	22	99.444%	25
CNN6	0	99.500%	125	99.760%	12	99.867%	6
CNN6	1	99.316%	171	99.840%	8	99.822%	8
CNN6	2	99.336%	166	99.780%	11	99.800%	9
CNN6	3	99.380%	155	99.800%	10	99.733%	12
CNN6	4	99.260%	185	99.760%	12	99.844%	7
VGG	0	99.708%	73	99.380%	31	99.578%	19
VGG	1	99.760%	60	99.360%	32	99.311%	31
VGG	2	99.740%	65	99.380%	31	99.422%	26
VGG	3	99.700%	75	99.360%	32	99.644%	16
VGG	4	99.672%	82	99.460%	27	99.644%	16
Resnet	0	99.848%	38	99.540%	23	99.511%	22
Resnet	1	99.892%	27	99.620%	19	99.556%	20
Resnet	2	99.884%	29	99.620%	19	99.556%	20
Resnet	3	99.860%	35	99.600%	20	99.511%	22
Resnet	4	99.892%	27	99.600%	20	99.511%	22
CNN BN	0	99.752%	62	99.440%	28	99.000%	45
CNN BN	1	99.864%	34	99.380%	31	99.267%	33
CNN BN	2	99.740%	65	99.420%	29	99.022%	44
CNN BN	3	99.916%	21	99.260%	37	98.911%	49
CNN BN	4	99.856%	36	99.420%	29	98.889%	50
INC3	0	99.844%	39	99.600%	20	99.733%	12
INC3	1	99.848%	38	99.640%	18	99.600%	18
INC3	2	99.676%	81	99.480%	26	99.556%	20
INC3	3	99.836%	41	99.560%	22	99.578%	19
INC3	4	99.668%	83	99.620%	19	99.622%	17
INC6	0	99.796%	51	99.740%	13	99.733%	12
INC6	1	99.840%	40	99.700%	15	99.778%	10
INC6	2	99.740%	65	99.720%	14	99.756%	11
INC6	3	99.792%	52	99.640%	18	99.756%	11
INC6	4	99.808%	48	99.680%	16	99.711%	13



Supplementary Figure 2. Visual representation of our 27 synthetic spectra classes to evaluate the neural network architectures according to the three identified misclassification causes. The first row consists of three triplets distinguishable by an additional minor peak. The second row contains 9 classes with similar peak positions and slightly different peak positions, and the last row depicts the three classes with overlapping peak intensities.



Supplementary Figure 3. Evaluation of the Batch-Normalization layer in select networks. Thus, we modified the model architectures to allow activating or deactivating the Batch-Normalization process in the CNN BN, CNN6, and Resnet models. We trained 5 different initializations for each model, both with and without Batch-Normalization. While the validation loss for the models without the normalization operations converges smoothly, fluctuations are observed for models with active Batch-Normalization layers.