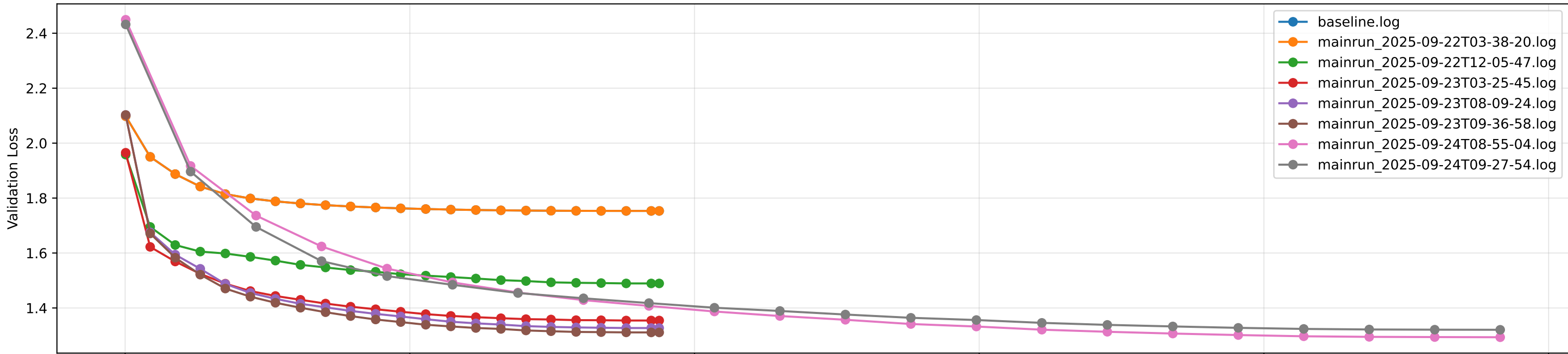
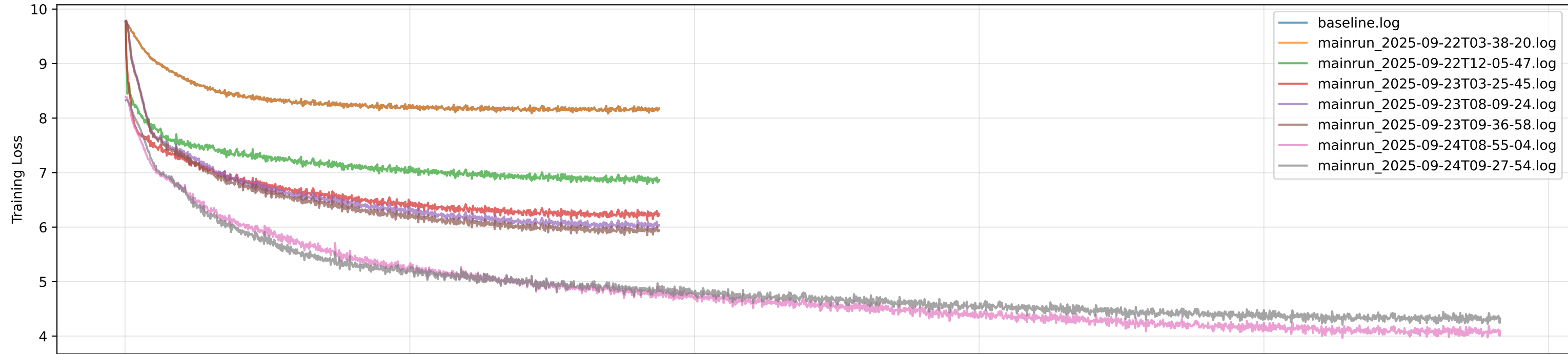


Validation Loss Curves



Training Loss



Model Meta Info Table

	n_layer	n_head	d_model	block_size	batch_size	vocab_size	dropout	lr	weight_decay	min_lr	warmup_ratio	epochs	parameters_count	device	final_val_loss
baseline.log	6	8	512	128	64	16000	0.1	0.006	0.0			7	27172864	cpu	1.7533
mainrun_2025-09-22T03-38-20.log	6	8	512	128	64	16000	0.1	0.006	0.0			7	27172864	cpu	1.7533
mainrun_2025-09-22T12-05-47.log	6	8	512	128	64	16000	0.1	0.006	0.0			7	27172864	mps	1.4893
mainrun_2025-09-23T03-25-45.log	6	8	512	128	64	16000	0.2012	0.0003	0.01			7	27172864	mps	1.3544
mainrun_2025-09-23T08-09-24.log	6	8	512	128	64	16000	0.153	0.0005	0.01			7	27172864	mps	1.327
mainrun_2025-09-23T09-36-58.log	6	8	512	128	64	16000	0.153	0.0005	0.01			7	27172864	mps	1.311
mainrun_2025-09-24T08-55-04.log	6	6	384	64	64	4000	0.051	0.0011	0.01	1e-05	0.1	7	12208128	mps	1.2935
mainrun_2025-09-24T09-27-54.log	4	3	192	64	64	4000	0.0603	0.0021	0.01	1e-05	0.1	7	2560128	mps	1.3206