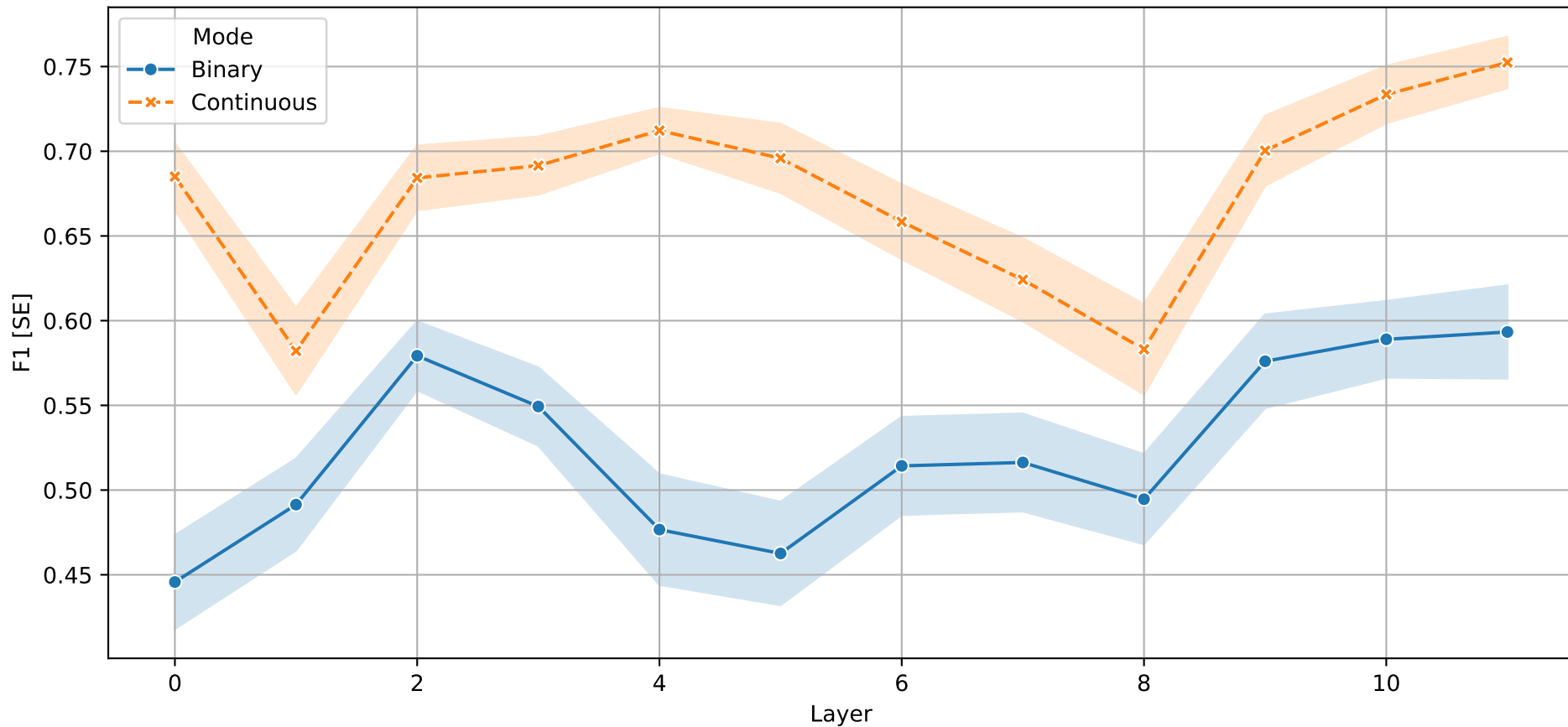
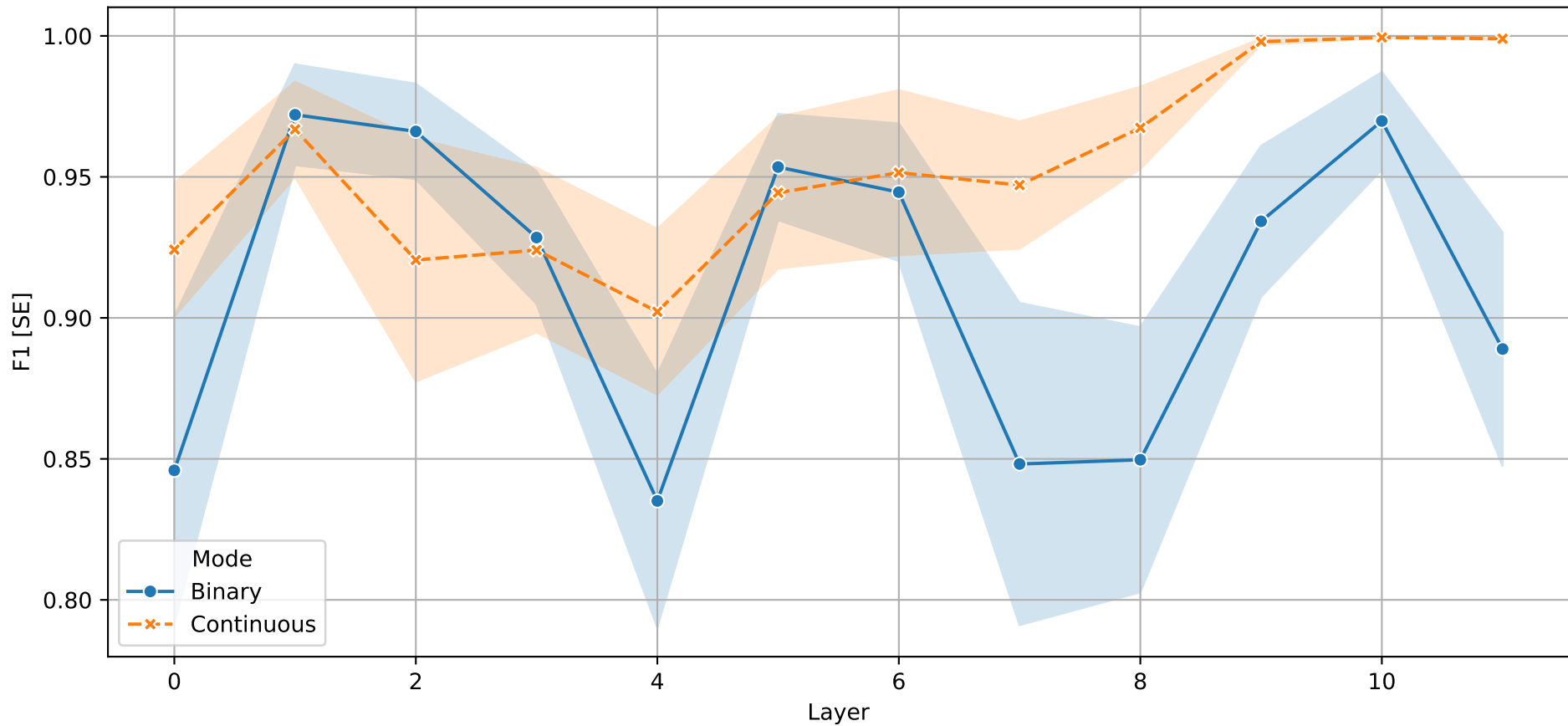


F1 per Layer - Single Neuron Probing

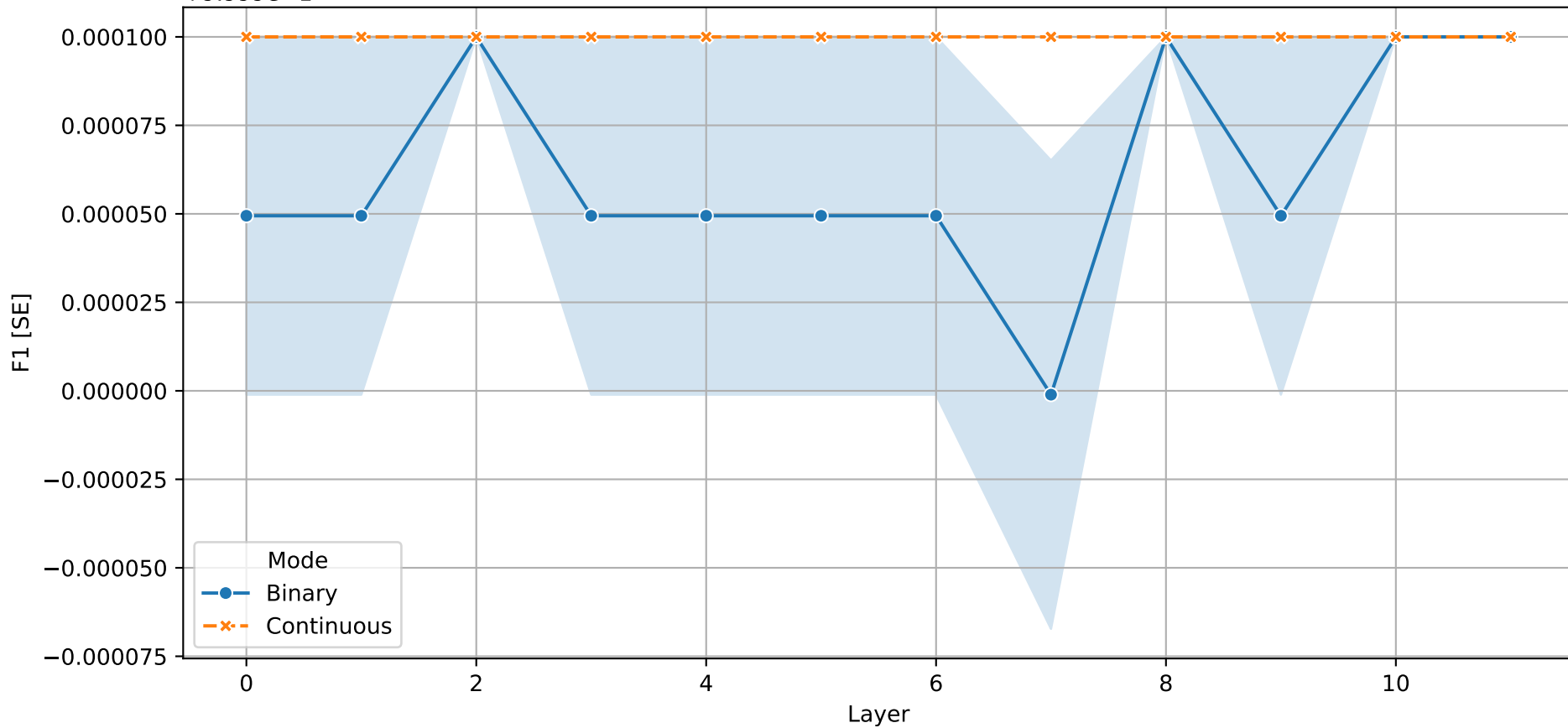


F1 per Layer - Top-K Neurons Probing

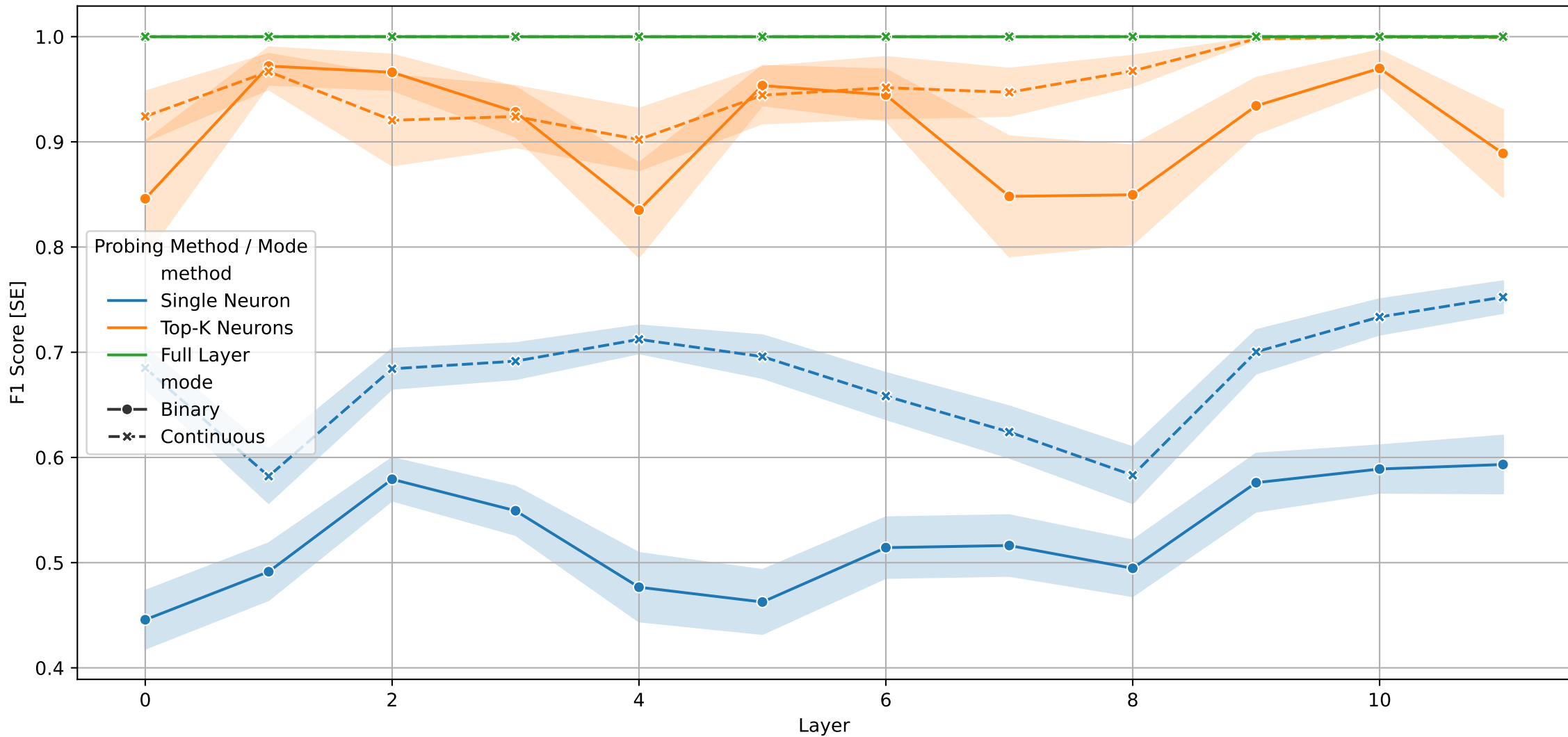


F1 per Layer - Full Layer Probing

+9.999e-1



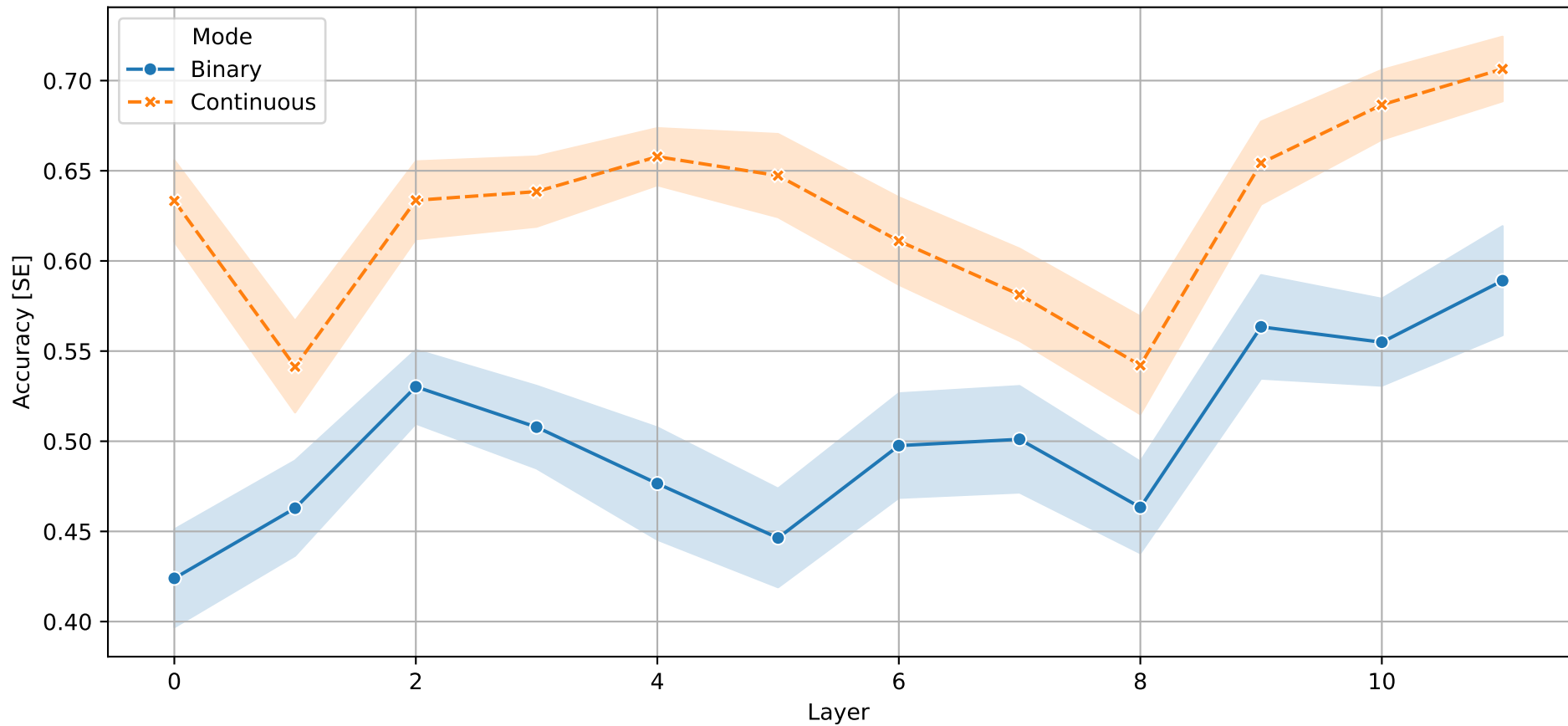
Overall F1 per Layer - All Methods



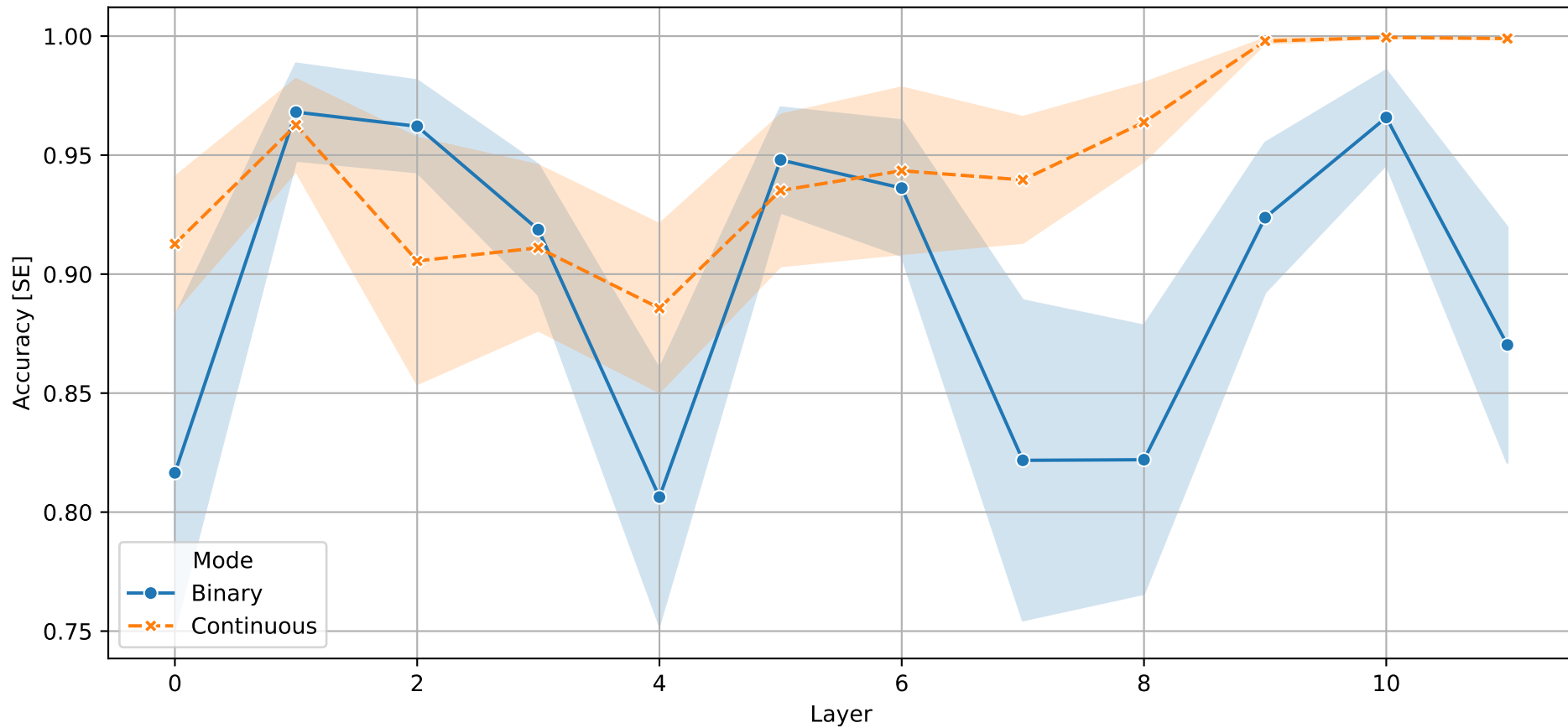
F1 Score Summary by Probing Method

Method	Metric	Binary	Continuous
Full Layer	f1_best_layer	2.0	0.0
Full Layer	f1_max	1.0	1.0
Full Layer	f1_mean	1.0	1.0
Full Layer	f1_std	0.0001	0.0
Single Neuron	f1_best_layer	11.0	11.0
Single Neuron	f1_max	1.0	1.0
Single Neuron	f1_mean	0.524	0.6752
Single Neuron	f1_std	0.246	0.189
Top-K Neurons	f1_best_layer	1.0	10.0
Top-K Neurons	f1_max	1.0	1.0
Top-K Neurons	f1_mean	0.9114	0.9537
Top-K Neurons	f1_std	0.1082	0.0705

Accuracy per Layer - Single Neuron Probing

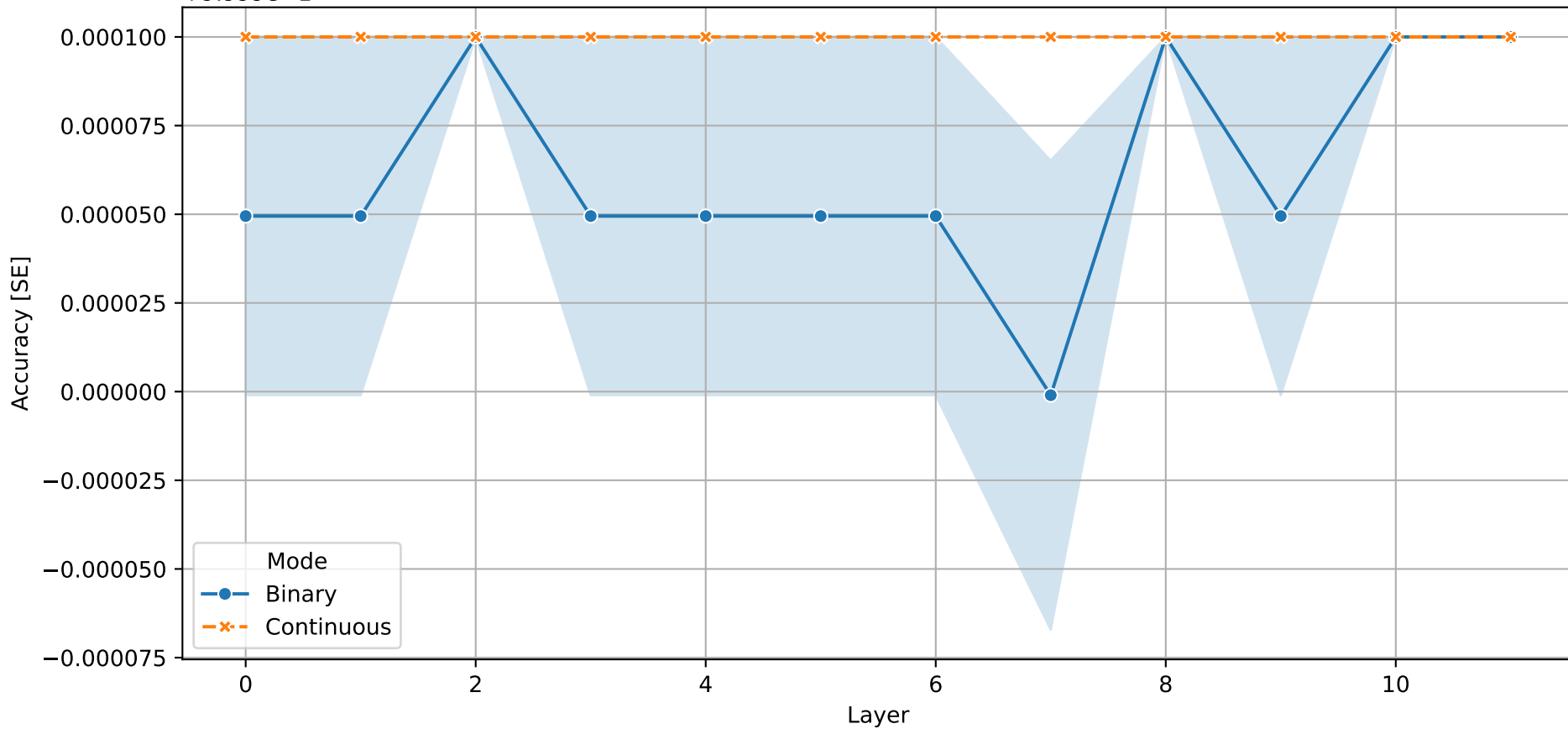


Accuracy per Layer - Top-K Neurons Probing

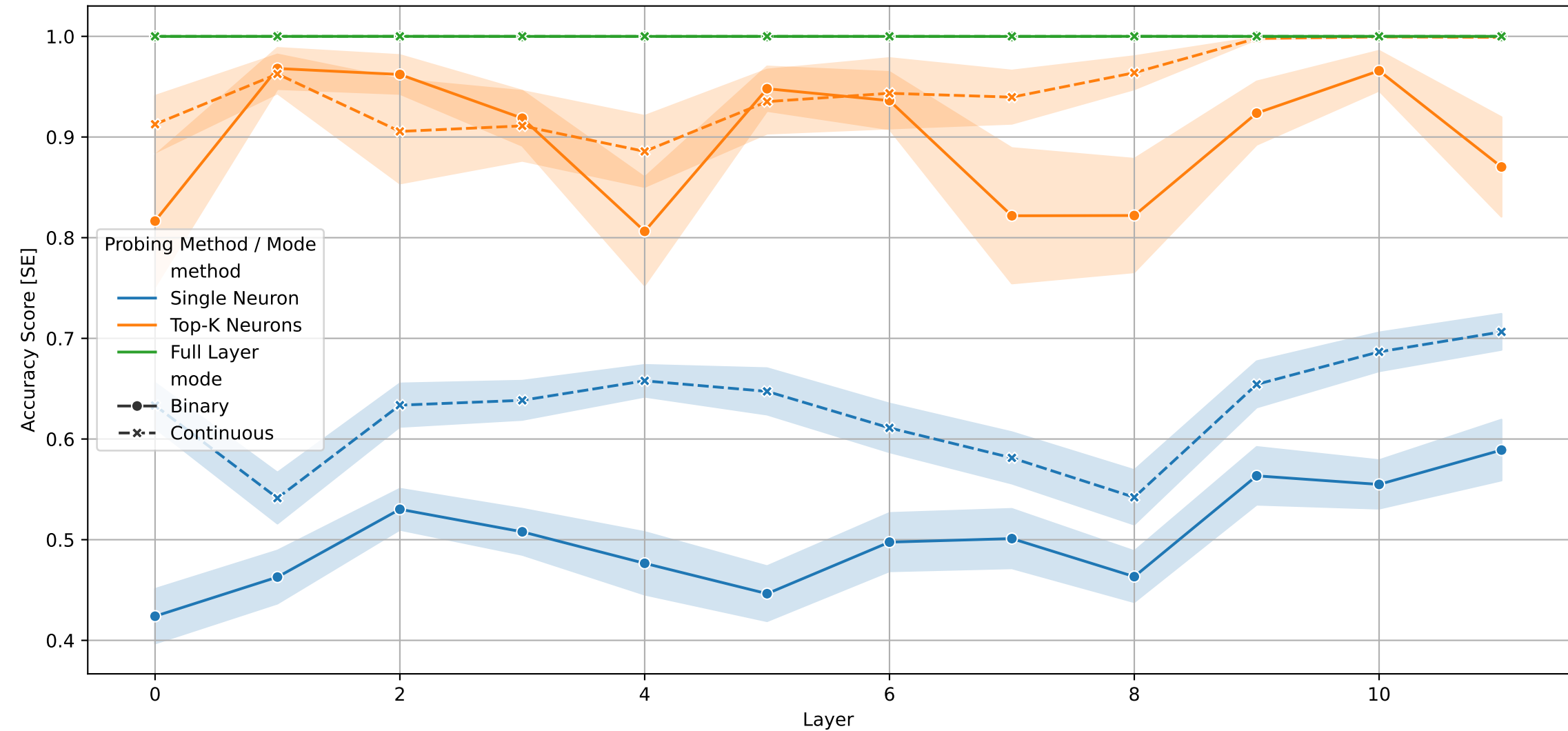


Accuracy per Layer - Full Layer Probing

+9.999e-1



Overall Accuracy per Layer - All Methods



Accuracy Score Summary by Probing Method

Method	Metric	Binary	Continuous
Full Layer	accuracy_best_layer	2.0	0.0
Full Layer	accuracy_max	1.0	1.0
Full Layer	accuracy_mean	1.0	1.0
Full Layer	accuracy_std	0.0001	0.0
Single Neuron	accuracy_best_layer	11.0	11.0
Single Neuron	accuracy_max	1.0	1.0
Single Neuron	accuracy_mean	0.5014	0.6278
Single Neuron	accuracy_std	0.2447	0.2041
Top-K Neurons	accuracy_best_layer	1.0	10.0
Top-K Neurons	accuracy_max	1.0	1.0
Top-K Neurons	accuracy_mean	0.8966	0.9463
Top-K Neurons	accuracy_std	0.1287	0.0837