

A Appendix

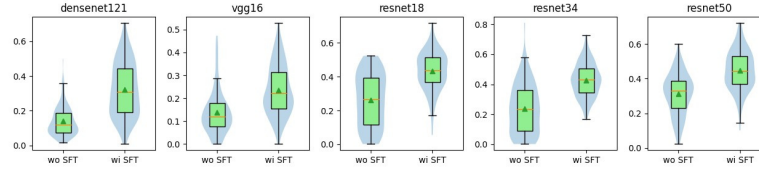


Fig. A1: IoU bewteen the predicted gaze map and the Grad-CAM maps of classification models that use SFT or not.

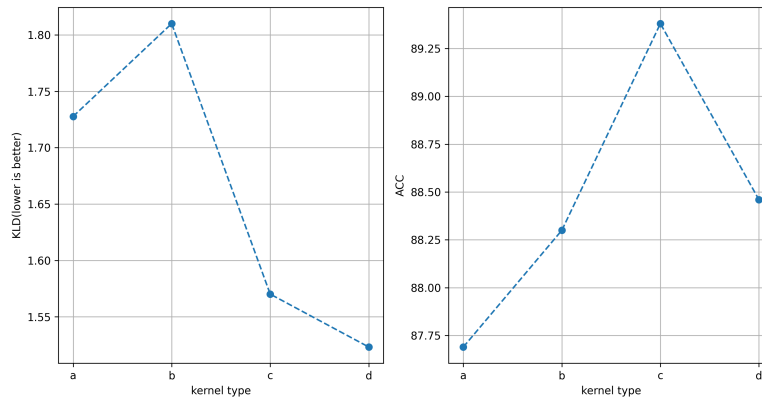


Fig. A2: Comparison of KLD and ACC using four distinct learnable kernels. Four distinct learnable kernels are shown in Fig. A3.

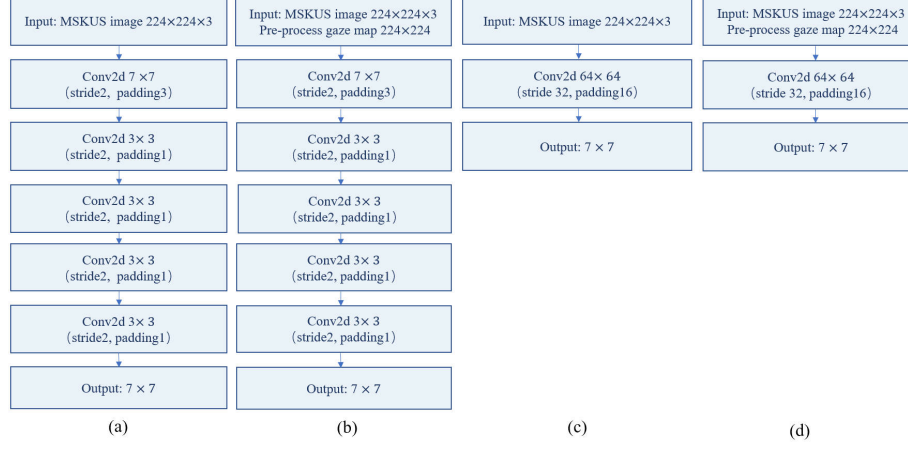


Fig. A3: Four different learnable kernel layer structure.

Table A1: Comparison with other sonographer attention-based adjusting mechanism and fixed Gaussian kernel (Five CNN backbone networks).

Models	ACC↑(%)	AUC↑	CC↑	SIM↑	KLD↓
ResNet18	85.072±2.647	0.918±0.012	0.259±0.039	0.152±0.014	2.141±0.102
ResNet18-TLS	89.131±3.325	0.968±0.005	0.404±0.004	0.281±0.003	1.787±0.017
ResNet18-Gaussian	86.768±1.841	0.957±0.007	0.504±0.014	0.243±0.007	1.622±0.035
ResNet18-SFT	89.535±2.554	0.970±0.005	0.592±0.027	0.286±0.011	1.463±0.041
ResNet34	84.923±2.464	0.925±0.012	0.263±0.029	0.141±0.005	2.139±0.043
ResNet34-TLS	87.563±1.890	0.947±0.018	0.376±0.006	0.252±0.004	1.951±0.043
ResNet34-Gaussian	89.698±4.234	0.964±0.009	0.460±0.023	0.227±0.003	1.697±0.032
ResNet34-SFT	91.237±0.925	0.967±0.013	0.559±0.027	0.266±0.009	1.518±0.030
ResNet50	83.385±5.705	0.924±0.028	0.247±0.025	0.247±0.025	2.168±0.063
ResNet50-TLS	89.617±3.055	0.967±0.011	0.402±0.028	0.298±0.020	2.133±0.232
ResNet50-Gaussian	86.307±3.344	0.937±0.025	0.530±0.048	0.329±0.034	1.517±0.102
ResNet50-SFT	90.615±2.142	0.974±0.008	0.603±0.018	0.367±0.015	1.331±0.046
DenseNet121	82.307±0.973	0.927±0.008	0.260±0.036	0.149±0.008	2.142±0.075
DenseNet121-TLS	89.455±1.621	0.965±0.010	0.369±0.011	0.239±0.007	1.991±0.062
DenseNet121-Gaussian	87.077±1.713	0.959±0.017	0.398±0.0497	0.222±0.014	1.791±0.089
DenseNet121-SFT	89.382±0.574	0.966±0.004	0.529±0.049	0.272±0.021	1.570±0.107
Vgg16	89.132±3.201	0.958±0.021	0.221±0.089	0.182±0.044	3.461±0.776
Vgg16-TLS	91.507±2.885	0.966±0.020	0.416±0.020	0.305±0.013	1.932±0.084
Vgg16-Gaussian	90.615±2.451	0.959±0.034	0.509±0.092	0.330±0.055	1.689±0.500
Vgg16-SFT	91.534±2.525	0.976±0.010	0.560±0.039	0.376±0.040	1.427±0.079