Learning How to MIMIC: Using Model Explanations to Guide Deep Learning Training

Matthew Watson, Bashar Awwad Shiekh Hasan, Noura Al Moubayed
Durham University
Durham, UK
{matthew.s.watson,bashar.awwad-shiekh-hasan,noura.al-moubayed}@durham.ac.uk

Table 1. Table reporting model accuracy and mean KLD/NSS similarity between GradCAM explanations and radiologist eye-gaze data.

Model Architecture	Seed	Accuracy	KLD	NSS
Baseline	1735	72.17	10.744	-0.497
	2948	74.34	6.114	0.174
	4235	72.61	8.902	-0.331
	4582	69.00	9.555	-0.272
	4678	74.00	13.349	-0.145
	5682	73.81	4.288	0.183
	7624	75.55	14.404	-0.858
	7626	73.69	14.064	-0.113
	9374	69.85	10.289	-0.078
	9576	73.09	7.197	-0.173
Improved UNet (current SOTA)	1735	75.29	5.363	0.469
	2948	70.58	13.031	-0.032
	4235	75.57	5.429	0.096
	4582	76.51	9.937	-0.324
	4678	75.77	7.266	-0.169
	5682	69.25	12.195	0.336
	7624	74.93	11.257	0.405
	7626	75.85	4.992	0.025
	9374	74.59	4.672	-0.777
	9576	72.97	9.265	-0.386
Normal Ensemble	1735	74.59	1.221	-0.070
	2948	78.90	2.287	0.359
	4235	78.90	3.285	0.360
	4582	74.11	2.378	0.324
	4678	**79.86**	3.884	-0.165
	5682	78.42	4.944	-0.653
	7624	76.51	2.117	0.269
	7626	75.12	1.2688	0.290
	9374	73.64	1.099	-0.249
	9576	76.99	1.740	1.111
Expl. Ensemble (Ours)	1735	74.11	1.340	0.640
	2948	76.51	1.025	0.577
	4235	76.99	**0.786**	**1.237**
	4582	76.51	0.967	1.157
	4678	75.60	1.808	0.758
	5682	73.16	1.388	0.666
	7624	73.16	1.263	1.170
	7626	77.46	1.267	0.566
	9374	78.94	0.820	1.176
	9576	76.03	0.908	1.011

Seed	Ensemble Size	KLD	NSS	Seed	Ensemble Size	KLD	NSS
1467	5	1.983	-2.515	1467	8	1.485	0.837
3942	5	2.785	-1.013	3942	8	1.701	-1.16
4635	5	1.936	0.279	4635	8	2.175	-1.145
8304	5	2.694	-2.151	8304	8	2.321	-1.163
5305	5	2.292	-2.302	5305	8	1.266	-0.842
5439	5	1.833	-1.489	5439	8	1.471	0.831
6395	5	2.302	-1.853	6395	8	6.55	-1.491
7098	5	1.811	-1.586	7098	8	2.503	0.556
2089	5	2.472	-2.995	2089	8	1.25	-1.045
3104	5	2.441	1.021	3104	8	1.559	-0.954
1467	7	2.193	0.298	1467	10	0.786	1.237
3942	7	1.991	-0.081	3942	10	0.967	1.157
4635	7	2.372	0.23	4635	10	1.808	0.758
8304	7	1.975	-2.031	8304	10	1.388	0.666
5305	7	2.382	0.023	5305	10	1.263	1.170
5439	7	2.476	3.741	5439	10	1.267	0.566
6395	7	1.313	-1.64	6395	10	0.820	1.176
7098	7	2.004	2.298	7098	10	0.908	1.011
2089	7	1.608	-0.069	2089	10	1.340	0.640
3104	7	1.541	1.22	3104	10	1.025	0.577

Table 2. Table reporting mean KLD/NSS similarity between GradCAM explanations and radiologist eye-gaze data of our Explanation Ensemble architecture with differing numbers of sub-models (i.e. ensemble size).

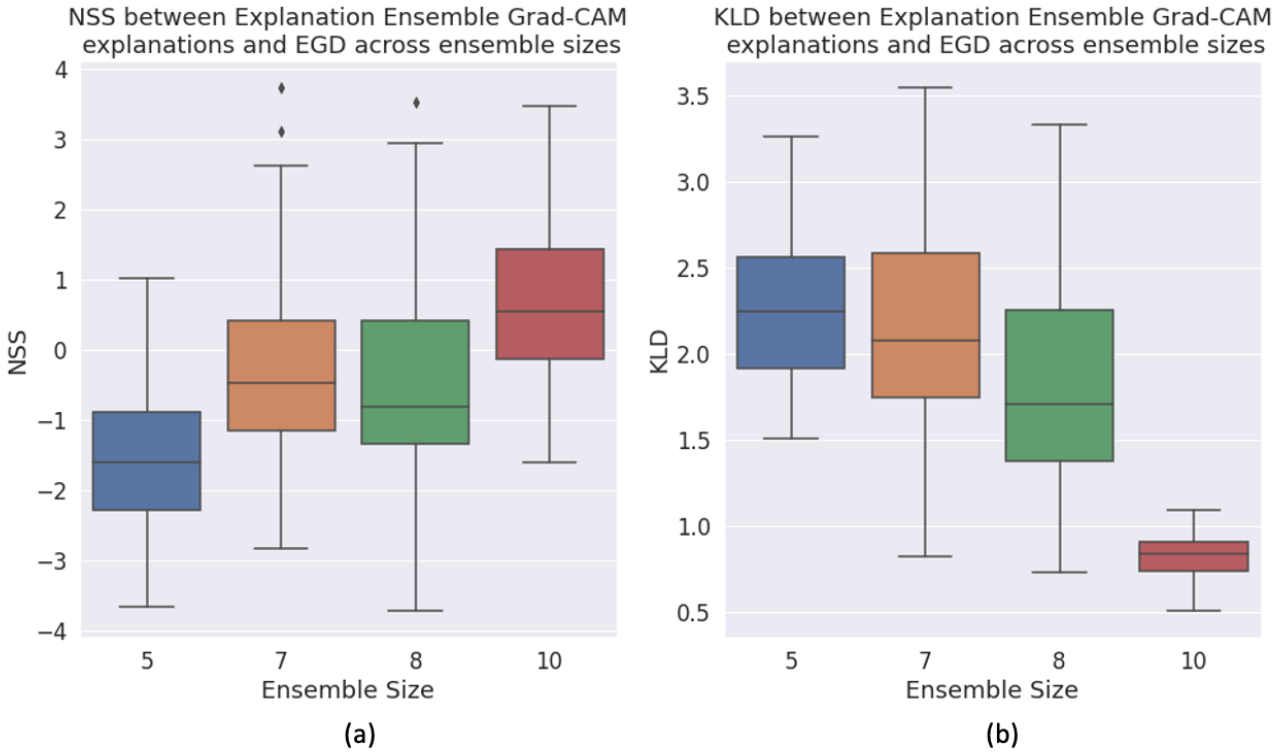

Figure 1. Boxplots of mean (a) NSS and (b) KLD between Grad-CAM explanations and radiologist EGD, across a range of ensemble sizes. For each ensemble size, 10 models with different random seeds were trained. Note that KLD is a divergence metric meaning smaller values are better.

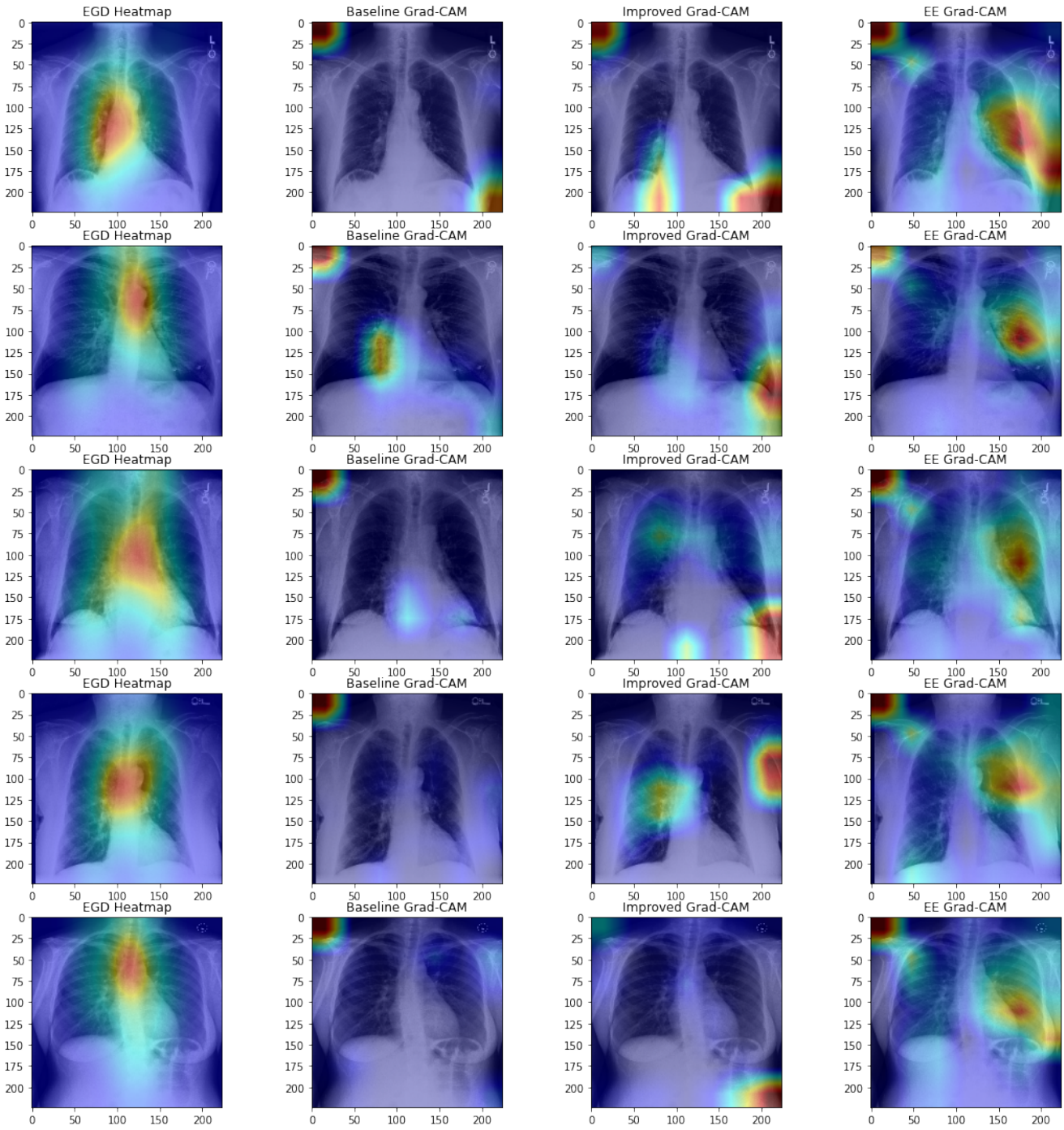

Figure 2. 5 random samples from the MIMIC-CXR-EGD dataset overlaid with the eye gaze data heatmaps and GradCAM explanations from the baseline, improved UNet and explanation ensemble models.

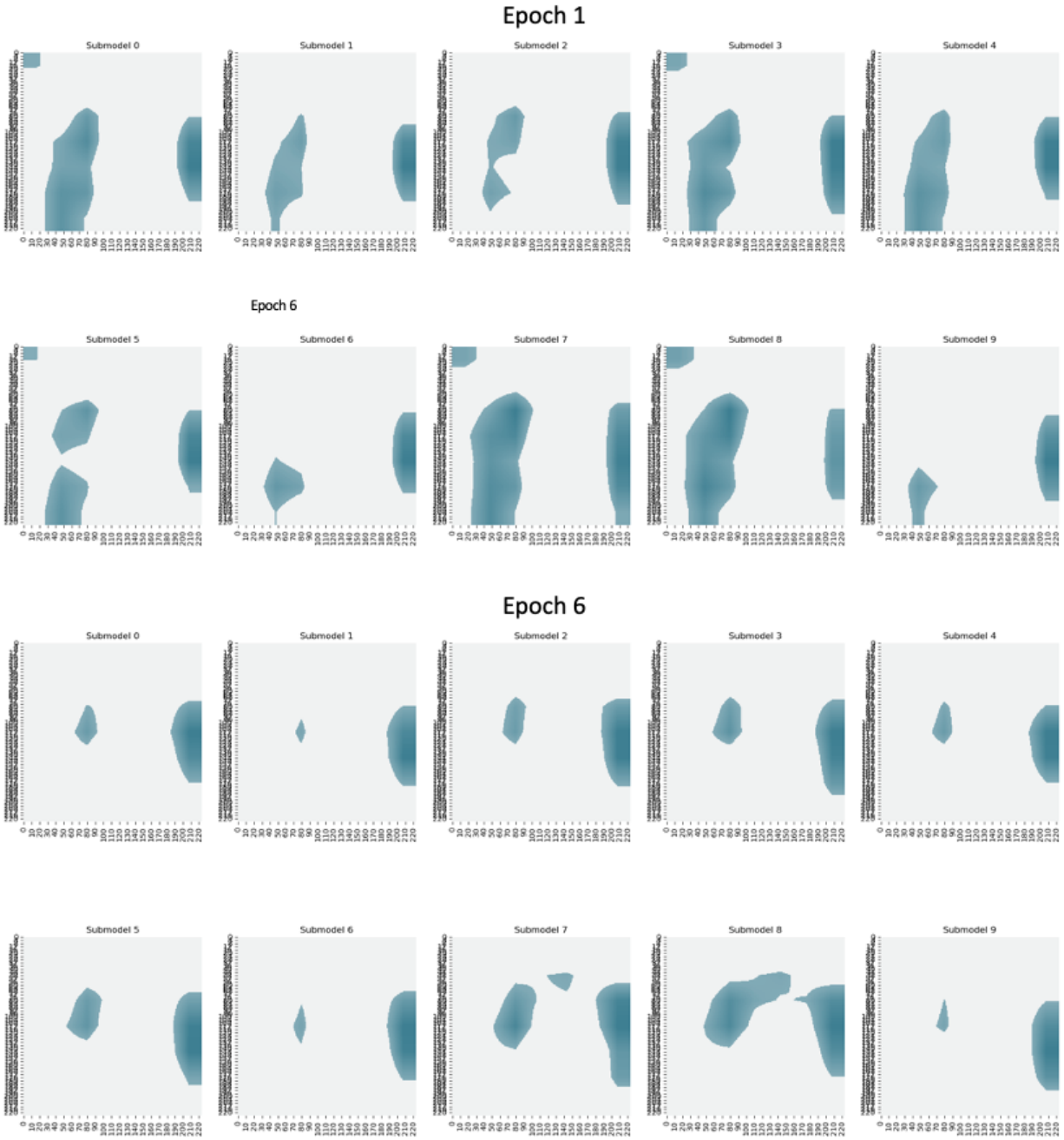

Figure 3. Average GradCAM values (across the validation split) of each sub-model of our Explanation Ensemble model, as training progresses over epochs 1 and 6. To aid with visualisation, only the most important 50% of pixels are shown. Sub-models start training with vastly different learned features, and as training progresses our training procedure encourages the sub-models to learn similar features. Joint with Figure 4 in the Supplementary Material, this is a larger version of Figure 4 in the main paper.

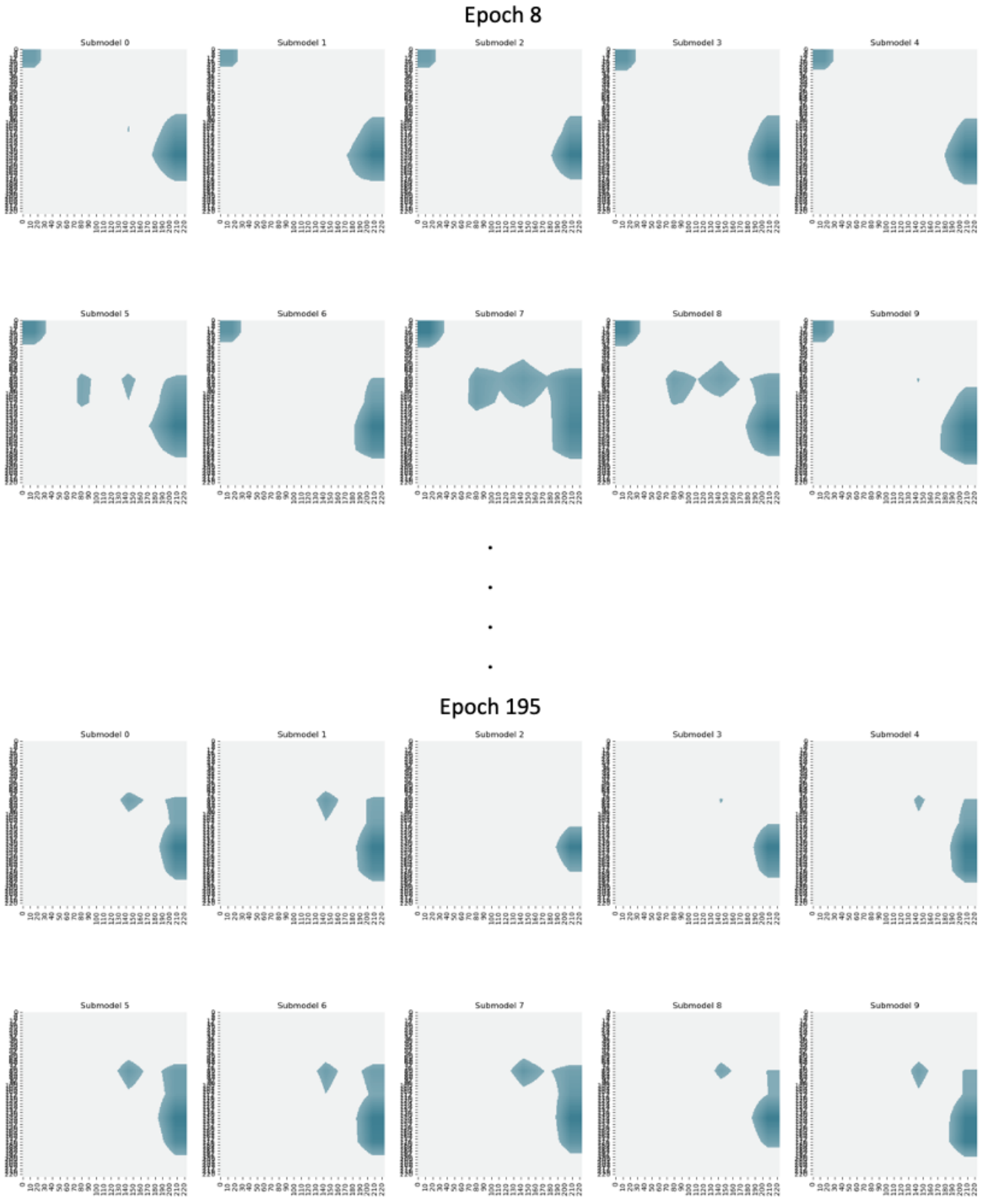

Figure 4. Average GradCAM values (across the validation split) of each sub-model of our Explanation Ensemble model, as training progresses over epochs 1 and 6. To aid with visualisation, only the most important 50% of pixels are shown. Sub-models start training with vastly different learned features, and as training progresses our training procedure encourages the sub-models to learn similar features. Joint with Figure 3 in the Supplementary Material, this is a larger version of Figure 4 in the main paper.