RESNET152

Single Task Learning

Dataset = **CIFAR10**, Epochs = **20,** Model Architecture = **ResNet152**

STL ResNet152 (CIFAR10) Task 1: 10 Classes Predictions								
Model #	**Non-ResNet Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 1024, 512, 256, 128, 10)	128	0	Adam	1.030	0.644	1.002	**0.664**
2	(2048, 1024, 512, 256, 128, 10)	64	0	Adam	1.147	0.594	1.218	0.570
3	(2048, 1024, 512, 256, 128, 10)	64	0.4	Adam	0.685	**0.778**	2.033	0.544
4	(2048, 1024, 512, 256, 128, 10)	128	0	SGD	0.778	0.728	2.671	0.301
5	(2048, 1024, 512, 256, 128, 10)	64	0	SGD	0.702	0.758	7.316	0.191
6	(2048, 1024, 512, 256, 128, 10)	64	0.4	SGD	0.949	0.666	3.825	0.215
Time per epoch: ~46 seconds								

Single Task Learning

Dataset = **CIFAR10**, Epochs = **20,** Model Architecture = **ResNet152**

STL ResNet152 (CIFAR10) Task 2: 2 Super Classes Predictions								
Model #	**Non-ResNet Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 512, 256, 128, 2)	128	0	Adam	0.068	**0.973**	0.311	0.928
2	(2048, 512, 256, 128, 2)	64	0	Adam	0.116	0.955	0.177	0.932
3	(2048, 512, 256, 128, 2)	64	0.4	Adam	0.157	0.942	0.162	**0.942**
4	(2048, 512, 256, 128, 2)	128	0	SGD	0.230	0.908	0.268	0.891
5	(2048, 512, 256, 128, 2)	64	0	SGD	0.231	0.905	0.243	0.907
6	(2048, 512, 256, 128, 2)	64	0.4	SGD	0.095	0.964	0.837	0.685
Time per epoch: ~46 seconds								

Single Task Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ResNet152**

STL ResNet152 (CIFAR100) **Task 1: 100 Classes Predictions**								
Model #	**Non-ResNet Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 1024, 512, 256, 128, 100)	128	0	Adam	1.912	**0.461**	2.995	**0.304**
2	(2048, 1024, 512, 256, 128, 100)	64	0	Adam	2.404	0.353	3.443	0.228
3	(2048, 1024, 512, 256, 128, 100)	64	0.4	Adam	2.544	0.328	695.16	0.009
4	(2048, 1024, 512, 256, 128, 100)	128	0	SGD	2.856	0.269	7.683	0.035
5	(2048, 1024, 512, 256, 128, 100)	64	0	SGD	2.940	0.261	4.522	0.080
6	(2048, 1024, 512, 256, 128, 100)	64	0.4	SGD	2.674	0.301	4.102	0.141
Time per epoch: ~ 46 seconds								

Single Task Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ResNet152**

STL ResNet152 (CIFAR100) **Task 2: 20 Super Classes Predictions**								
Model #	**Dense Layers (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 512, 256, 128, 20)	128	0	Adam	1.350	**0.571**	1.981	**0.430**
2	(2048, 512, 256, 128, 20)	64	0	Adam	1.464	0.541	2.343	0.396
3	(2048, 512, 256, 128, 20)	64	0.4	Adam	1.603	0.496	2.201	0.369
4	(2048, 512, 256, 128, 20)	128	0	SGD	1.896	0.404	4.362	0.100
5	(2048, 512, 256, 128, 20)	64	0	SGD	1.901	0.402	2.534	0.264
6	(2048, 512, 256, 128, 20)	64	0.4	SGD	1.468	0.538	3.281	0.207
Time per epoch: ~ 52 seconds								

Multi Tasks Learning

Dataset = **CIFAR10**, Epochs = **20,** Model Architecture = **ResNet152**

MTL ResNet152 (CIFAR10) Task 1: 10 Classes Predictions Task 2: 2 Super Classes Predictions									
Model #	**Dense Layers (Nodes in Each Layer) - 2 Branches**	**Batch Size**	**Gamma**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss (Task 1, Task 2)**	**Train Acc (Task 1, Task 2)**	**Test Loss (Task 1, Task 2)**	**Test Acc (Task 1, Task 2)**
1	(2048, 1024, 512, 256, 128, 10) (2048, 512, 256, 128, 2)	128	0.5	0	Adam	0.633, 0.078	0.784, 0.972	1.132, 0.200	0.653, 0.938
2	(2048, 1024, 512, 256, 128, 10) (2048, 512, 256, 128, 2)	128	0.4	0	Adam	0.587, 0.070	**0.802, 0.974**	1.046, 0.201	**0.688, 0.949**
3	(2048, 1024, 512, 256, 128, 10) (2048, 512, 256, 128, 2)	128	0.6	0	Adam	0.798, 0.090	0.721, 0.966	1.083, 0.156	0.637, 0.946
4	(2048, 1024, 512, 256, 128, 10) (2048, 512, 256, 128, 2)	128	0.4	0.4	Adam	1.192, 0.154	0.561, 0.941	1.292, 0.165	0.539, 0.937
5	(2048, 1024, 512, 256, 128, 10) (2048, 512, 256, 128, 2)	128	0.4	0.4	SGD	1.246, 0.154	0.535, 0.940	1.526, 0.307	0.450, 0.874
Time per epoch: ~ 46 seconds									

Multi Tasks Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ResNet152**

MTL ResNet152 (CIFAR100) Task 1: 100 Classes Predictions Task 2: 20 Super Classes Predictions									
Model #	**Dense Layers (Nodes in Each Layer) - 2 Branches**	**Batch Size**	**Gamma**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss (Task 1, Task 2)**	**Train Acc (Task 1, Task 2)**	**Test Loss (Task 1, Task 2)**	**Test Acc (Task 1, Task 2)**
1	(2048, 1024, 512, 256, 128, 100) (2048, 512, 256, 128, 20)	128	0.5	0	Adam	2.325, 1.271	**0.355, 0.588**	3.045, 1.799	**0.271, 0.476**
2	(2048, 1024, 512, 256, 128, 100) (2048, 512, 256, 128, 20)	128	0.4	0	Adam	2.383, 1.330	0.349, 0.574	3.031, 1.804	0.244, 0.450
3	(2048, 1024, 512, 256, 128, 100) (2048, 512, 256, 128, 20)	128	0.6	0	Adam	2.582, 1.439	0.309, 0.546	2.997, 1.783	0.251, 0.466
4	(2048, 1024, 512, 256, 128, 100) (2048, 512, 256, 128, 20)	128	0.4	0.4	Adam	3.237, 1.973	0.185, 0.381	3.425, 2.116	0.165, 0.349
5	(2048, 1024, 512, 256, 128, 100) (2048, 512, 256, 128, 20)	128	0.4	0.4	SGD	3.042, 1.719	0.201, 0.453	4.232, 2.630	0.093, 0.250
Time per epoch: ~ 47 seconds									

Vision Transformer

Single Task Learning

Dataset = **CIFAR10**, Epochs = **20,** Model Architecture = **ViT**

STL ViT (CIFAR10) Task 1: 10 Classes Predictions								
Model #	**Non-ViT Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 1024, 512, 256, 128, 8)	256	0	Adam	0.027	**0.991**	2.153	0.641
2	(2048, 1024, 512, 256, 128, 8)	128	0	Adam	0.042	0.987	2.076	0.657
3	(2048, 1024, 512, 256, 128, 8)	256	0.5	Adam	0.950	0.684	1.015	**0.659**
4	(2048, 1024, 512, 256, 128, 8)	256	0	SGD	0.567	0.805	2.714	0.466
5	(2048, 1024, 512, 256, 128, 8)	128	0	SGD	0.059	0.986	2.747	0.535
6	(2048, 1024, 512, 256, 128, 8)	256	0.5	SGD	1.901	0.262	2.026	0.268
Time per epoch: ~ 63 seconds								

Single Task Learning

Dataset = **CIFAR10**, Epochs = **10,** Model Architecture = **ViT**

STL ViT (CIFAR10) Task 2: 2 Classes Predictions								
Model #	**Non-ViT Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 512, 256, 128, 2)	256	0	Adam	0.074	0.971	0.675	0.811
2	(2048, 512, 256, 128, 2)	128	0	Adam	0.031	**0.989**	0.449	0.912
3	(2048, 512, 256, 128, 2)	256	0.5	Adam	0.167	0.938	0.179	**0.933**
4	(2048, 512, 256, 128, 2)	256	0	SGD	0.214	0.914	0.258	0.896
5	(2048, 512, 256, 128, 2)	128	0	SGD	0.157	0.938	0.218	0.912
6	(2048, 512, 256, 128, 2)	256	0.5	SGD	0.408	0.824	0.375	0.838
Time per epoch: ~ 63 seconds								

Single Task Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ViT**

STL ViT (CIFAR100) **Task 1: 100 Classes Predictions**								
Model #	**Non-ViT Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 1024, 512, 256, 128, 8)	256	0	Adam	0.079	**0.977**	6.613	0.300
2	(2048, 1024, 512, 256, 128, 8)	128	0	Adam	0.129	0.964	5.448	**0.302**
3	(2048, 1024, 512, 256, 128, 8)	256	0.5	Adam	4.148	0.046	4.043	0.058
4	(2048, 1024, 512, 256, 128, 8)	256	0	SGD	2.180	0.451	8.547	0.035
5	(2048, 1024, 512, 256, 128, 8)	128	0	SGD	0.582	0.862	4.652	0.232
6	(2048, 1024, 512, 256, 128, 8)	256	0.5	SGD	4.606	0.010	4.604	0.012
Time per epoch: ~ 63 seconds								

Single Task Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ViT**

STL ViT (CIFAR100) **Task 2: 20 Classes Predictions**								
Model #	**Non-ViT Layer (Nodes in Each Layer)**	**Batch Size**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss**	**Train Acc**	**Test Loss**	**Test Acc**
1	(2048, 512, 256, 128, 2)	256	0	Adam	0.036	**0.989**	4.432	**0.458**
2	(2048, 512, 256, 128, 2)	128	0	Adam	0.061	0.982	3.970	0.448
3	(2048, 512, 256, 128, 2)	256	0.5	Adam	2.047	0.371	2.024	0.382
4	(2048, 512, 256, 128, 2)	256	0	SGD	1.202	0.634	3.488	0.300
5	(2048, 512, 256, 128, 2)	128	0	SGD	0.233	0.943	2.841	0.396
6	(2048, 512, 256, 128, 2)	256	0.5	SGD	2.729	0.149	2.632	0.191
Time per epoch: ~ 63 seconds								

Multi Tasks Learning

Dataset = **CIFAR10**, Epochs = **20,** Model Architecture = **ViT**

MTL ViT (CIFAR10) Task 1: 10 Classes Predictions Task 2: 2 Super Classes Predictions									
Model #	**Dense Layers (Nodes in Each Layer) - 2 Branches**	**Batch Size**	**Gamma**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss (Task 1, Task 2)**	**Train Acc (Task 1, Task 2)**	**Test Loss (Task 1, Task 2)**	**Test Acc (Task 1, Task 2)**
1	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.5	0.2	Adam	0.091, 0.010	**0.972, 0.997**	1.860, 0.334	0.672, 0.940
2	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.4	0.2	Adam	0.102, 0.013	0.968, 0.996	1.682, 0.326	0.670, 0.943
3	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.6	0.2	Adam	0.085, 0.014	0.974, 0.995	1.696, 0.305	0.667, 0.939
4	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	128	0.5	0.2	Adam	0.123, 0.016	0.964, 0.995	1.610, 0.284	**0.675, 0.937**
5	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	128	0.5	0.2	SGD	1.251, 0.202	0.547, 0.920	1.251, 0.205	0.553, 0.921
Time per epoch: ~ 63 seconds									

Multi Tasks Learning

Dataset = **CIFAR100**, Epochs = **20,** Model Architecture = **ViT**

MTL ViT (CIFAR100) Task 1: 100 Classes Predictions Task 2: 20 Super Classes Predictions									
Model #	**Dense Layers (Nodes in Each Layer) - 2 Branches**	**Batch Size**	**Gamma**	**Dropout (Dense Layer)**	**Optimizer**	**Train Loss (Task 1, Task 2)**	**Train Acc (Task 1, Task 2)**	**Test Loss (Task 1, Task 2)**	**Test Acc (Task 1, Task 2)**
1	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.5	0.2	Adam	0.732, 0.038	**0.788, 0.861**	3.692, 0.155	**0.335, 0.491**
2	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.4	0.2	Adam	0.737, 0.036	0.790, 0.870	3.787, 0.162	0.330, 0.479
3	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	256	0.6	0.2	Adam	0.729, 0.038	0.787, 0.858	3.667, 0.156	0.334, 0.487
4	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	128	0.5	0.2	Adam	0.944, 0.043	0.731, 0.838	3.469, 0.150	0.333, 0.488
5	(2048, 1024, 512, 256, 128, 8) (2048, 512, 256, 128, 2)	128	0.5	0.2	SGD	3.679, 0.194	0.135, 0.160	3.483, 0.175	0.181, 0.231
Time per epoch: ~ 63 seconds									