

# Rebuttal

Table 1: Ablation Study on Fashion MNIST: (A) dimensionality of latent embeddings ordered by depth, (B) dimensionality of MLP and Router hidden layer ordered by depth, (C) Increasing dimensionality, (D) number of hidden layers of MLP, (E) number of hidden layers of router, (F) pre-determined maximum depth of tree.

	$z_{\text{depth}}$	$MLP_{\text{depth}}$	$MLP_{\text{layers}}$	$Router_{\text{layers}}$	$depth$	Acc	NMI	ELBO
base	8, 8, 8, 8, 8, 8	128, 128, 128, 128, 128	1	2	6	$63.5 \pm 4.1$	$63.9 \pm 2.2$	$-239.2 \pm 0.3$
(A)	2, 2, 2, 2, 2, 2					$49.8 \pm 4.0$	$51.8 \pm 2.7$	$-249.5 \pm 0.4$
(A)	4, 4, 4, 4, 4, 4					$60.0 \pm 4.7$	$61.0 \pm 3.7$	$-242.6 \pm 0.4$
(A)	16, 16, 16, 16, 16, 16					$61.0 \pm 4.5$	$63.0 \pm 2.4$	$-238.8 \pm 0.4$
(A)	32, 32, 32, 32, 32, 32					$62.7 \pm 4.4$	$63.5 \pm 2.8$	$-238.7 \pm 0.4$
(B)		16, 16, 16, 16, 16				$58.3 \pm 3.5$	$60.6 \pm 2.1$	$-241.5 \pm 0.3$
(B)		32, 32, 32, 32, 32				$60.6 \pm 3.6$	$62.7 \pm 2.5$	$-240.4 \pm 0.4$
(B)		64, 64, 64, 64, 64				$62.6 \pm 3.8$	$63.5 \pm 1.4$	$-239.6 \pm 0.3$
(B)		256, 256, 256, 256, 256				$62.3 \pm 4.9$	$63.4 \pm 3.9$	$-239.2 \pm 0.4$
(C)	64, 32, 16, 8, 4, 2	128, 64, 32, 16, 8				$56.7 \pm 4.0$	$59.9 \pm 3.1$	$-239.2 \pm 0.4$
(C)	64, 32, 16, 8, 4, 2	256, 128, 64, 32, 16				$60.9 \pm 5.3$	$62.6 \pm 2.4$	$-238.6 \pm 0.3$
(C)	128, 64, 32, 16, 8, 4	128, 64, 32, 16, 8				$58.5 \pm 4.4$	$61.3 \pm 2.5$	$-239.1 \pm 0.4$
(C)	128, 64, 32, 16, 8, 4	256, 128, 64, 32, 16				$62.6 \pm 3.4$	$63.6 \pm 1.5$	$-238.8 \pm 0.3$
(C)	256, 128, 64, 32, 16, 8	256, 128, 64, 32, 16				$61.1 \pm 3.7$	$61.7 \pm 2.9$	$-238.9 \pm 0.3$
(C)	256, 128, 64, 32, 16, 8	512, 256, 128, 64, 32				$60.5 \pm 4.7$	$61.1 \pm 3.1$	$-239.1 \pm 0.3$
(D)			2			$60.3 \pm 5.3$	$61.6 \pm 3.5$	$-239.0 \pm 0.3$
(D)			3			$61.4 \pm 4.2$	$62.9 \pm 3.2$	$-239.0 \pm 0.3$
(E)				1		$59.6 \pm 4.0$	$62.3 \pm 2.0$	$-239.2 \pm 0.4$
(E)				3		$62.1 \pm 4.4$	$63.6 \pm 3.1$	$-239.1 \pm 0.4$
(F)	8, 8, 8, 8, 8	128, 128, 128, 128			5	$64.0 \pm 4.6$	$64.4 \pm 2.7$	$-239.2 \pm 0.3$
(F)	8, 8, 8, 8, 8, 8, 8	128, 128, 128, 128, 128, 128			7	$61.0 \pm 4.9$	$62.9 \pm 2.9$	$-239.2 \pm 0.3$
(F)	8, 8, 8, 8, 8, 8, 8, 8	128, 128, 128, 128, 128, 128, 128, 128			8	$62.1 \pm 5.2$	$63.6 \pm 2.7$	$-239.1 \pm 0.4$

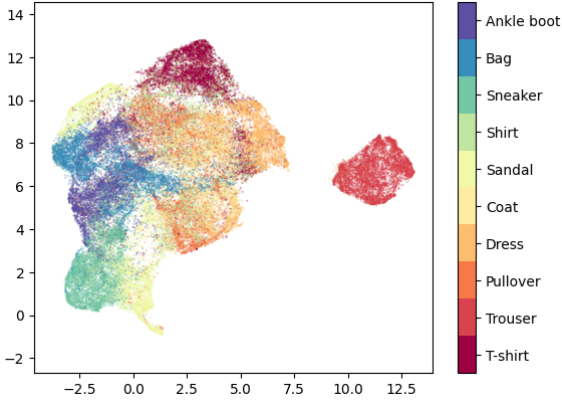


Figure 1: UMAP visualization of root embedding for Fashion MNIST.



Figure 2: Training ELBO after every split for Fashion MNIST.