iScience, Volume *24*

Supplemental information

Miscell: An efficient self-supervised learning approach for dissecting single-cell transcriptome

Hongru Shen, Yang Li, Mengyao Feng, Xilin Shen, Dan Wu, Chao Zhang, Yichen Yang, Meng Yang, Jiani Hu, Jilei Liu, Wei Wang, Qiang Zhang, Fangfang Song, Jilong Yang, Kexin Chen, and Xiangchun Li

SUPPLEMENTARY MATERIALS

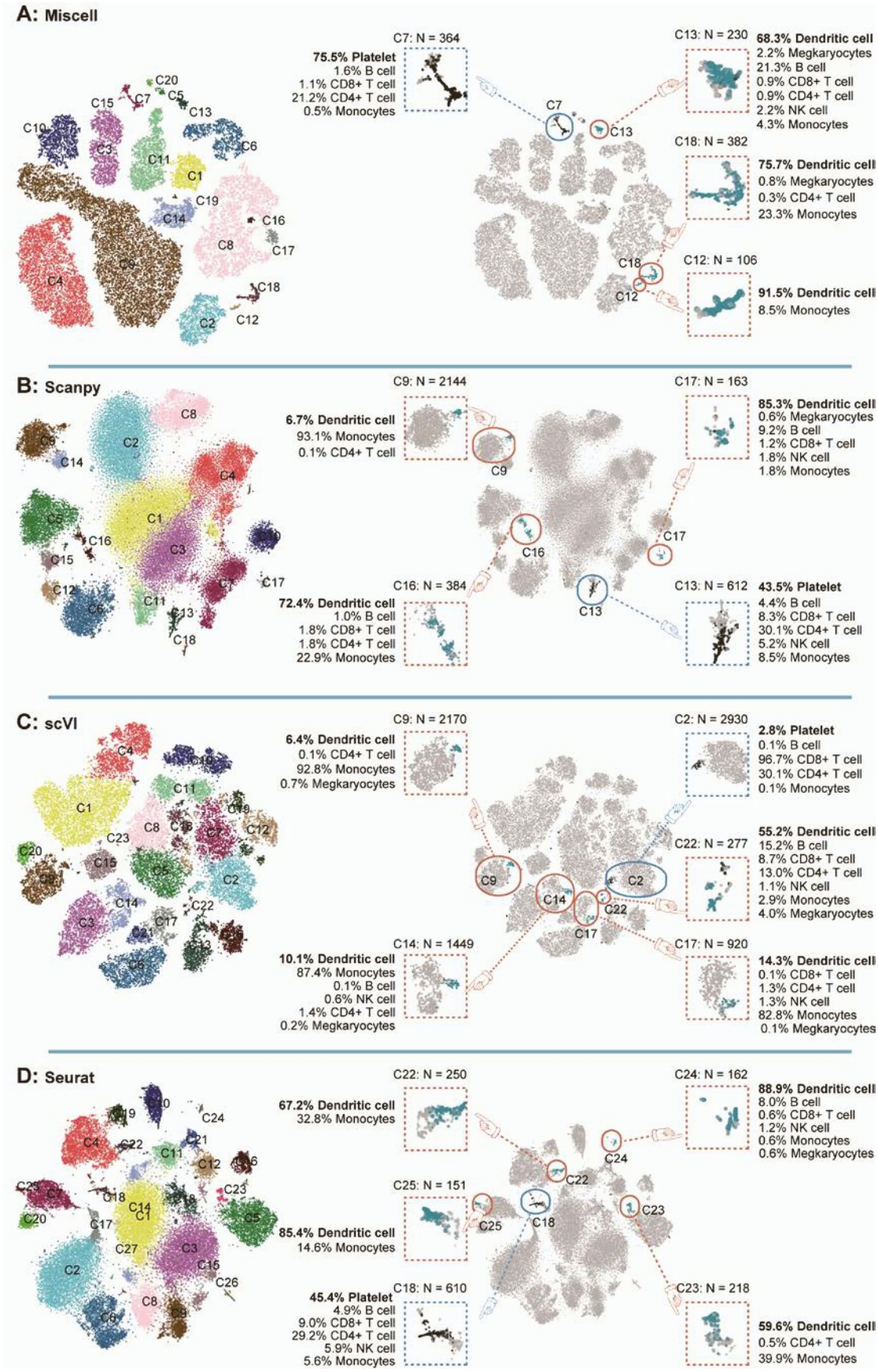


Figure S1. Identification of cell types with small numbers such as platelet and dendritic cell by (A) Miscell, (B) Scanpy, (C) scVI and (D) Seurat, related to Figure 2. All cell clusters (left-side) detected by different methods and cell clusters with small numbers are highlighted (right-side).

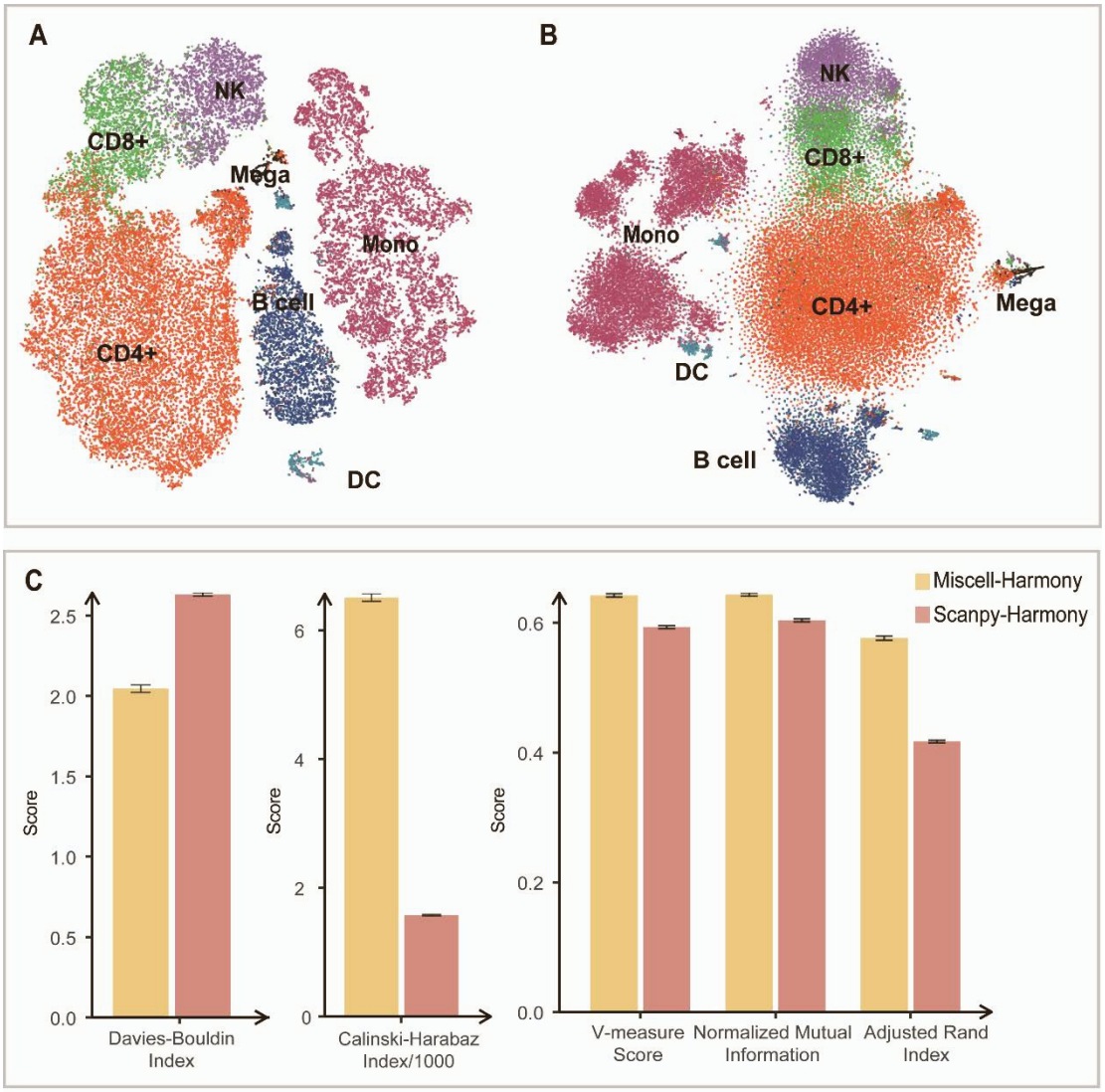


Figure S2. The clustering performances of Miscell versus Scanpy on batch-corrected *PBMC* dataset, related to Figure 2. (**A**) The t-SNE visualization for Miscell; (**B**) The t-SNE visualization for Scanpy; (**C**) The clustering metrics of Miscell versus Scanpy.

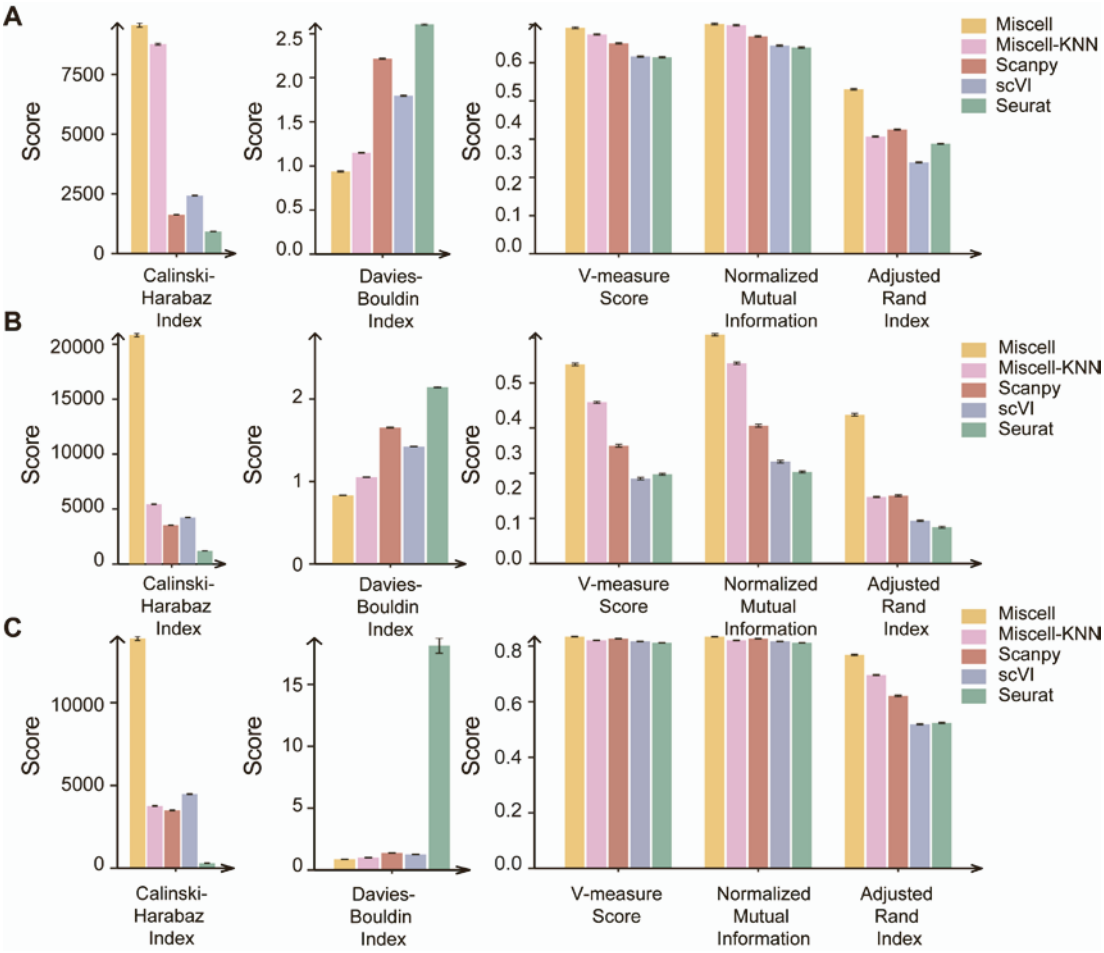


Figure S3. Clustering performance of Miscell, Scanpy-KNN, Scanpy, scVI and Seurat in (A) the *PBMC*, (B) *Smart-seq* and (C) *Muris* dataset, related to Figure 2. Scanpy-KNN indicates replacement of principal component analysis results with latent features of Miscell.

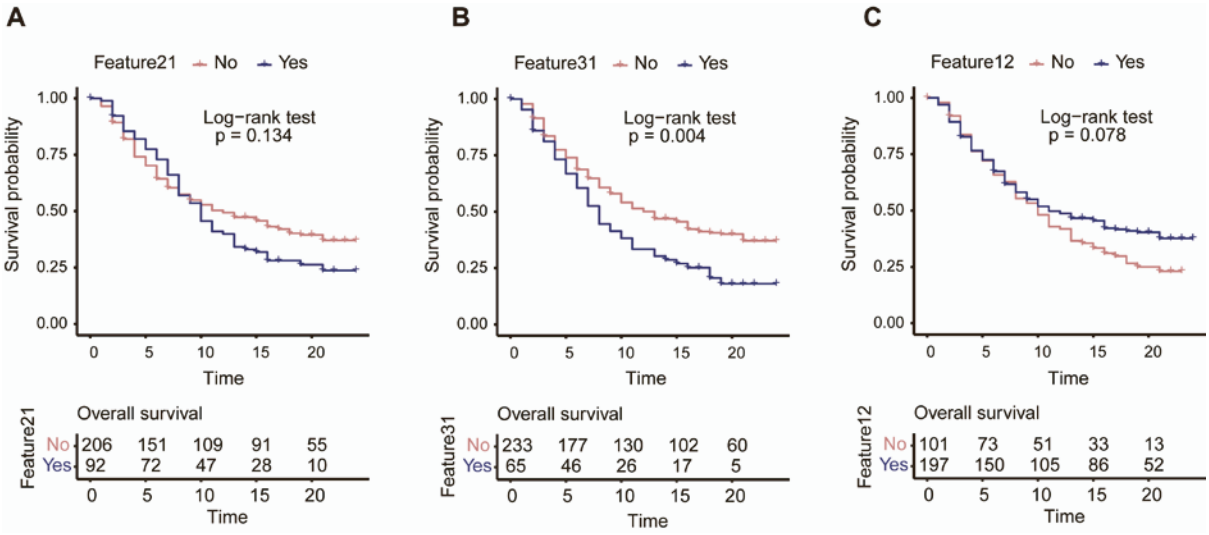


Figure S4. The association between ICB therapy response related latent features extracted by Miscell and overall survival in urothelial carcinoma clinical trials, related to Figure 4.

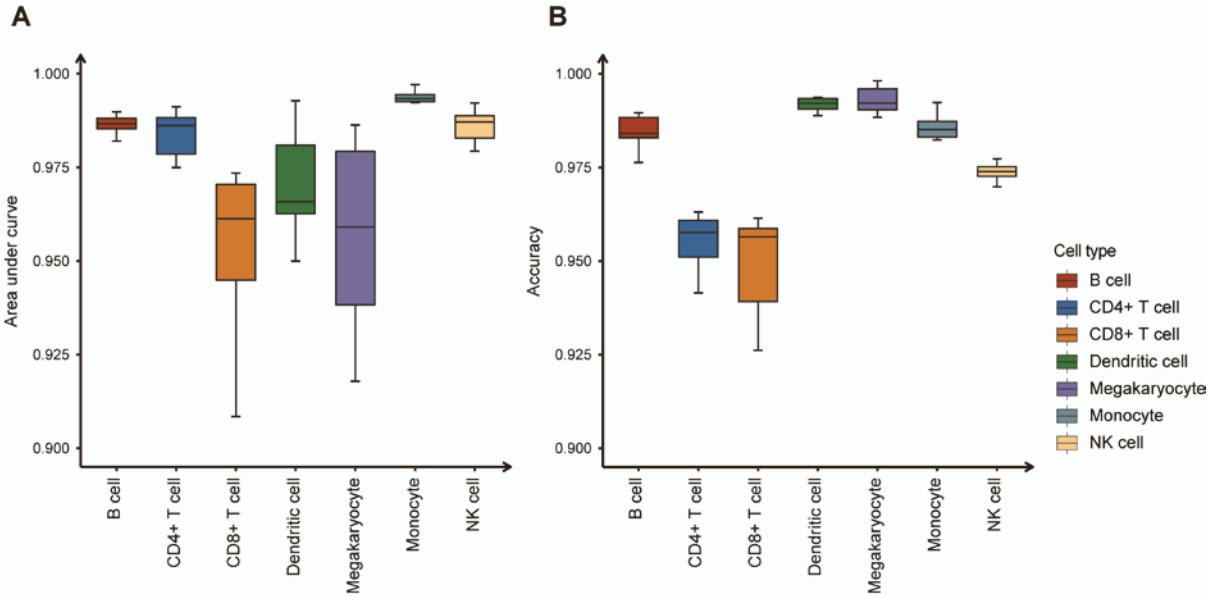


Figure S5. The performance of cell type identification in Miscell, related to Figure 1. (A) Area under the receiver operating curve and **(B)** accuracy from 10-fold cross-validation were shown.

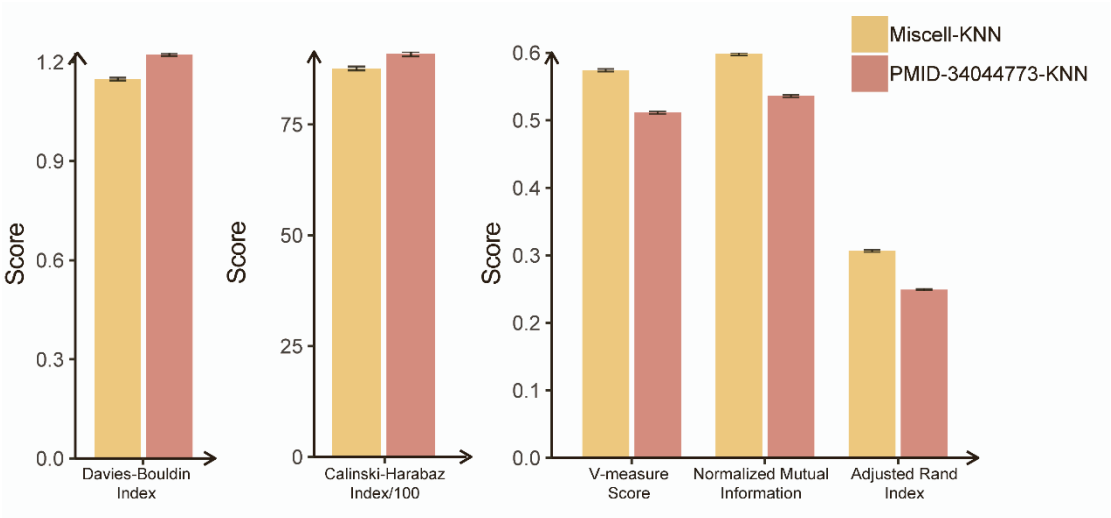


Figure S6. The clustering performances of Miscell-KNN versus the tool proposed in PMID-34044773 on *PBMC* dataset with the same hyper-parameters in KNN clustering, related to Figure 2.

Table S1. Data source information of collected single cell, related to STAR Methods.

GEO accession	Samples	Tumor type	Citation	Accession URL
GSE103322	5902	HNSC	Puram SV, Tirosh I, Parikh AS, Patel AP et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell 2017 Dec 14;171(7):1611-1624.e24.	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE103322&format=file&file=GSE103322%5FHNSCC%5Fall%5Fdata%2Etxt%2Egz
GSE89567	6341	Astrocytoma	Venteicher AS, Tirosh I, Hebert C, Yizhak K et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science 2017 Mar 31;355(6332).	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE89567&format=file&file=GSE89567%5FIDH%5FA%5Fprocessed%5Fdata%2Etxt%2Egz
GSE108989	11138	COAD	Zhang L, Yu X, Zheng L, Zhang Y et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature 2018 Dec;564(7735):268-272.	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989
GSE98638	5063	LIHC	Zheng C, Zheng L, Yoo JK, Guo H et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell 2017 Jun 15;169(7):1342-1356.e16.	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98638
GSE102130	4058	GBMLGG	Filbin MG, Tirosh I, Hovestadt V, Shaw ML et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. Science 2018 Apr 20;360(6386):331-335.	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE102130&format=file&file=GSE102130%5FK27Mprojec

				t%2ERSEM%2Evh20170621%2Etxt%2Egz
GSE146771	10468	COAD	Zhang L, Li Z, Skrzypczynska KM, Fang Q et al. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell 2020 Apr 16;181(2):442-459.e29.	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146771
GSE99254	12346	LCA	Guo X, Zhang Y, Zheng L, Zheng C et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med 2018 Jul;24(7):978-985.	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254
GSE72056	4645	SKCM	Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 2016 Apr 8;352(6282):189-96.	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE72056&format=file&file=GSE72056%5Fmelanoma%5Fsingle%5Fcell%5Frevised%5Fv2%2Etxt%2Egz
GSE115978	7186	SKCM	Jerby-Arnon L, Shah P, Cuoco MS, Rodman C et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell 2018 Nov 1;175(4):984-997.e24.	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115978
GSE70630	4347	GBMLGG	Tirosh I, Venteicher AS, Hebert C, Escalante LE et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature 2016 Nov 10;539(7628):309-313.	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE70630&format=file&file=GSE70630%5FOG%5Fprocessed%5Fdata%5Fv2%2Etxt%2Egz

Table S2. The clustering performance of Miscell with different parameter on the *Smart-seq* dataset, related to STAR Methods.

Queue size (q)	Momentum coefficient (θ)	Davies-Bouldin Index	Calinski-Harabaz Index	V_measure Score	Normalized Mutual Information	Adjusted Rand Index
1024	0.99	0.74	30550.34	0.43	0.43	0.28
1024	0.999	0.83	20843.48	0.44	0.51	0.33
1024	0.9999	1.04	11990.71	0.40	0.40	0.55
2048	0.999	0.78	27186.49	0.42	0.42	0.25
4096	0.999	0.78	27186.49	0.42	0.42	0.25

Table S3. Clustering metrics of different methods on the *PBMC* dataset, related to Figure 2.

Assessment	Method	Mean	Standard deviation	95% confidence interval
Adjusted Rand Index	Miscell	0.431	0.002	(0.43,0.431)
	Miscell-KNN	0.307	0.002	(0.306,0.307)
	Scanpy	0.325	0.001	(0.325,0.325)
	scVI	0.239	0.001	(0.239,0.239)
	Seurat	0.288	0.001	(0.288,0.288)
Normalized Mutual Information	Miscell	0.601	0.002	(0.601,0.602)
	Miscell-KNN	0.598	0.002	(0.598,0.598)
	Scanpy	0.569	0.002	(0.569,0.569)
	scVI	0.545	0.002	(0.545,0.545)
	Seurat	0.540	0.002	(0.54,0.54)
V-measure Score	Miscell	0.591	0.002	(0.591,0.591)
	Miscell-KNN	0.574	0.002	(0.574,0.574)
	Scanpy	0.551	0.002	(0.55,0.551)
	scVI	0.516	0.002	(0.516,0.516)
	Seurat	0.514	0.002	(0.514,0.514)
Calinski-Harabaz Index	Miscell	9564.303	87.846	(9558.852,9569.755)
	Miscell-KNN	8762.874	40.394	(8760.367,8765.381)
	Scanpy	1627.366	9.561	(1626.773,1627.959)
	scVI	2430.310	13.418	(2429.477,2431.142)
	Seurat	927.235	5.023	(926.923,927.547)
Davies-Bouldin Index	Miscell	0.936	0.007	(0.936,0.937)
	Miscell-KNN	1.148	0.004	(1.148,1.149)
	Scanpy	2.217	0.007	(2.216,2.217)
	scVI	1.797	0.007	(1.796,1.797)
	Seurat	2.607	0.009	(2.607,2.608)

Table S4. Clustering metrics of different methods on the *Smart-seq* dataset, related to Figure 2.

Assessment index	Method	Mean	Standard deviation	95% confidence interval
Adjusted Rand	Miscell	0.330	0.003	(0.33,0.33)

Index	Miscell-KNN	0.147	0.001	(0.147,0.148)
	Scanpy	0.150	0.002	(0.15,0.15)
	scVI	0.094	0.001	(0.094,0.095)
	Seurat	0.080	0.002	(0.08,0.08)
Normalized Mutual Information	Miscell	0.508	0.003	(0.508,0.508)
	Miscell-KNN	0.444	0.002	(0.444,0.444)
	Scanpy	0.306	0.003	(0.305,0.306)
	scVI	0.226	0.003	(0.226,0.226)
	Seurat	0.203	0.002	(0.203,0.203)
V-measure Score	Miscell	0.442	0.003	(0.442,0.442)
	Miscell-KNN	0.358	0.002	(0.357,0.358)
	Scanpy	0.261	0.003	(0.261,0.261)
	scVI	0.188	0.003	(0.188,0.188)
	Seurat	0.198	0.002	(0.198,0.198)
Calinski-Harabaz Index	Miscell	20843.479	148.891	(20834.24,20852.71)
	Miscell-KNN	5487.240	24.157	(5485.741,5488.74)
	Scanpy	3591.719	18.161	(3590.592,3592.846)
	scVI	4288.770	18.434	(4287.626,4289.913)
	Seurat	1259.716	7.894	(1259.226,1260.206)
Davies-Bouldin Index	Miscell	0.825	0.003	(0.825,0.825)
	Miscell-KNN	1.045	0.003	(1.045,1.045)
	Scanpy	1.645	0.006	(1.644,1.645)
	scVI	1.416	0.004	(1.415,1.416)
	Seurat	2.134	0.008	(2.134,2.135)

Table S5. Cell types in the Tabula Muris dataset, related to Figure 2.

Cell type	Leukocyte	Immature T cell	Dendritic cell	Early pro-B cell
	B cell	Fibroblast	Monocyte	DN1 thymic pro-T cell
	T cell	Luminal epithelial cell of mammary gland	Lung endothelial cell	Kidney cell
	Bladder Urothelial cell	Skeletal muscle satellite cell	Immature B cell	Endothelial cell of hepatic sinusoid

	Stromal cell	Endothelial cell	Kidney capillary Endothelial cell	Duct epithelial cell
	Hematopoietic Precursor cell	Late pro-B cell	Kidney loop of Henle Ascending limb Epithelial cell	Ciliated columnar cell of Tracheobronchial tree
	Mesenchymal cell	Granulocytopoietic cell	Cardiac muscle cell	Mast cell
	Keratinocyte	Macrophage	Kidney collecting duct epithelial cell	Basophil
	Proerythroblast	Granulocyte	Erythroblast	Mesangial cell
	Bladder cell	Neuroendocrine cell	Endocardial cell	Mesenchymal stem cell
	Basal cell of epidermis	Natural killer cell	Type II pneumocyte	Epithelial cell
	Blood cell	kidney proximal straight tubule epithelial cell	Promonocyte	Langerhans cell
	Fraction A pre-pro B cell	Non-classical monocyte	Classical monocyte	Myeloid cell
	Hepatocyte	Basal cell	Alveolar macrophage	

Table S6. Clustering metrics of different methods on the *Muris* dataset, related to Figure 2.

Assessment index	Method	Mean	Standard deviation	95% confidence interval

Adjusted Rand Index	Miscell	0.769	0.003	(0.769,0.769)
	Miscell-KNN	0.696	0.002	(0.696,0.696)
	Scanpy	0.621	0.003	(0.621,0.621)
	scVI	0.519	0.002	(0.519,0.519)
	Seurat	0.524	0.002	(0.524,0.524)
Normalized Mutual Information	Miscell	0.834	0.001	(0.834,0.834)
	Miscell-KNN	0.821	0.001	(0.821,0.821)
	Scanpy	0.827	0.001	(0.827,0.827)
	scVI	0.817	0.001	(0.817,0.817)
	Seurat	0.812	0.001	(0.812,0.812)
V-measure Score	Miscell	0.834	0.001	(0.834,0.834)
	Miscell-KNN	0.821	0.001	(0.821,0.821)
	Scanpy	0.827	0.001	(0.827,0.827)
	scVI	0.817	0.001	(0.817,0.817)
	Seurat	0.812	0.001	(0.812,0.812)
Calinski-Harabaz Index	Miscell	13886.872	119.332	(13879.467,13894.277)
	Miscell-KNN	3767.196	20.558	(3765.92,3768.471)
	Scanpy	3498.473	17.984	(3497.357,3499.589)
	scVI	4479.667	20.561	(4478.391,4480.943)
	Seurat	305.999	4.129	(305.743,306.255)
Davies-Bouldin Index	Miscell	0.867	0.003	(0.866,0.867)
	Miscell-KNN	1.011	0.003	(1.01,1.011)
	Scanpy	1.373	0.005	(1.373,1.374)
	scVI	1.260	0.004	(1.26,1.261)
	Seurat	18.136	0.615	(18.097,18.174)

Table S7. The association between the latent features learned by Miscell and prognosis after controlling for Age, Sex and TNM stage in the *Smart-seq* dataset, related to Figure 4.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	q.value
Feature30	-0.377	0.686	0.091	-4.129	0.000	0.002
Feature54	0.175	1.192	0.049	3.589	0.000	0.011
Feature6	0.189	1.208	0.056	3.372	0.001	0.014
Feature27	-0.175	0.839	0.053	-3.322	0.001	0.014
Feature7	-0.228	0.796	0.075	-3.026	0.002	0.023
Feature10	-0.164	0.849	0.054	-3.045	0.002	0.023

Feature12	0.199	1.221	0.064	3.108	0.002	0.023
Feature20	-0.154	0.857	0.053	-2.904	0.004	0.030
Feature21	-0.182	0.834	0.064	-2.839	0.005	0.032
Feature4	-0.135	0.874	0.053	-2.534	0.011	0.066
Feature26	-0.349	0.706	0.136	-2.565	0.010	0.066
Feature1	-0.121	0.886	0.051	-2.357	0.018	0.080
Feature3	0.114	1.121	0.049	2.348	0.019	0.080
Feature9	-0.134	0.875	0.056	-2.384	0.017	0.080
Feature28	-0.158	0.854	0.068	-2.327	0.020	0.080
Feature49	-0.132	0.876	0.055	-2.409	0.016	0.080
Feature29	0.123	1.131	0.054	2.299	0.021	0.081
Feature19	0.120	1.127	0.055	2.179	0.029	0.104
Feature25	0.090	1.094	0.050	1.798	0.072	0.243
Feature22	0.083	1.086	0.051	1.634	0.102	0.327
Feature15	0.249	1.283	0.173	1.441	0.150	0.427
Feature16	0.073	1.076	0.051	1.427	0.154	0.427
Feature23	0.072	1.074	0.049	1.447	0.148	0.427
Feature17	0.065	1.067	0.054	1.210	0.226	0.537
Feature33	0.551	1.735	0.450	1.224	0.221	0.537
Feature36	0.551	1.735	0.450	1.224	0.221	0.537
Feature41	0.551	1.735	0.450	1.224	0.221	0.537
Feature18	0.066	1.068	0.056	1.174	0.241	0.550
Feature13	-0.052	0.949	0.053	-0.990	0.322	0.624
Feature32	0.055	1.057	0.052	1.049	0.294	0.624
Feature35	0.392	1.480	0.411	0.952	0.341	0.624
Feature38	0.392	1.480	0.411	0.952	0.341	0.624
Feature40	0.056	1.057	0.053	1.057	0.291	0.624
Feature43	0.392	1.480	0.411	0.952	0.341	0.624
Feature47	0.392	1.480	0.411	0.952	0.341	0.624
Feature44	0.179	1.196	0.210	0.851	0.395	0.701
Feature5	0.043	1.044	0.055	0.782	0.434	0.751
Feature8	-0.034	0.967	0.057	-0.593	0.553	0.885
Feature11	-0.035	0.966	0.058	-0.602	0.547	0.885
Feature52	-0.031	0.969	0.052	-0.607	0.544	0.885
Feature2	0.031	1.031	0.055	0.561	0.575	0.897
Feature24	0.012	1.013	0.051	0.246	0.806	0.901
Feature31	0.029	1.029	0.058	0.501	0.617	0.901
Feature34	0.138	1.148	0.357	0.386	0.700	0.901
Feature37	0.138	1.148	0.357	0.386	0.700	0.901
Feature39	0.138	1.148	0.357	0.386	0.700	0.901
Feature42	0.138	1.148	0.357	0.386	0.700	0.901

Feature45	-0.105	0.900	0.306	-0.345	0.730	0.901
Feature46	0.138	1.148	0.357	0.386	0.700	0.901
Feature48	0.138	1.148	0.357	0.386	0.700	0.901
Feature51	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature53	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature55	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature56	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature57	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature58	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature59	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature60	-0.021	0.979	0.051	-0.407	0.684	0.901
Feature61	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature62	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature63	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature64	-0.024	0.976	0.151	-0.160	0.873	0.901
Feature14	0.003	1.003	0.054	0.063	0.950	0.952
Feature50	0.003	1.003	0.051	0.061	0.952	0.952

Table S8. Top-15 marker genes of CD4+ T cell and CD8+ T cell clusters identified by three methods, related to Figure 3.

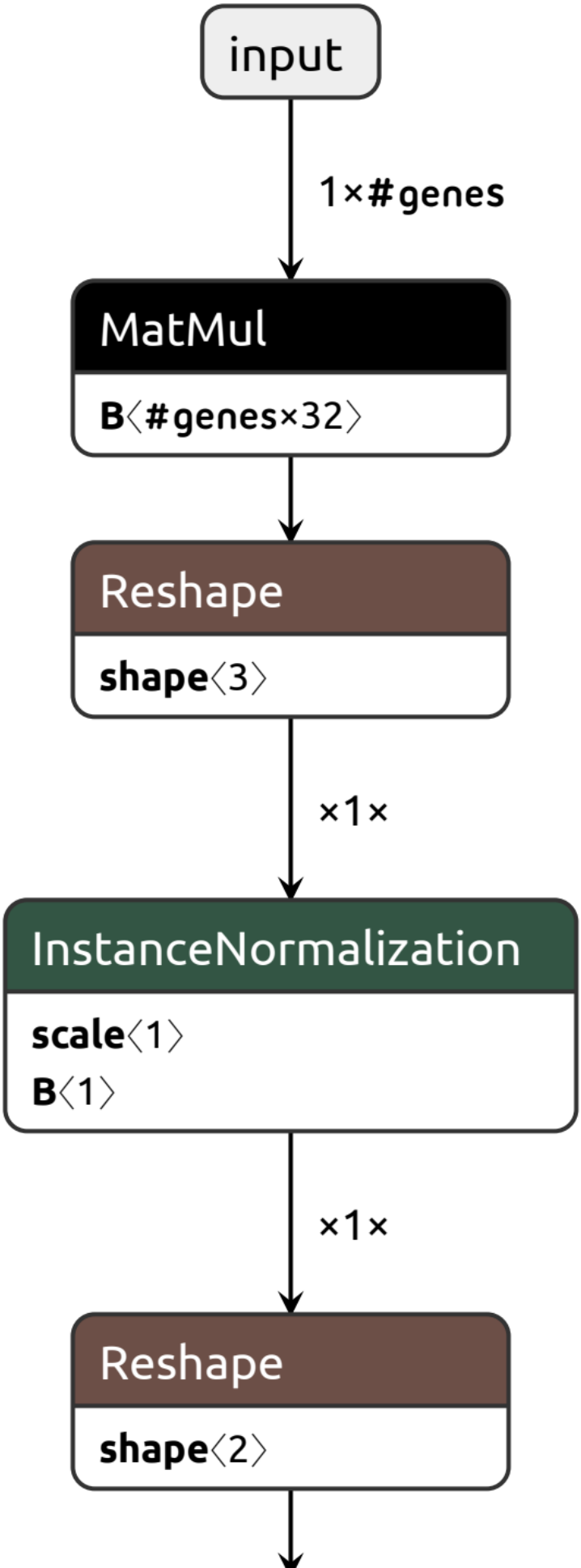
CD4+ T cell			CD8+ T cell		
Miscell	Scanpy	Seurat	Miscell	Scanpy	Seurat
CD4	*CD4*	*TNFRSF4*	*CD8A*	*CD8A*	*KLRK1*
CD40LG	*LTB*	*CD4*	*KLRK1*	*KLRK*	*CD8A*
FBLN7	*IL2RA*	*IL2RA*	*CD8B*	*CD8B*	*KLRC4-KLRK1*
TNFRSF25	*TNFRSF4*	*CD40LG*	*KLRD1*	*NKG7*	*NKG7*
TBC1D4	*TNFRSF25*	*LTB*	*KLRC4*	*CCL5*	*CCL5*
SLAMF1	*PIM2*	*FOXP3*	*KLRC3*	*KLRD1*	*CTSW*
TMEM173	*TBC1D4*	*CD28*	*KLRC2*	*CTSW*	*CD8B*
KLRB1	*ICOS*	*FBLN7*	*KLRC1*	*GZMA*	*GZMB*
ICOS	*CD40LG*	*TNFRSF25*	*CTSW*	*SLAMF7*	*GZMH*
CCR4	*GPR183*	*CORO1B*	*AOAH*	*CCL4*	*CCL4*
FOXP3	*CTLA4*	*TBC1D4*	*ABCB1*	*AOAH*	*KLRD1*
ICA1	*FOXP3*	*CTSB*	*CCL5*	*GZMH*	*GZMA*
TBXAS1	*TMEM173*	*PIM2*	*XCL1*	*PRF1*	*KLRC2*

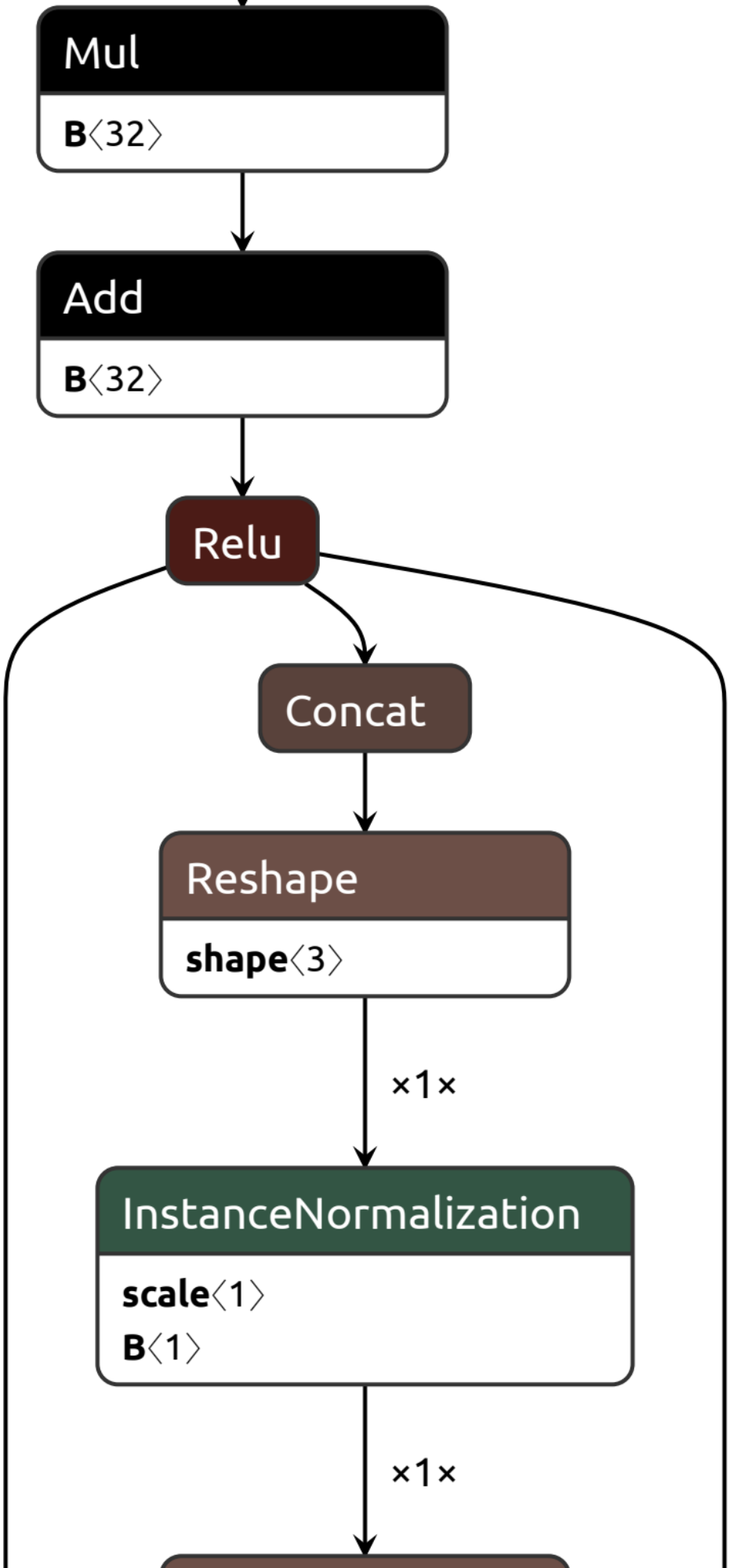
PIK3IP1	*SELL*	*IL6R*	*CD160*	*KLRC4*	*TARP*
FAAH2	*SLAMF1*	*CTLA4*	*CD55*	*CST7*	*KLRC3*

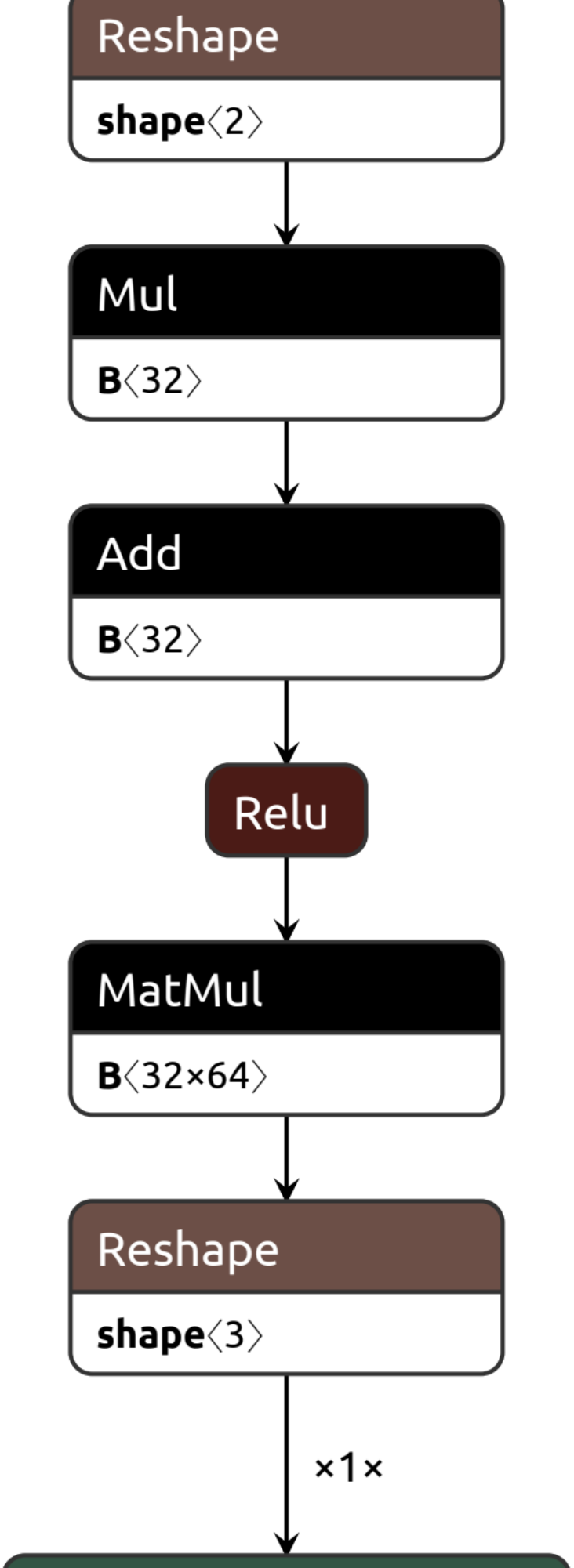
Table S9. Features and capabilities for the methods, related to STAR Methods, related to STAR Methods.

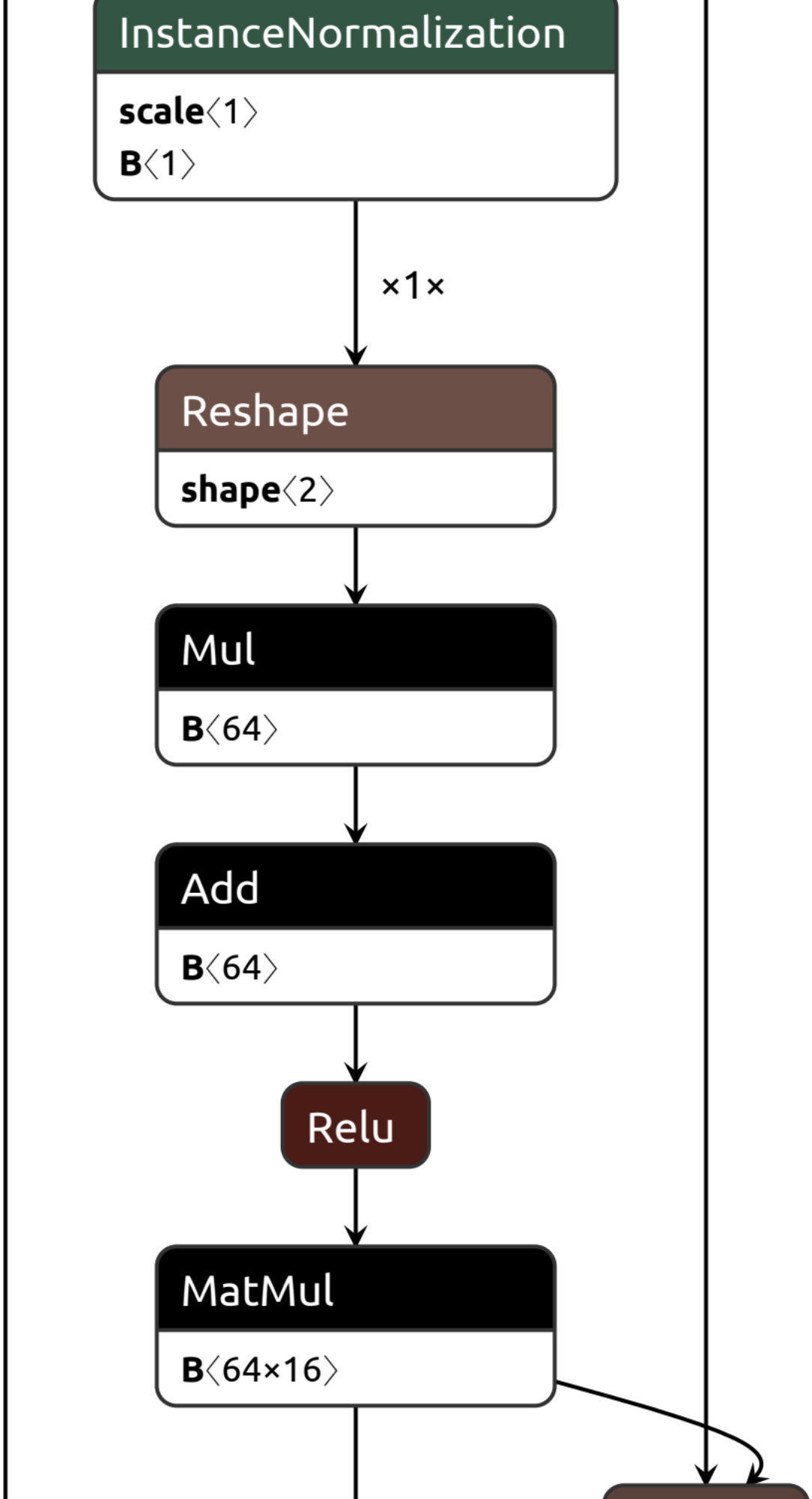
	Miscell	Scanpy	scVI	Seurat
Core procedure	Self-supervised learning	PCA	Deep generative model	PCA
Cluster delineation	DBSCAN	KNN	KNN	SNN
Marker gene detection strategy	Integrated gradient	t-test/ Wilcoxon test	-	MAST/ t-test/ Wilcoxon test

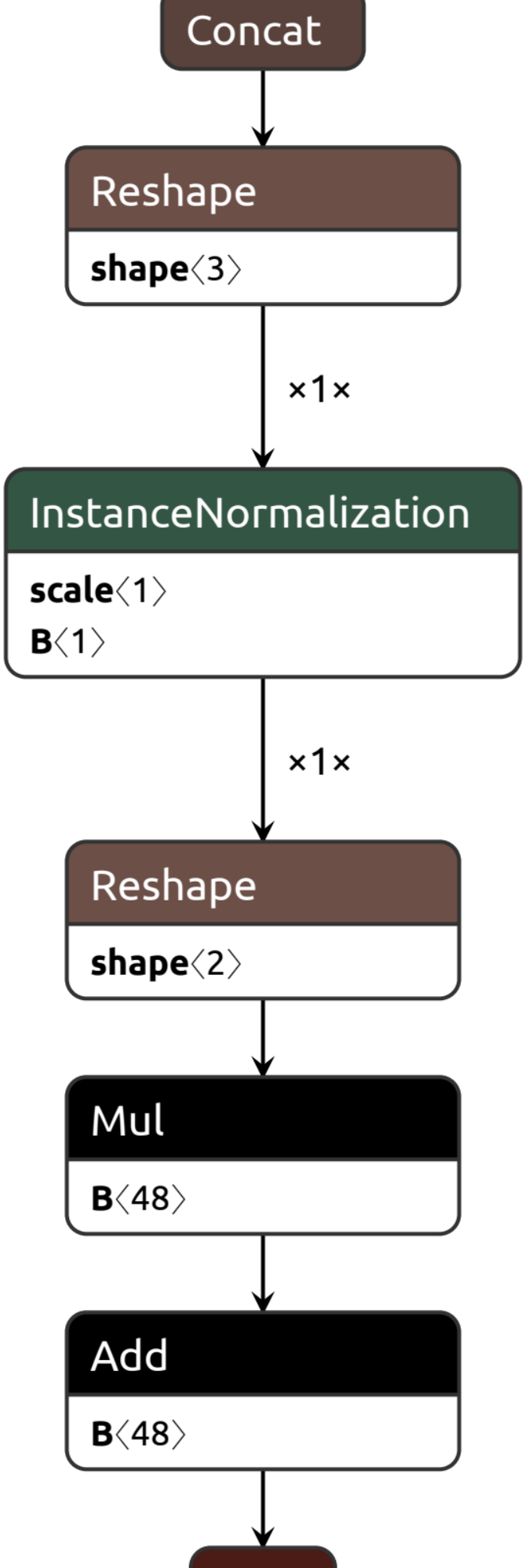
Data S1. The feature encoder used in the self-supervised learning framework is a DenseNet with 21 layers, related to STAR Methods.

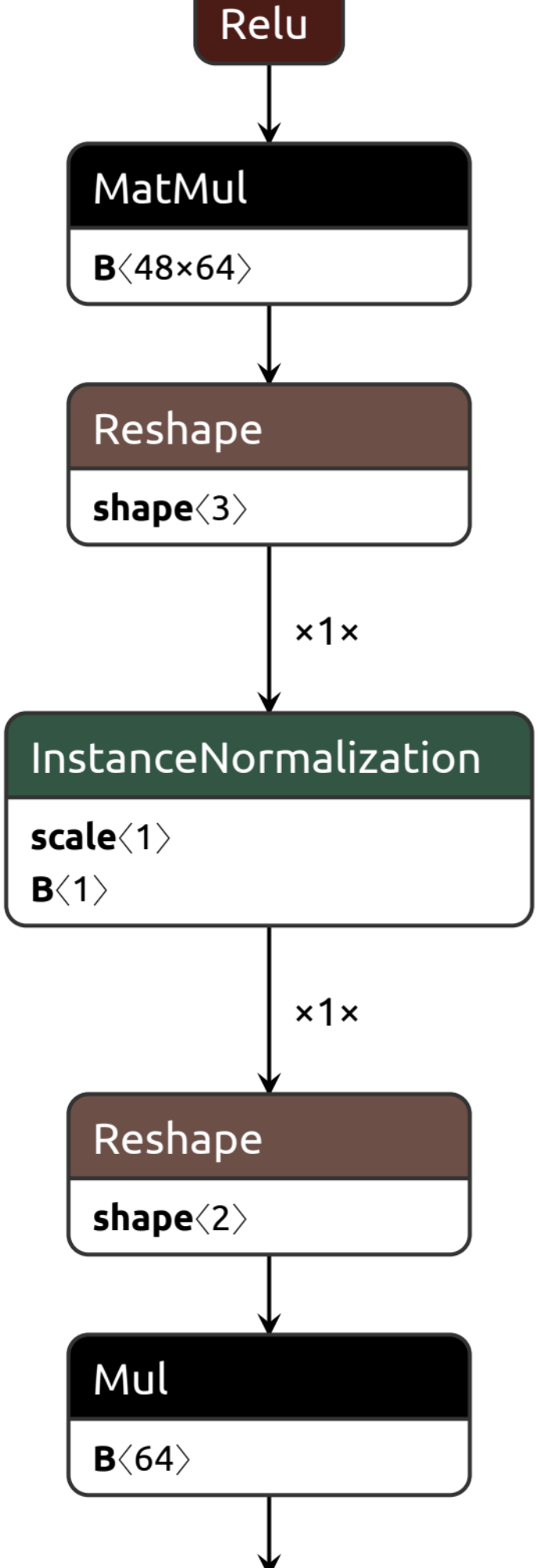

input
1×#genes
MatMul
B⟨#genes×32⟩
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩

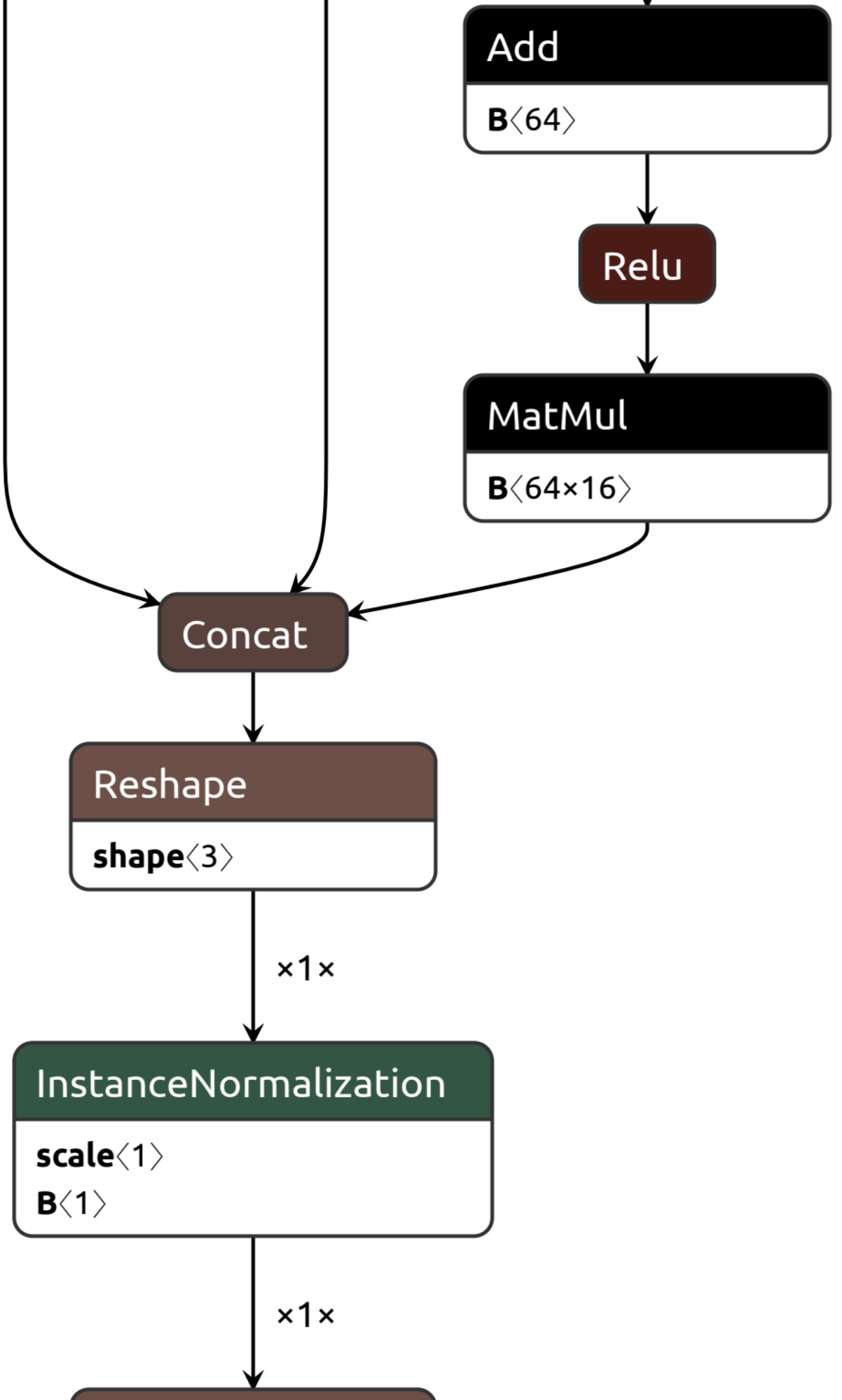

Mul
B⟨32⟩
Add
B⟨32⟩
Relu
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

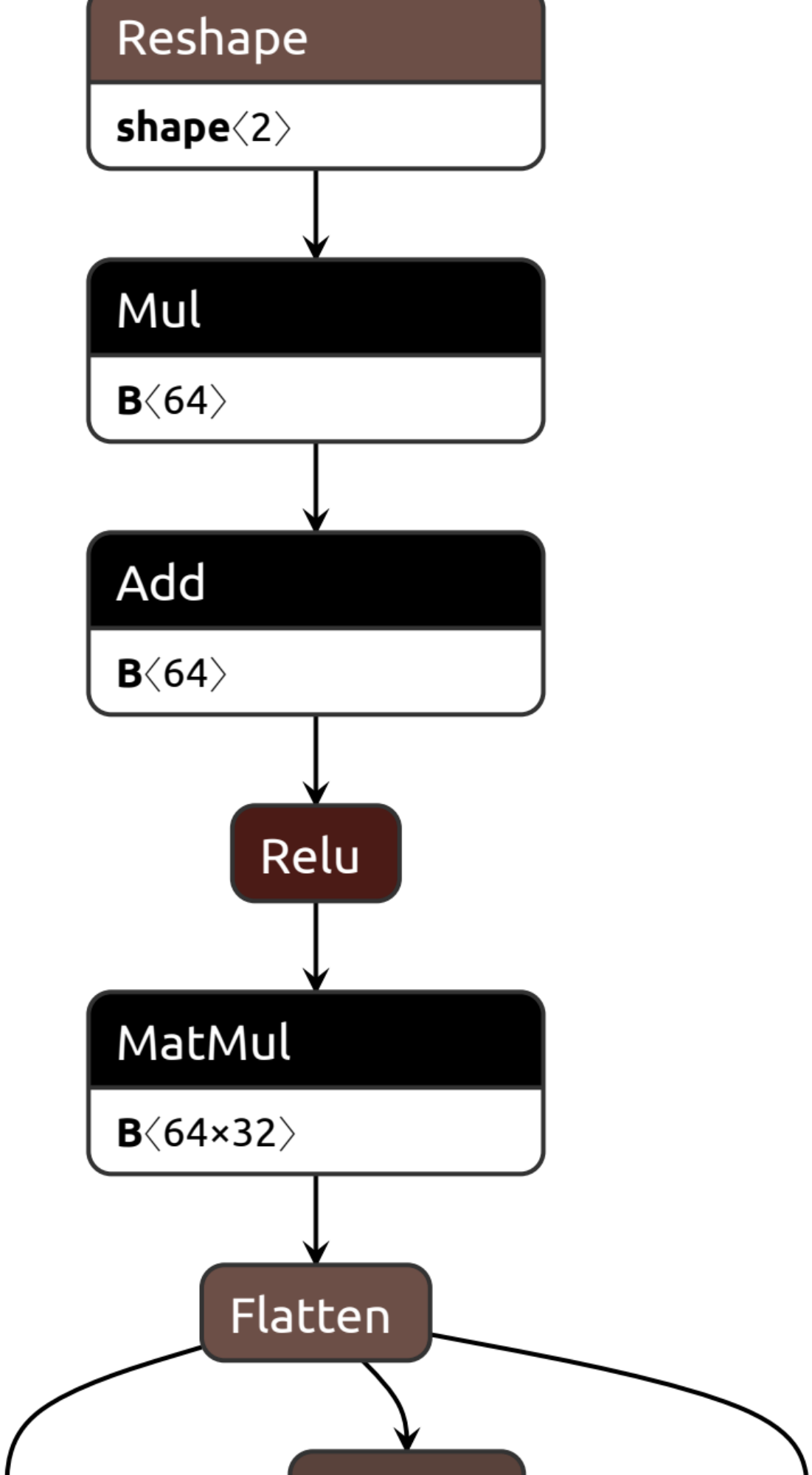

Reshape
shape⟨2⟩
Mul
B⟨32⟩
Add
B⟨32⟩
Relu
MatMul
B⟨32×64⟩
Reshape
shape⟨3⟩
×1×

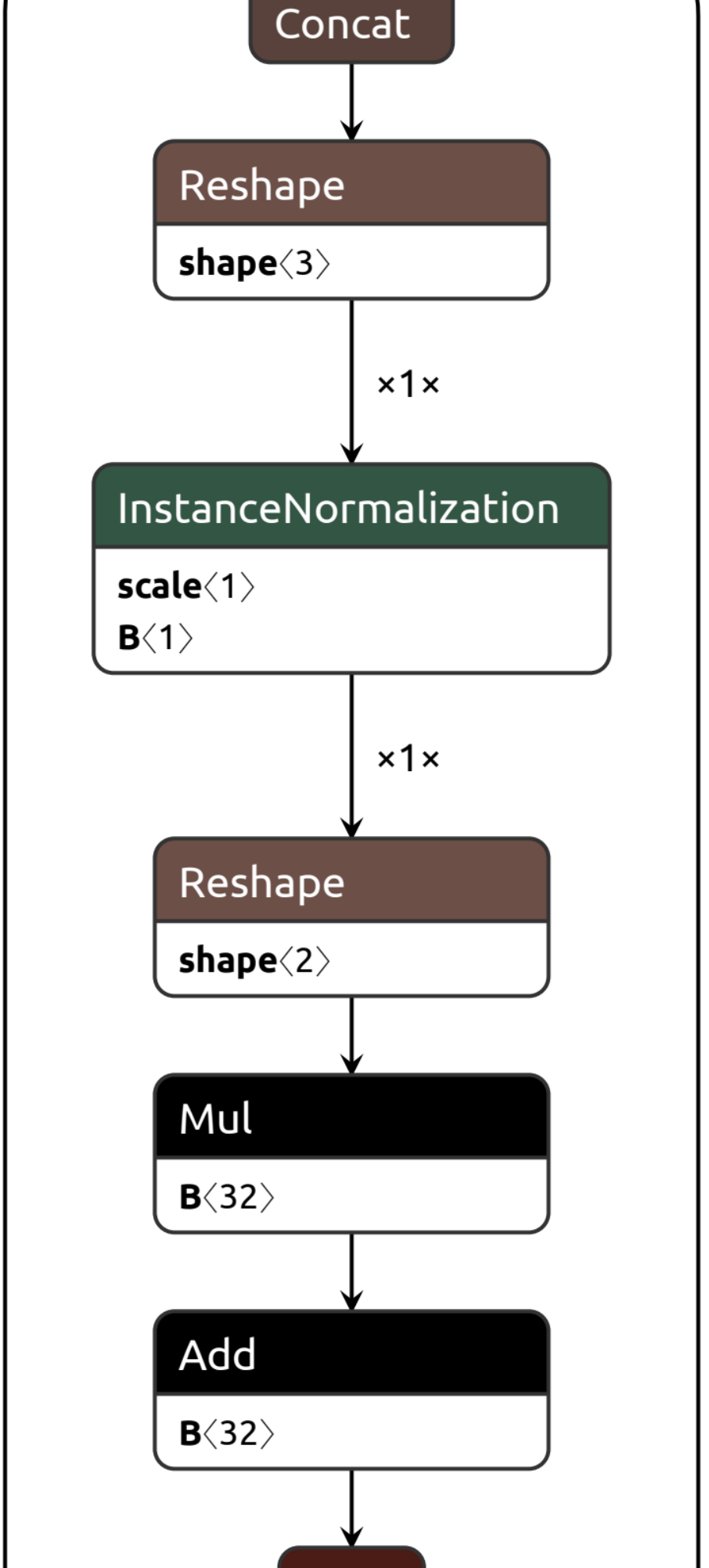

InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩

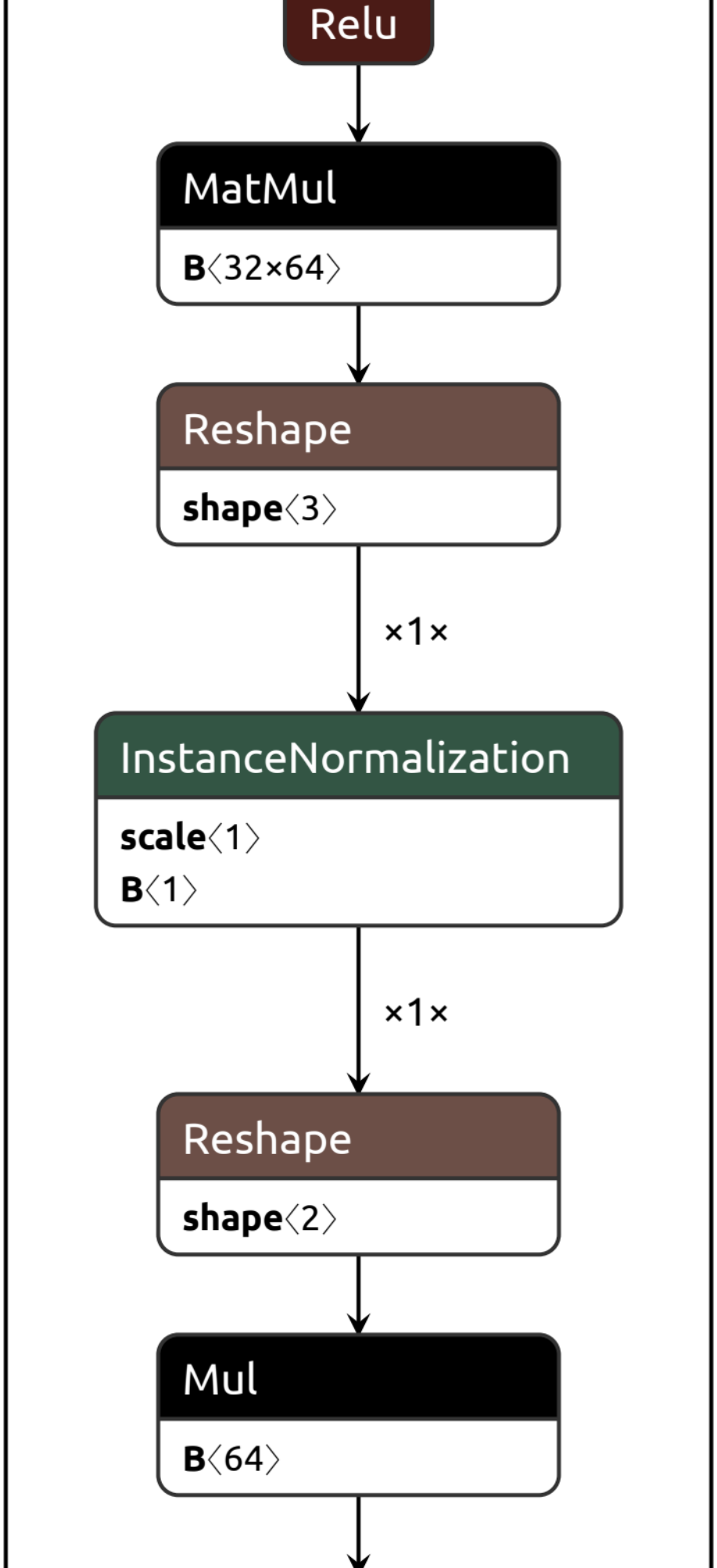

Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨48⟩
Add
B⟨48⟩

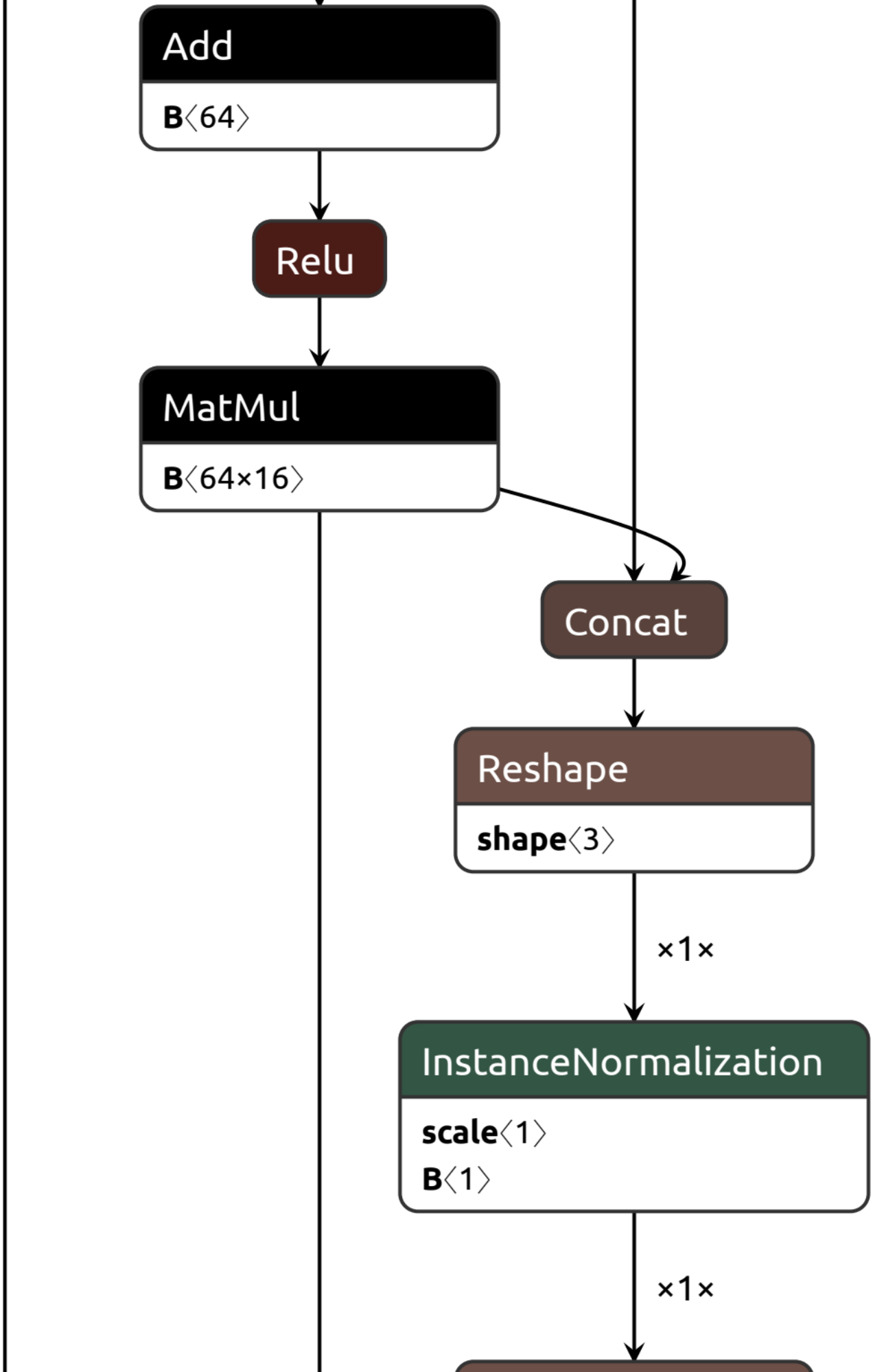

Relu
MatMul
B⟨48×64⟩
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩

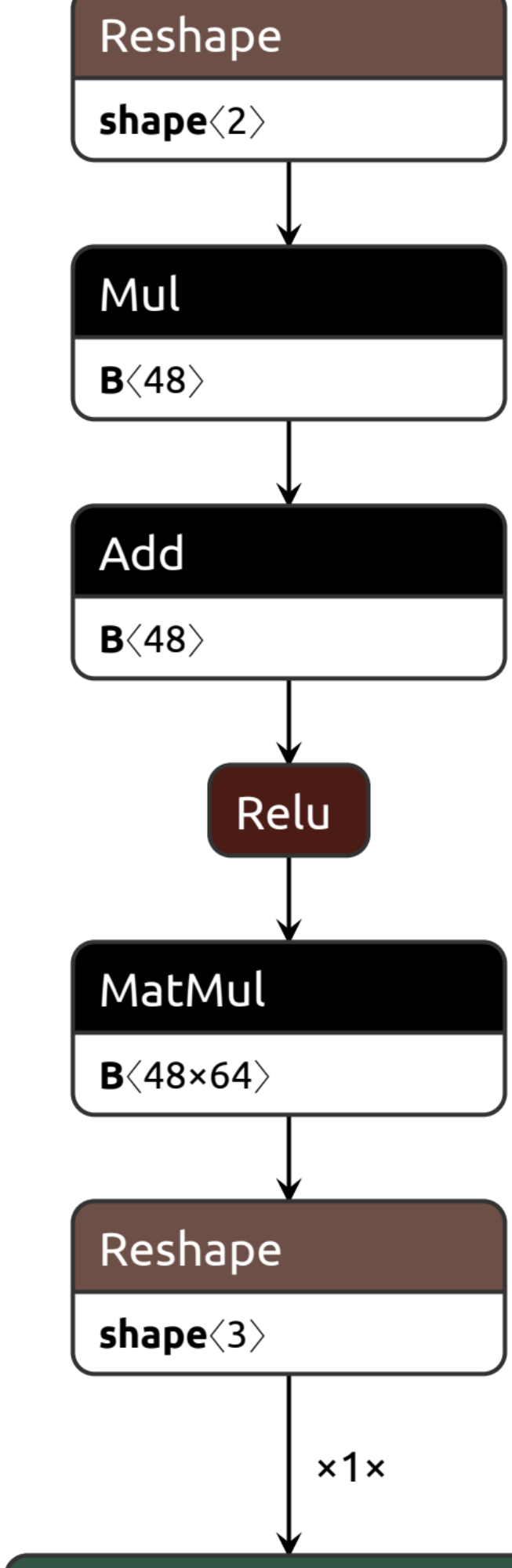

Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

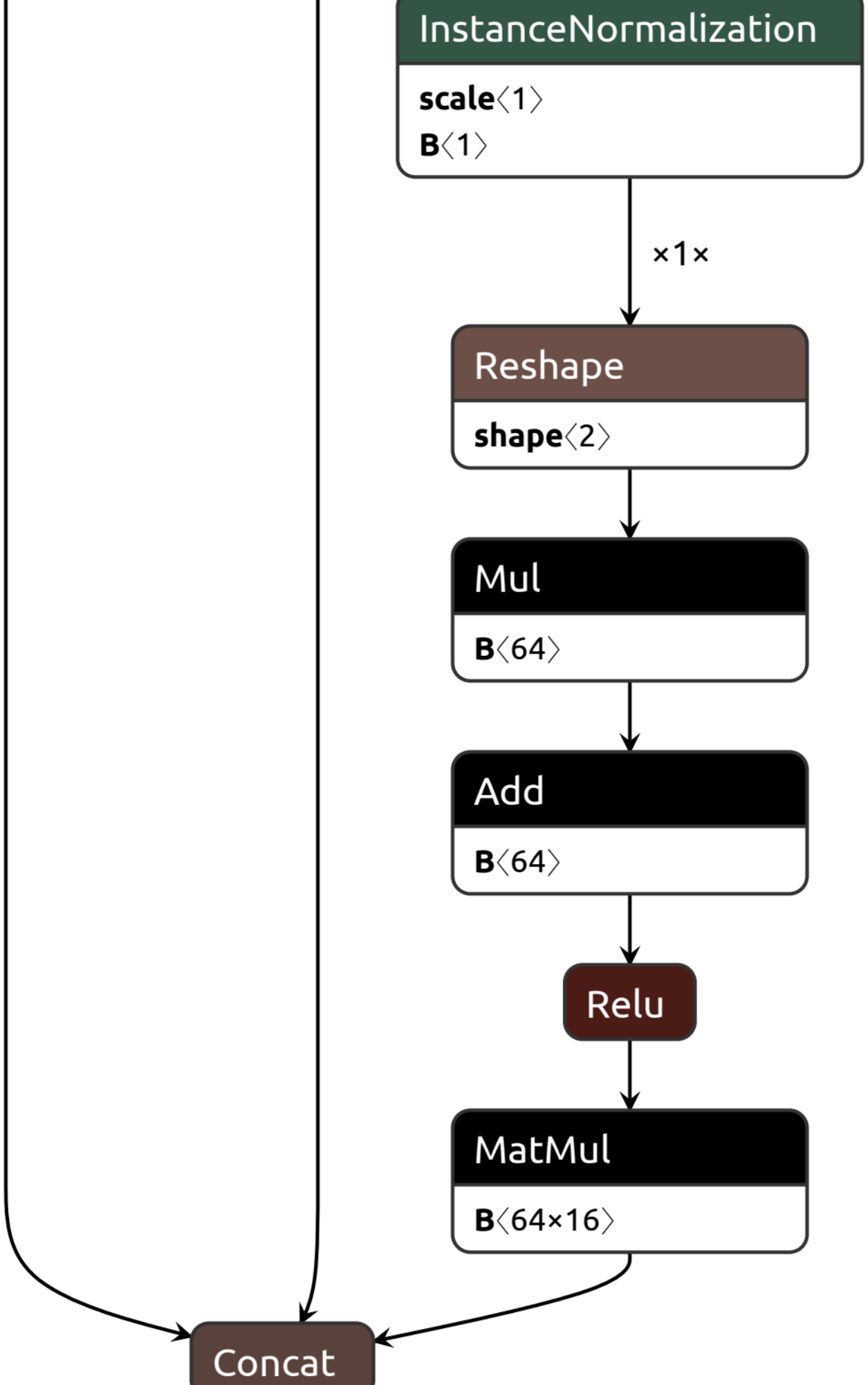

Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×32⟩
Flatten

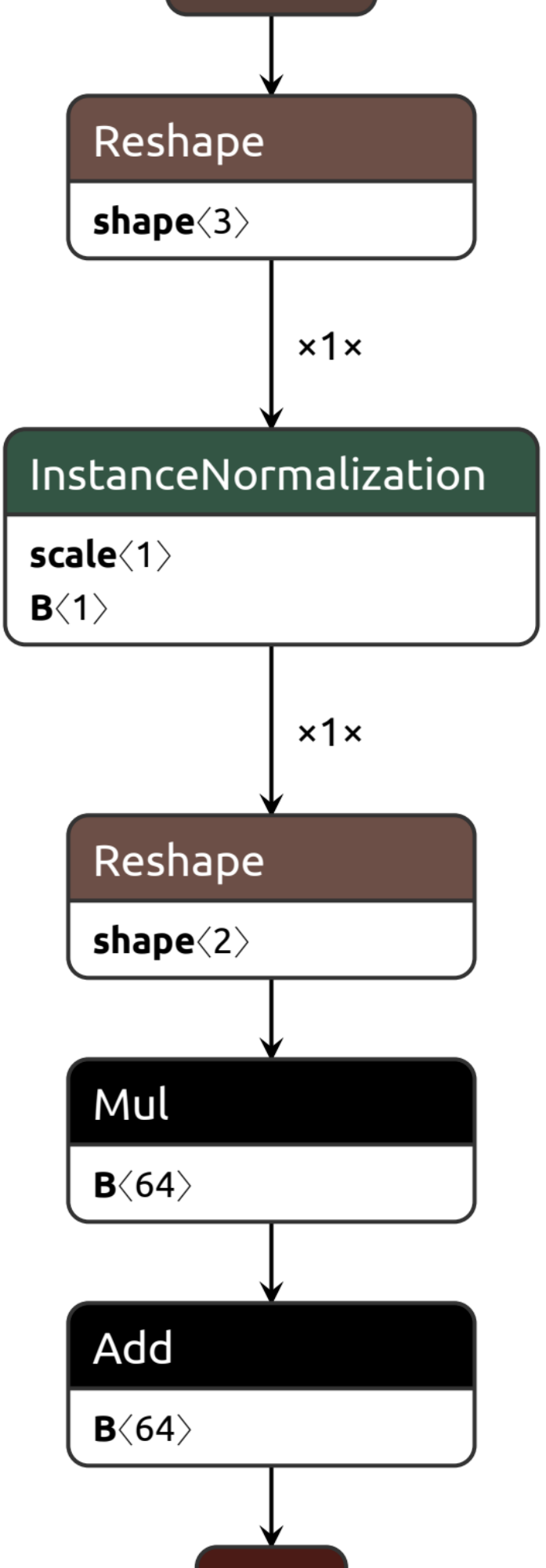

Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨32⟩
Add
B⟨32⟩

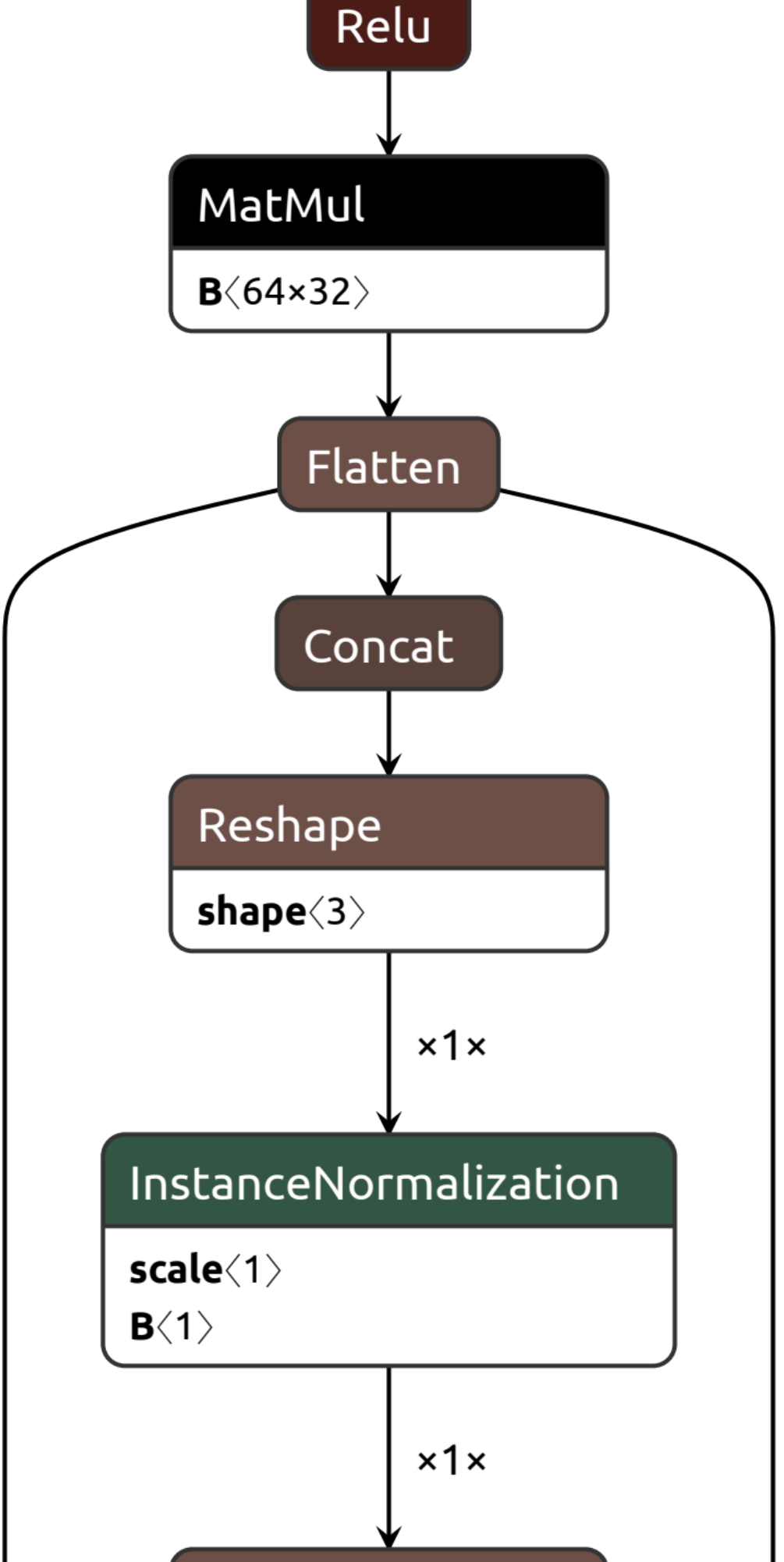

Relu
MatMul
B⟨32×64⟩
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩

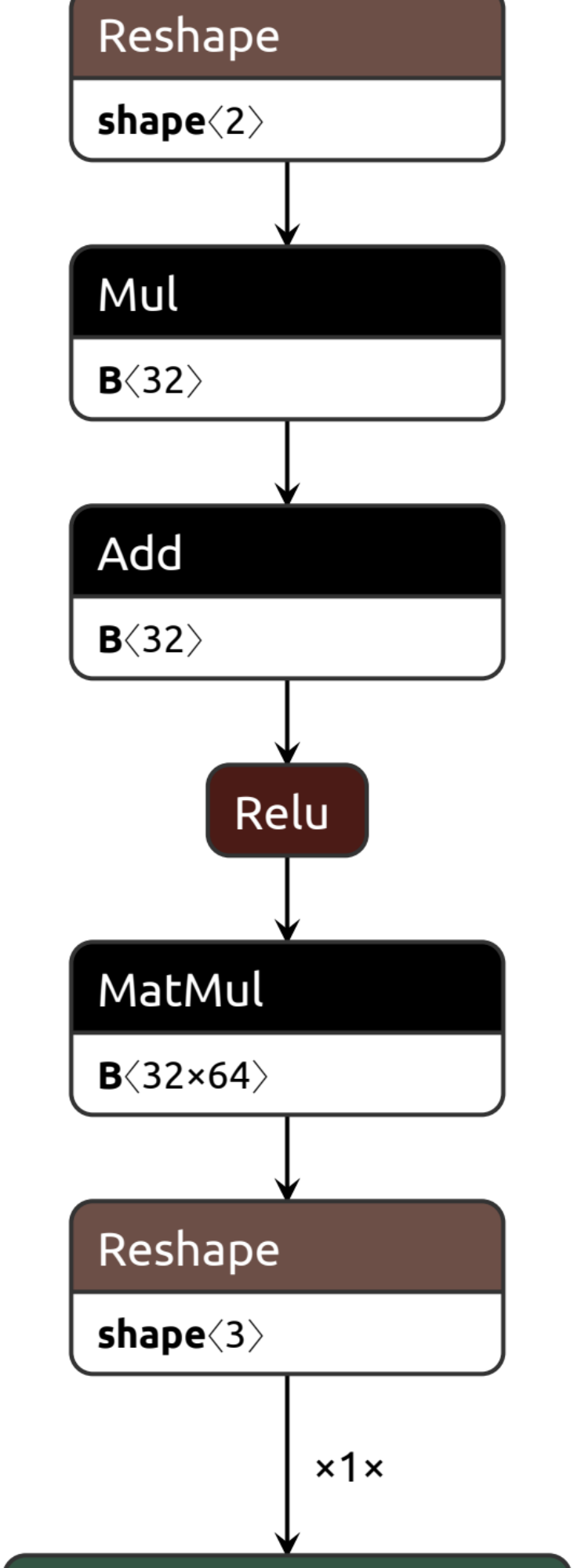

Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

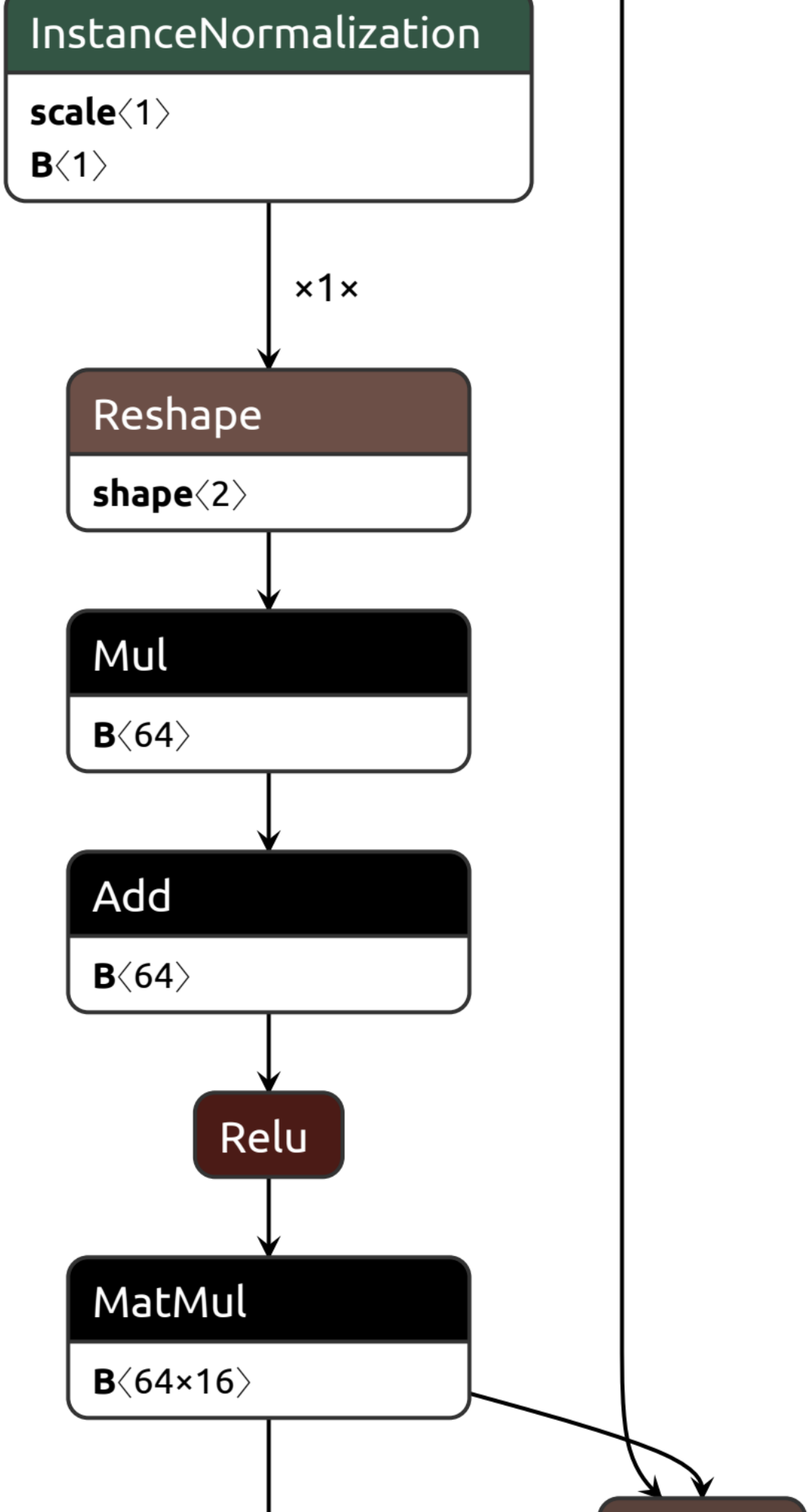

Reshape
shape⟨2⟩
Mul
B⟨48⟩
Add
B⟨48⟩
Relu
MatMul
B⟨48×64⟩
Reshape
shape⟨3⟩
×1×

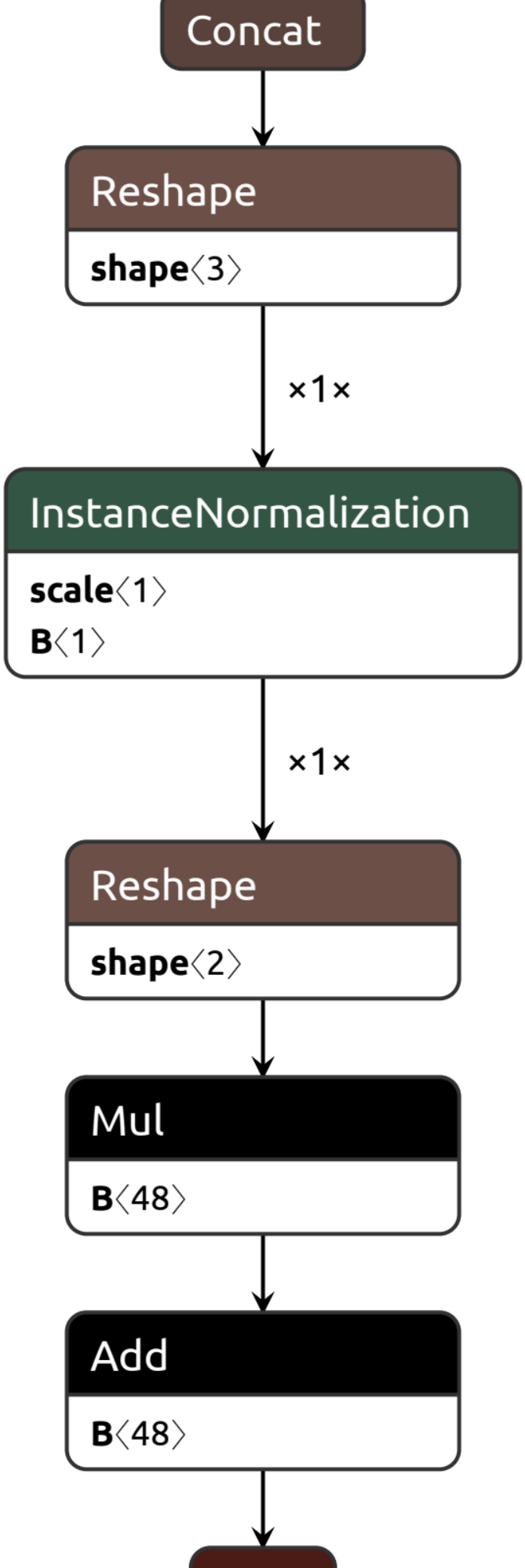

InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat

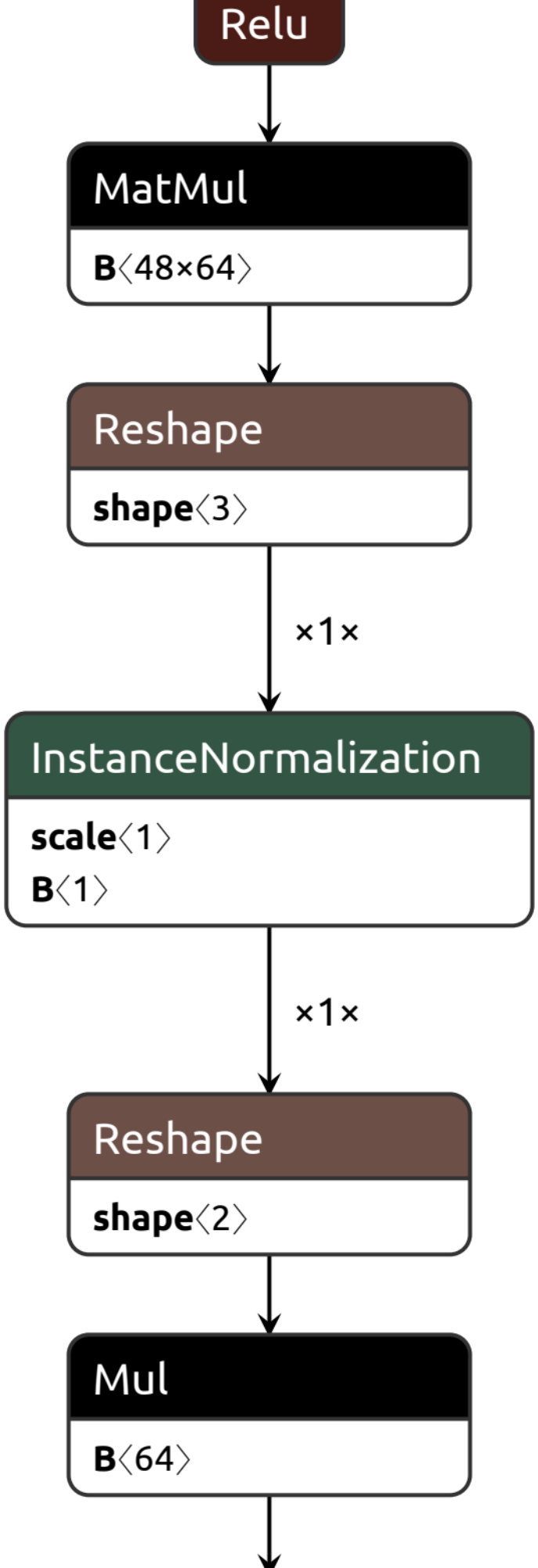

Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩

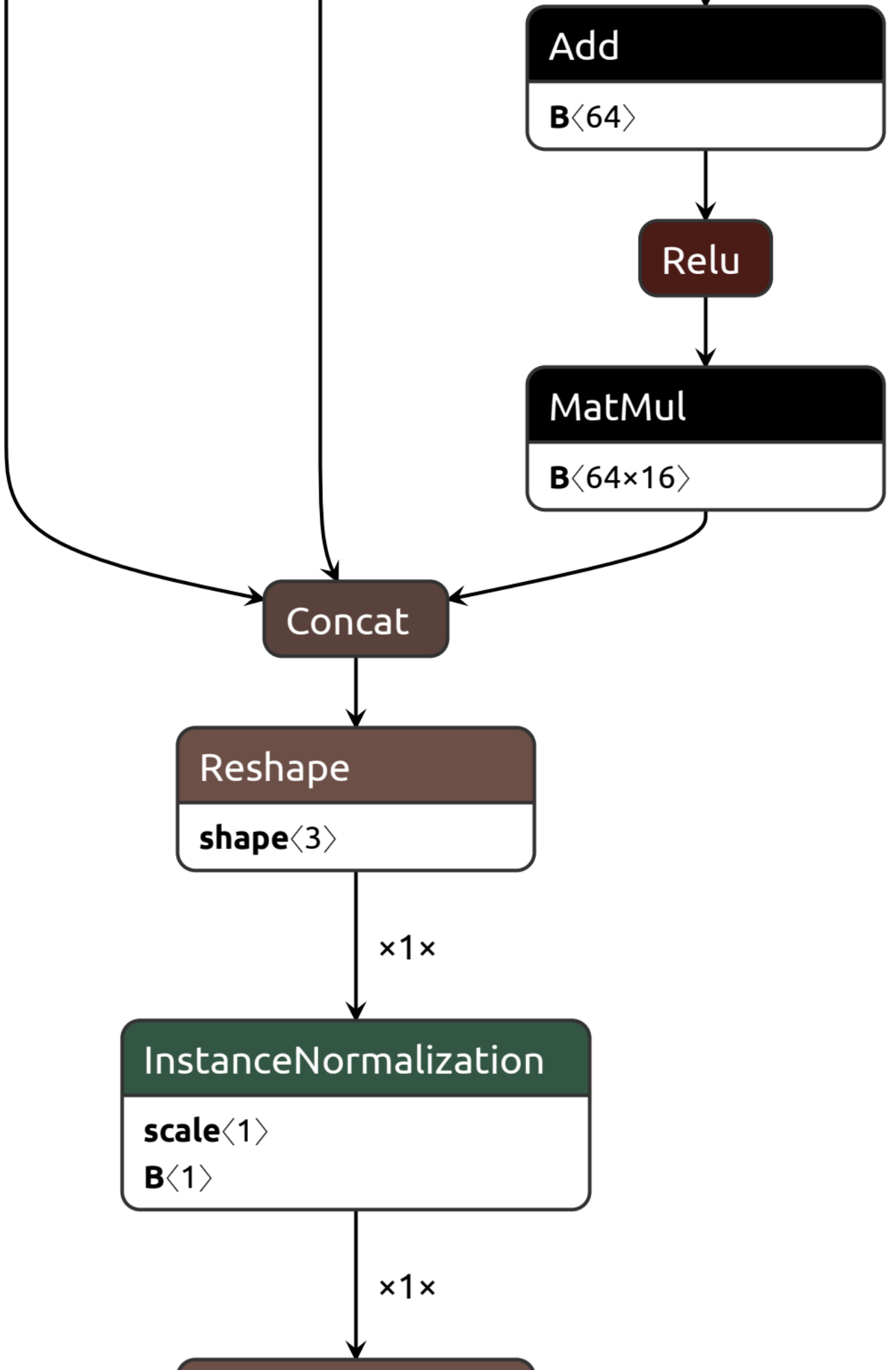

Relu
MatMul
B⟨64×32⟩
Flatten
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

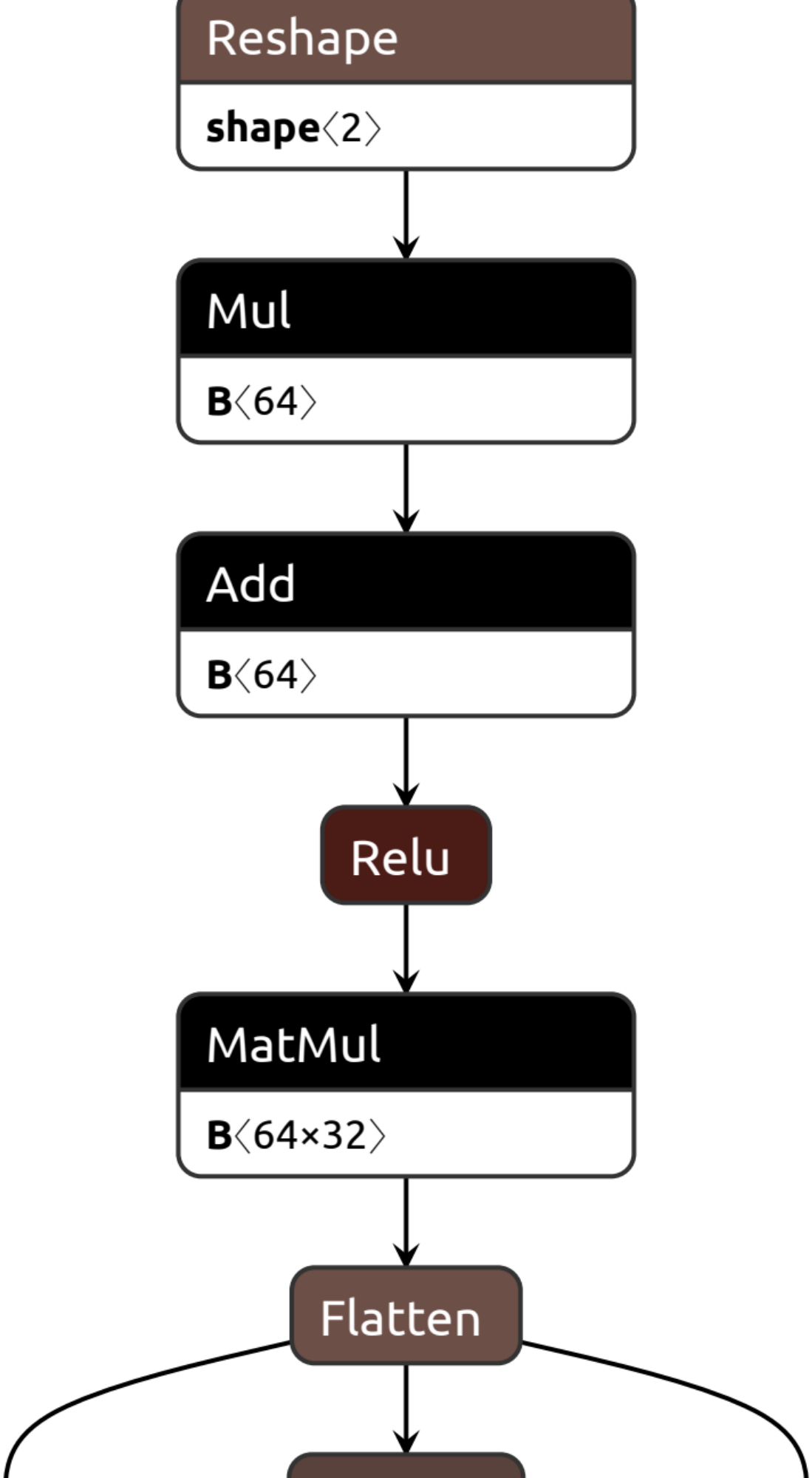

Reshape
shape⟨2⟩
Mul
B⟨32⟩
Add
B⟨32⟩
Relu
MatMul
B⟨32×64⟩
Reshape
shape⟨3⟩
×1×

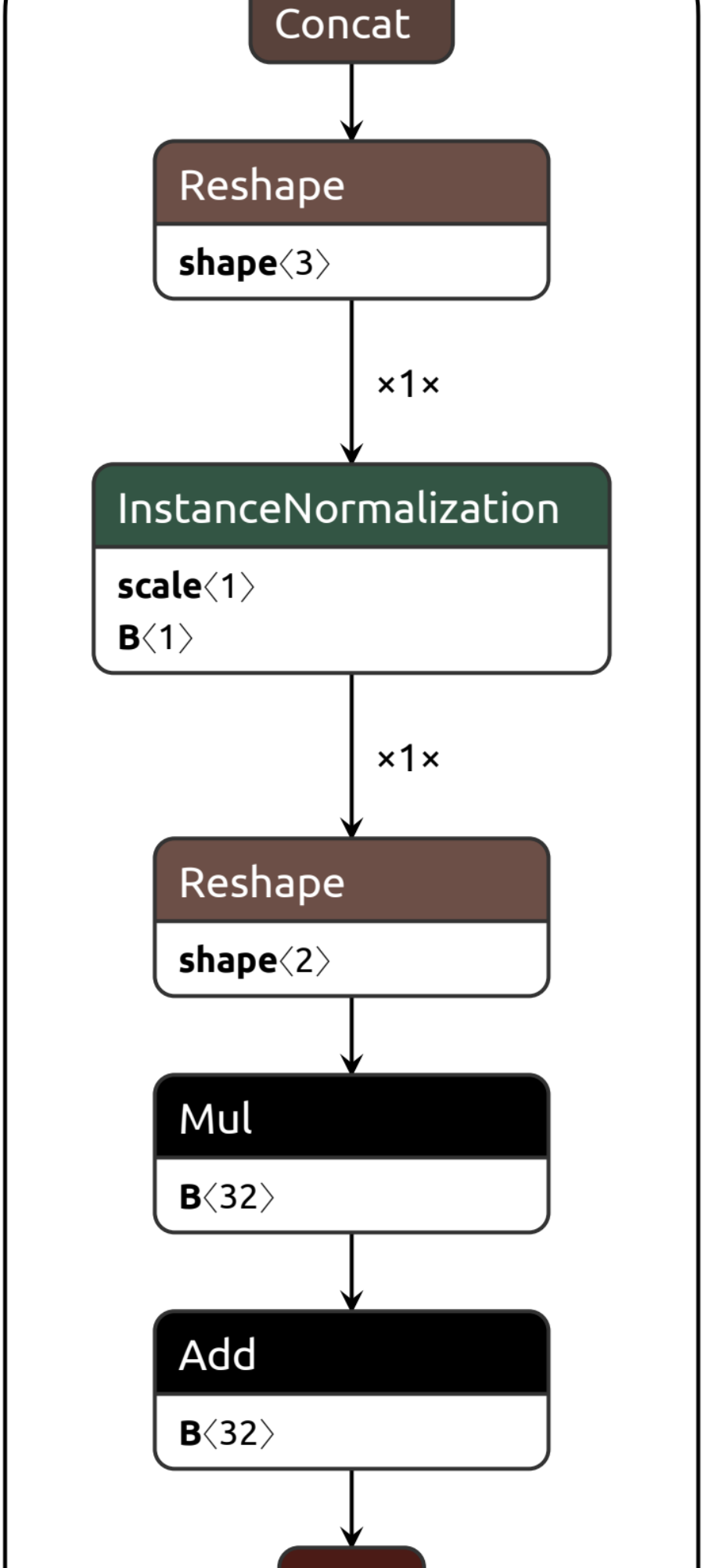

InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩

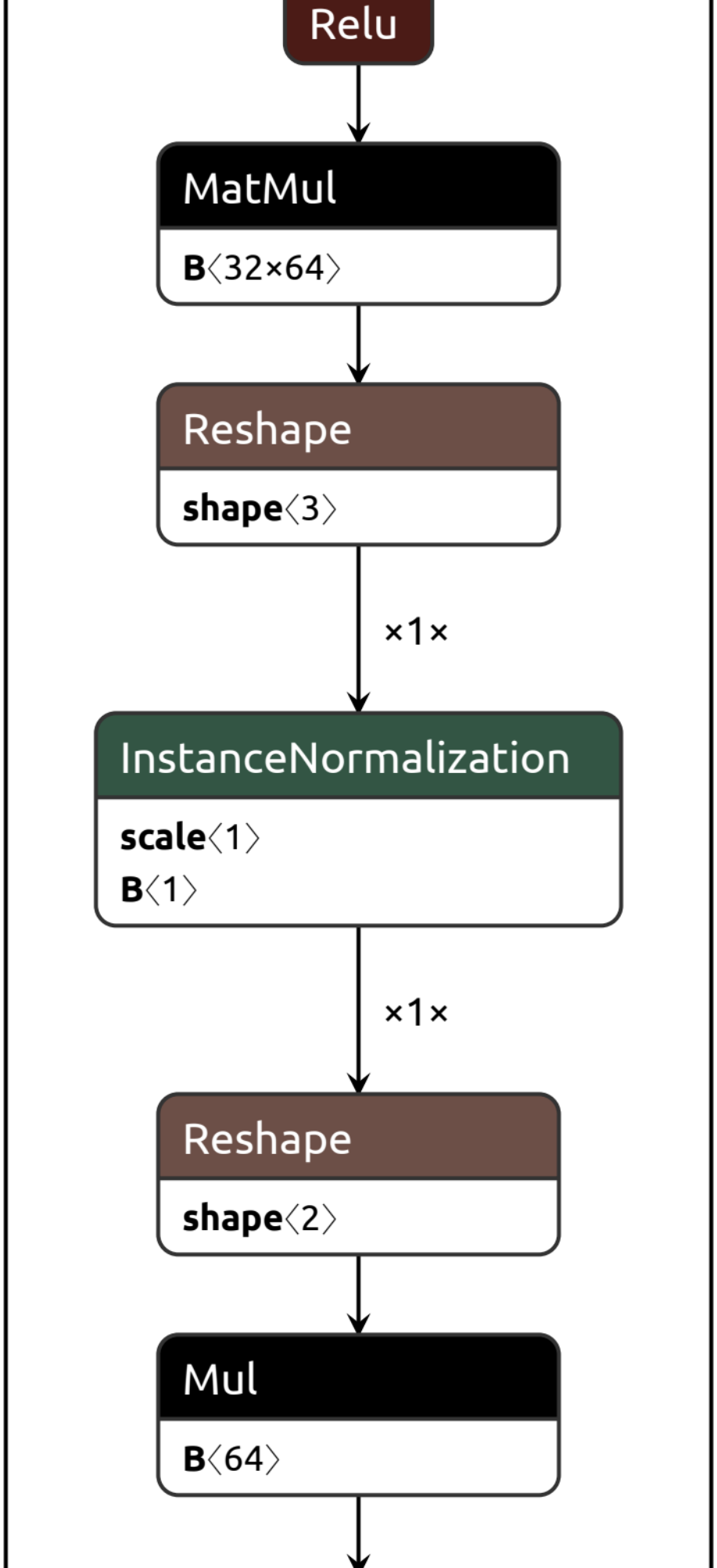

Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨48⟩
Add
B⟨48⟩

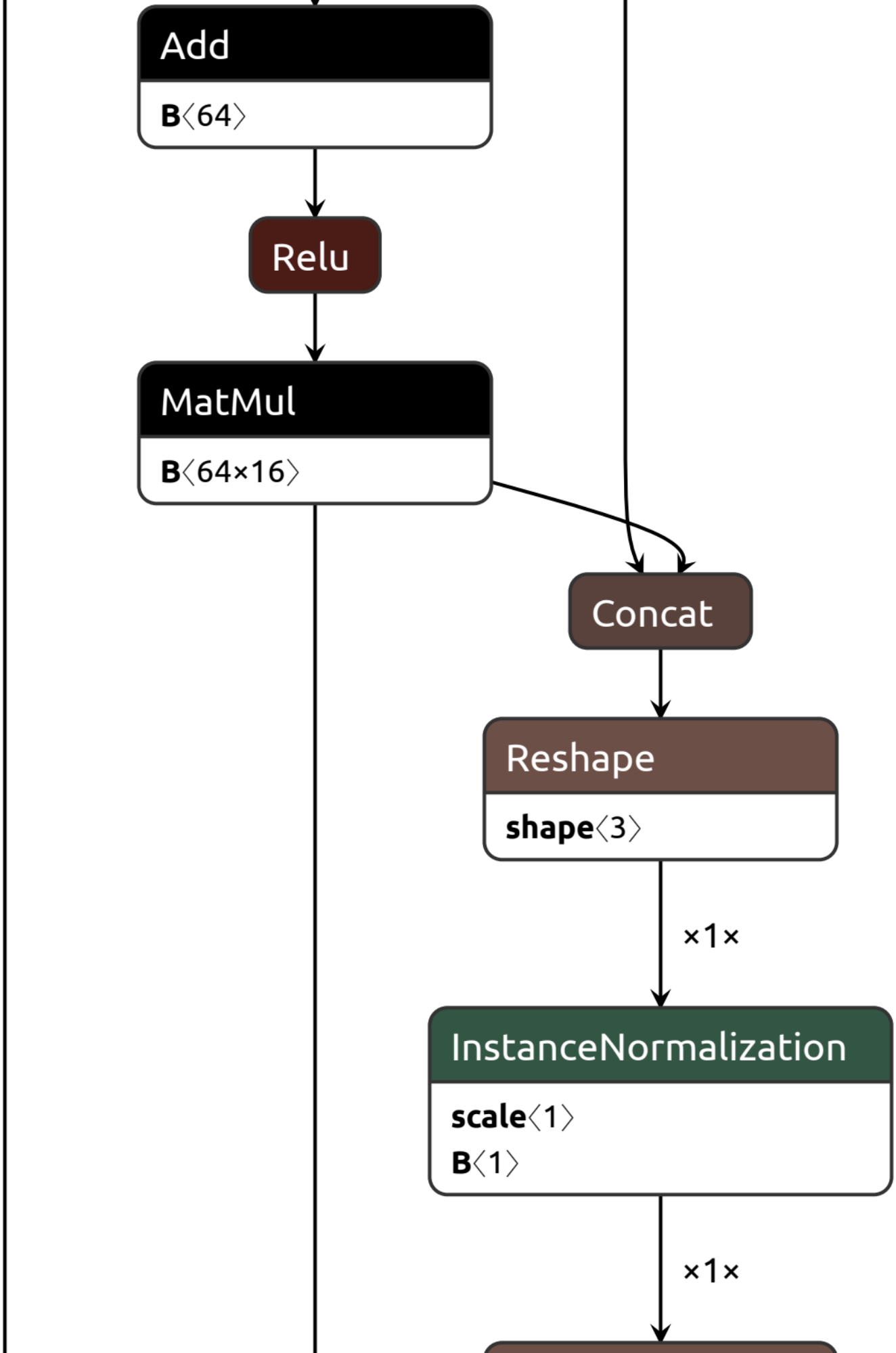

Relu
MatMul
B⟨48×64⟩
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩

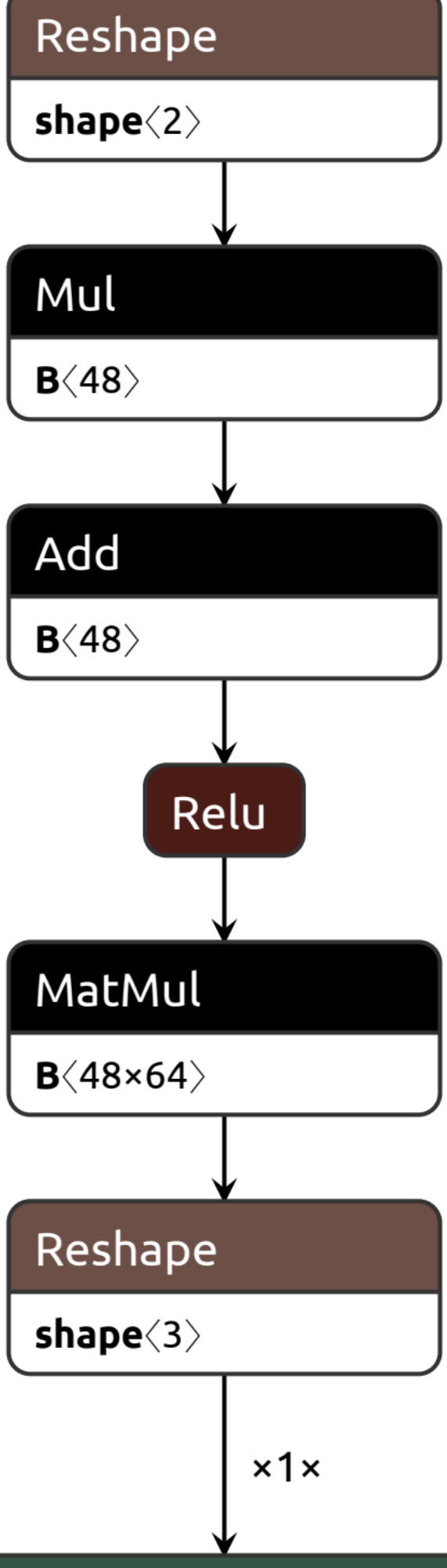

Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

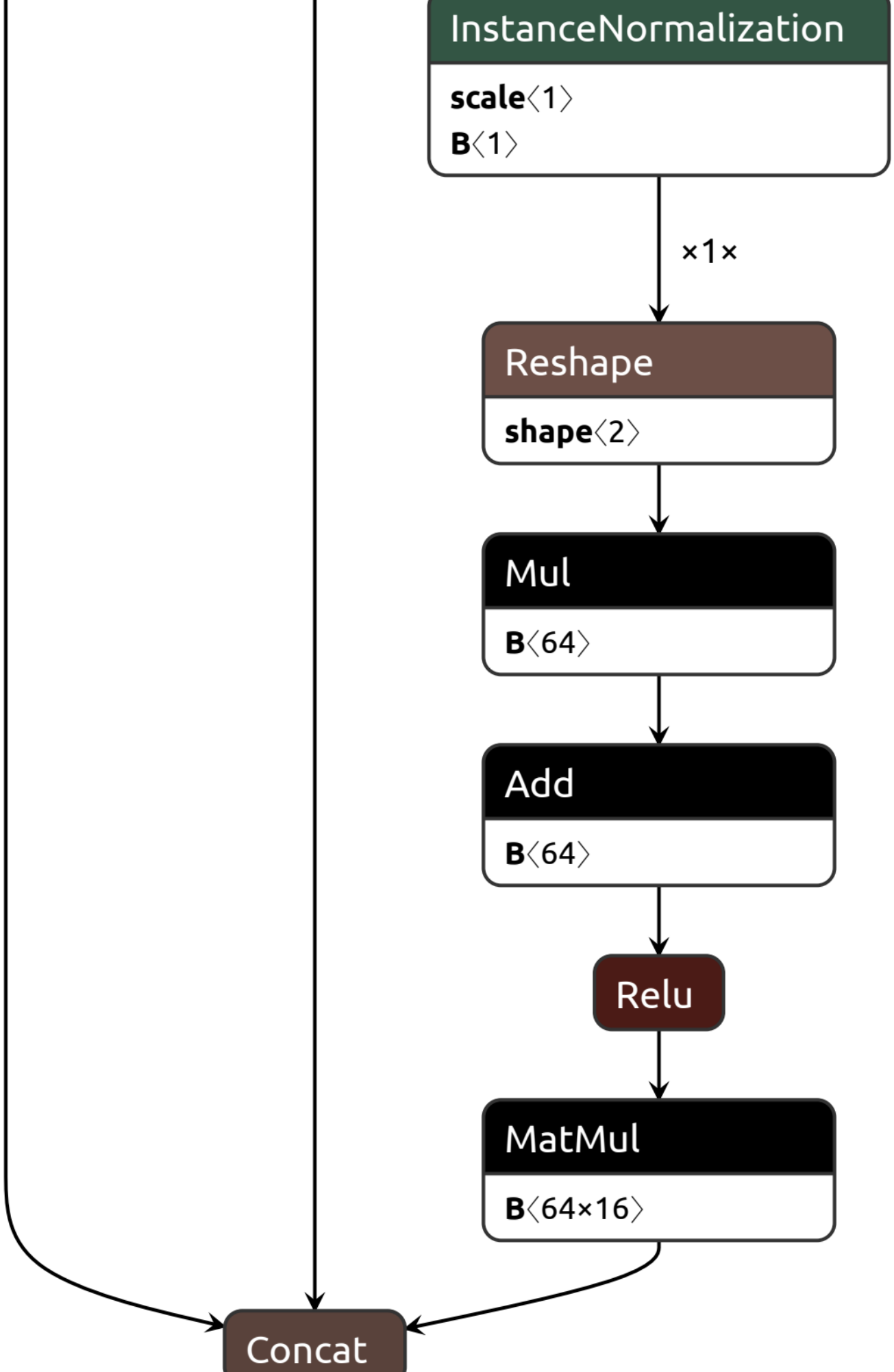

Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×32⟩
Flatten

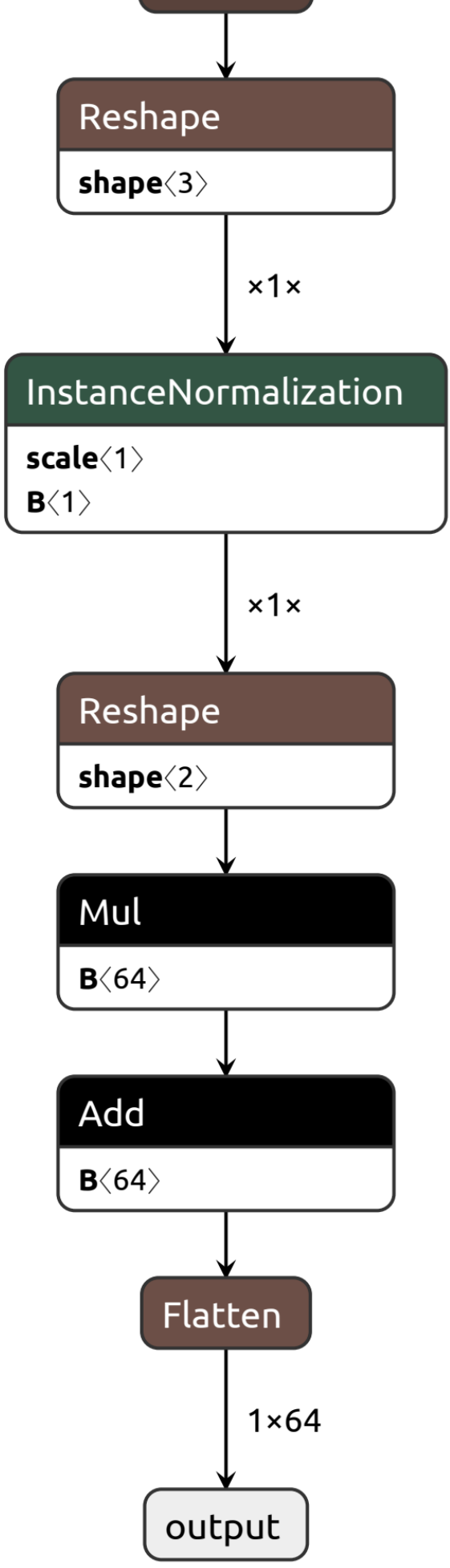

Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨32⟩
Add
B⟨32⟩


Relu
MatMul
B⟨32×64⟩
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩

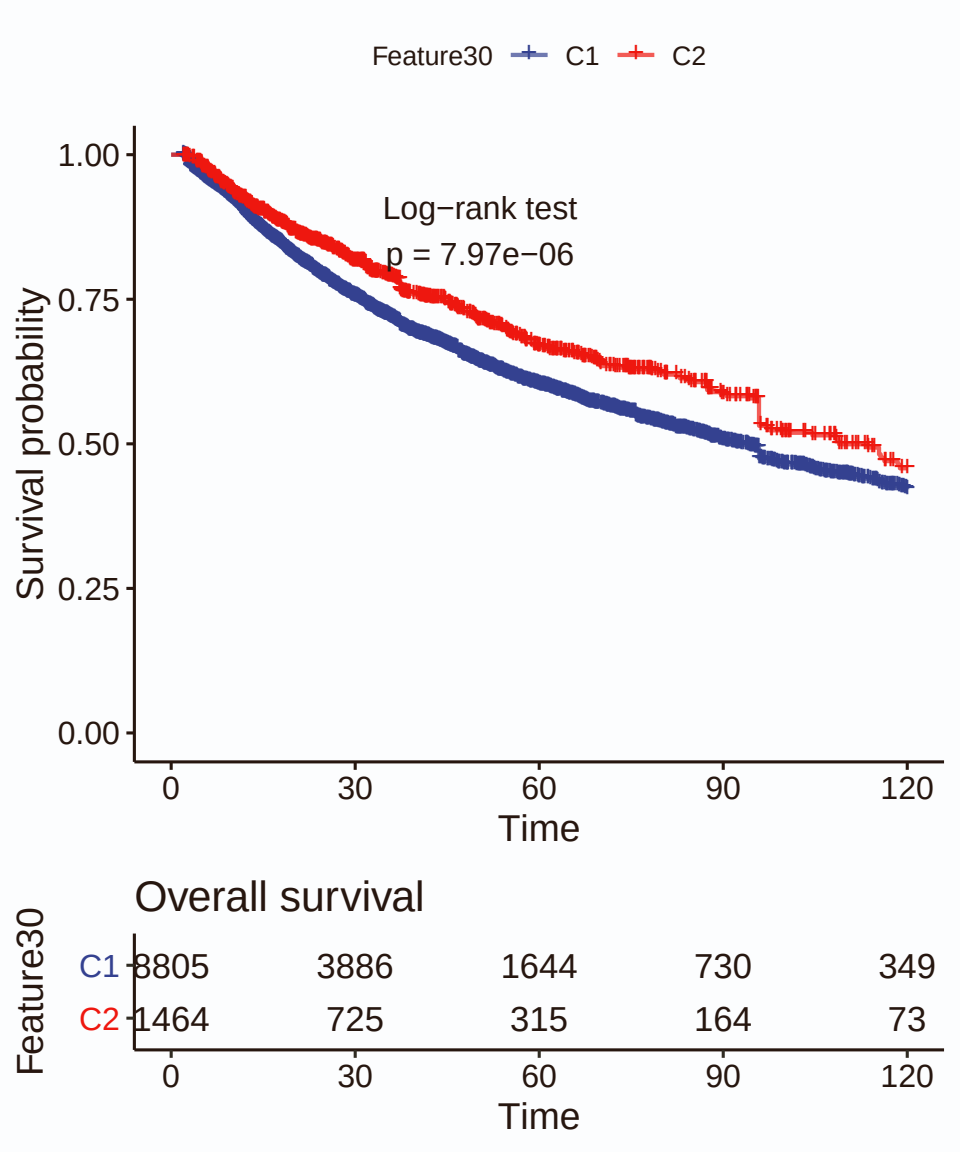

Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat
Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×

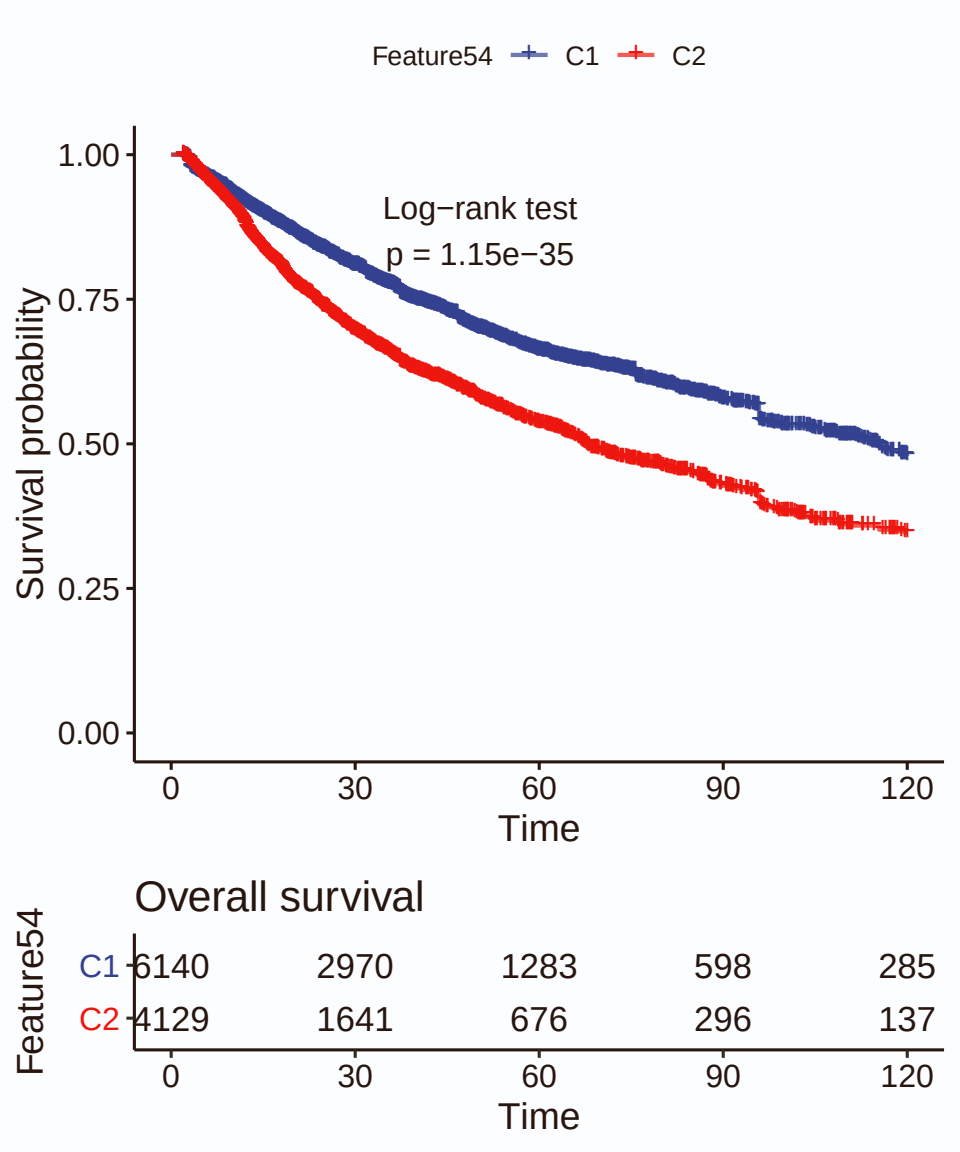

Reshape
shape⟨2⟩
Mul
B⟨48⟩
Add
B⟨48⟩
Relu
MatMul
B⟨48×64⟩
Reshape
shape⟨3⟩
×1×

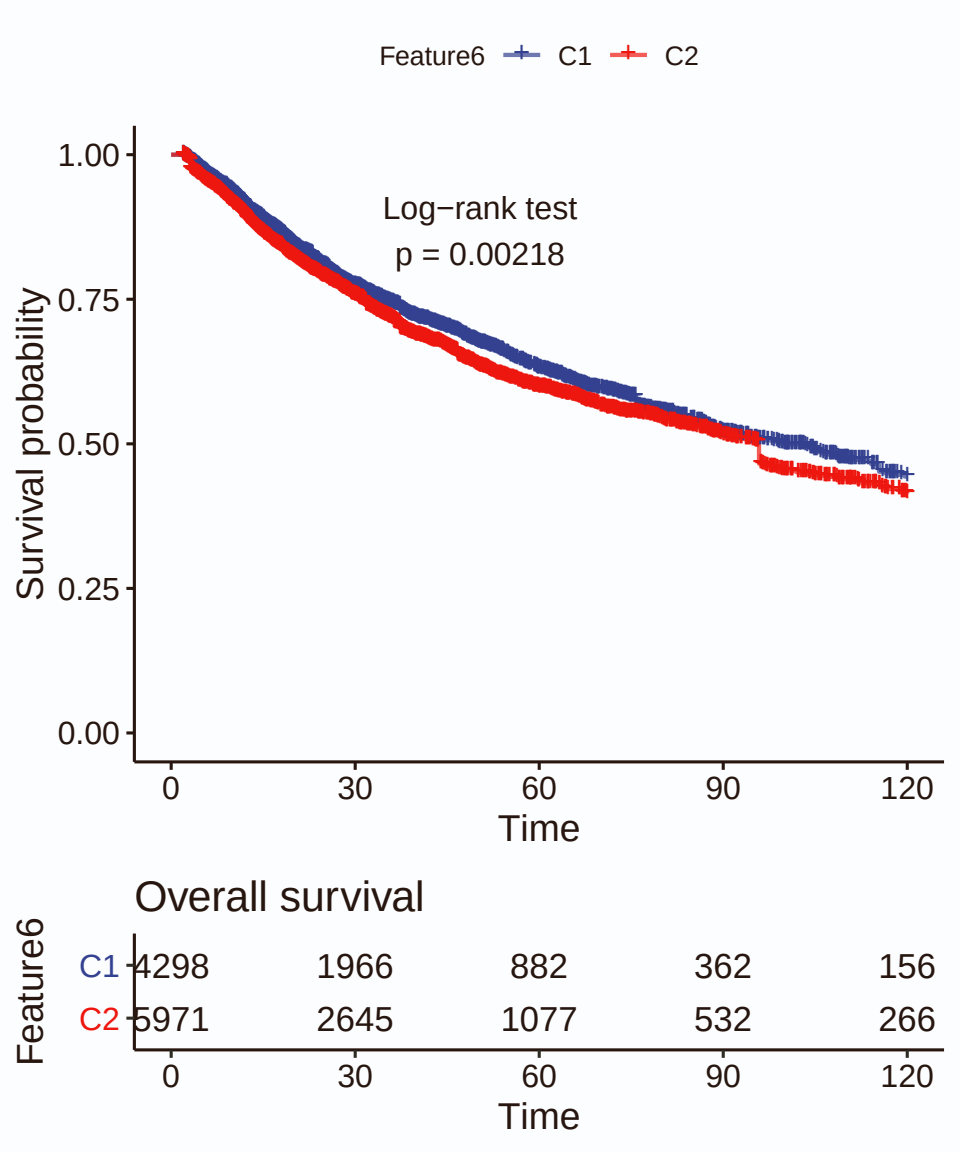

InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Relu
MatMul
B⟨64×16⟩
Concat

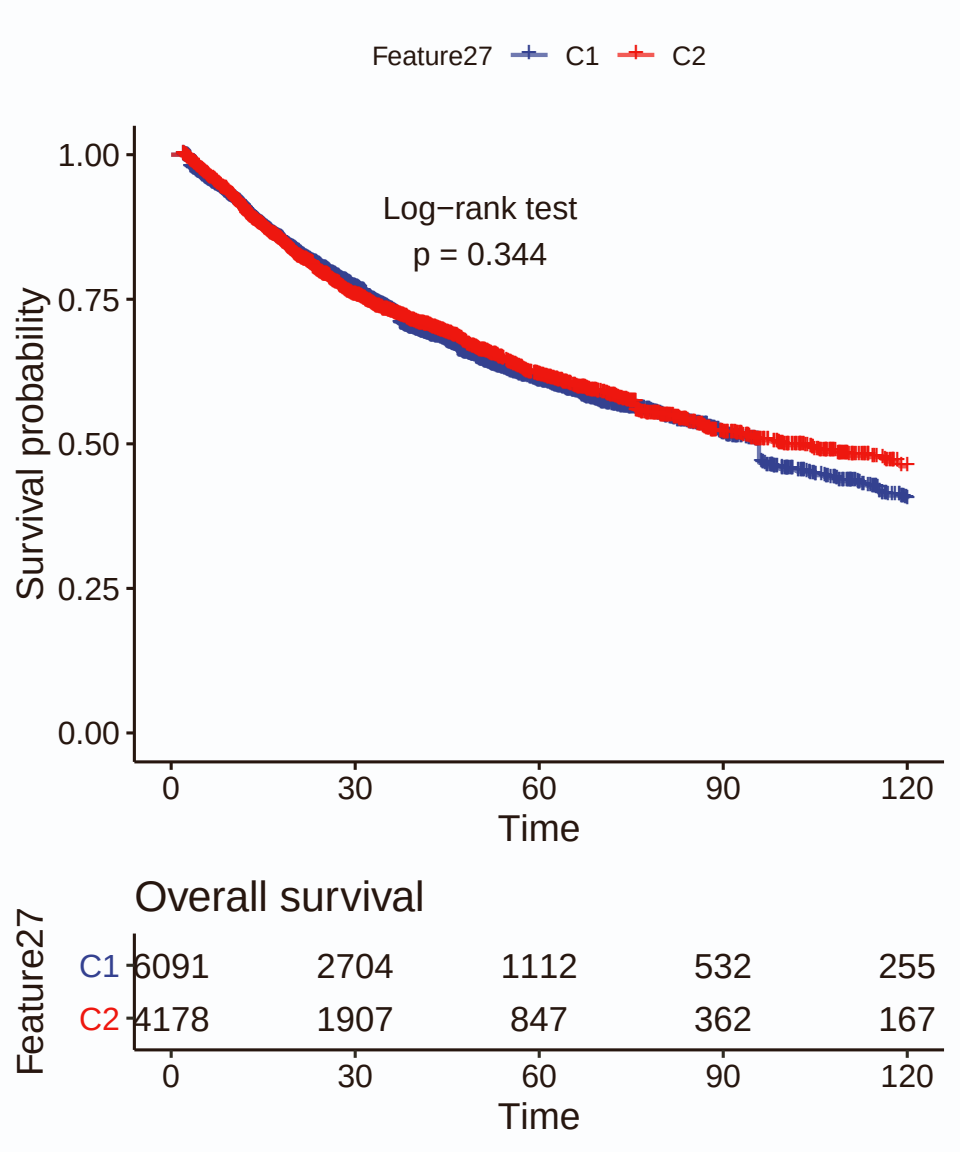

Reshape
shape⟨3⟩
×1×
InstanceNormalization
scale⟨1⟩
B⟨1⟩
×1×
Reshape
shape⟨2⟩
Mul
B⟨64⟩
Add
B⟨64⟩
Flatten
1×64
output

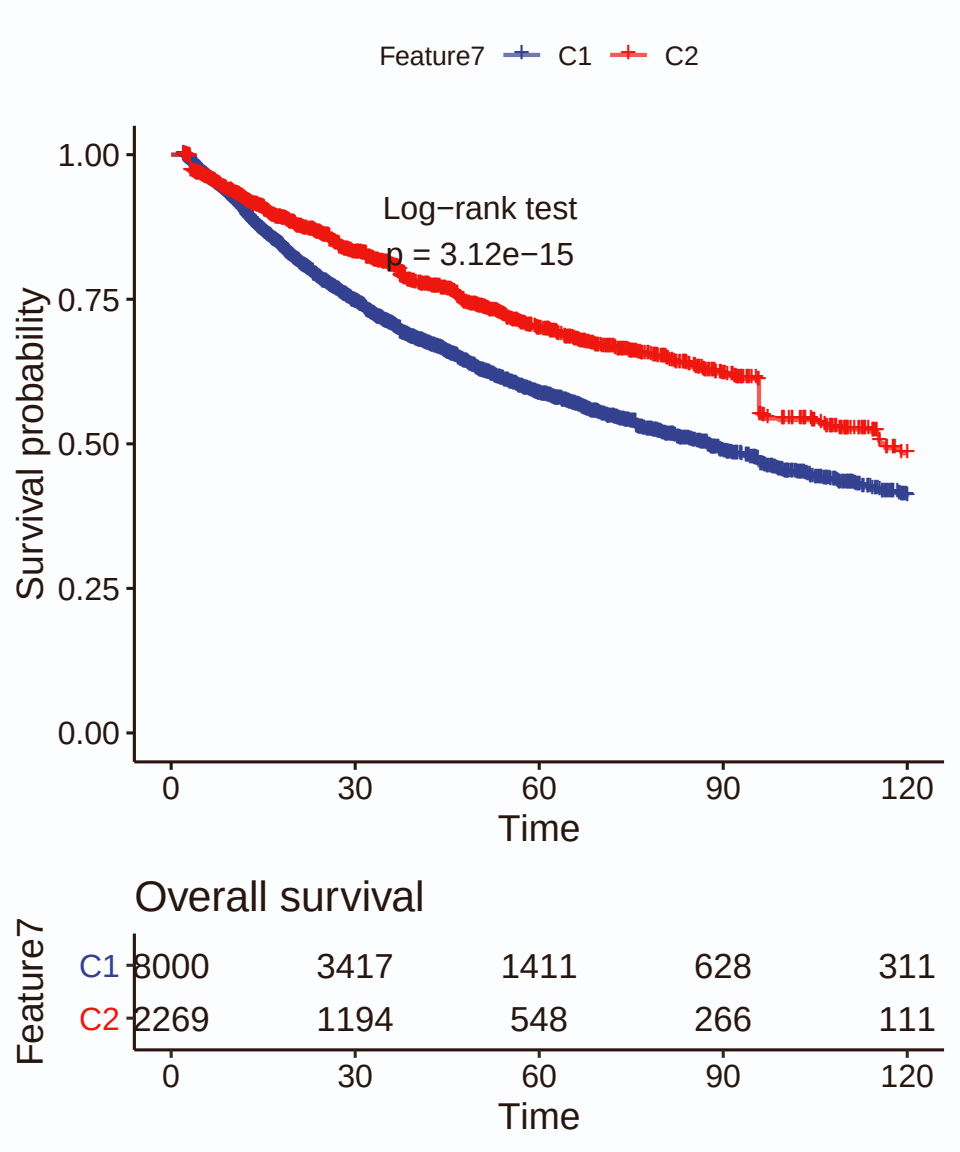
Extended Data2. The association between the latent features learned by Miscell and prognosis in the *Smart-seq* dataset, related to Figure 4.

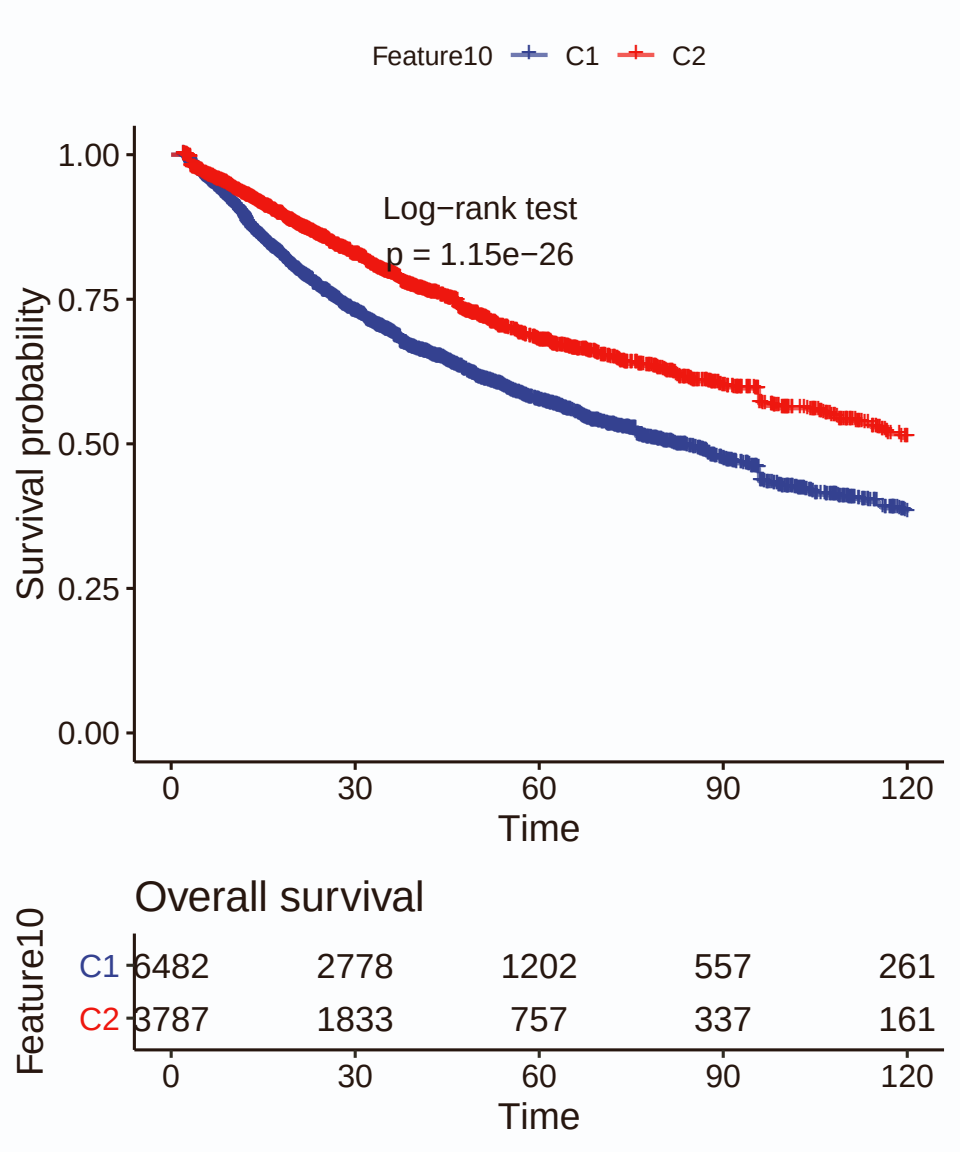

Feature30
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 7.97e−06
0
30
60
90
120
Time
Overall survival
Feature30
C1 8805 3886 1644 730 349
C2 1464 725 315 164 73
0
30
60
90
120
Time

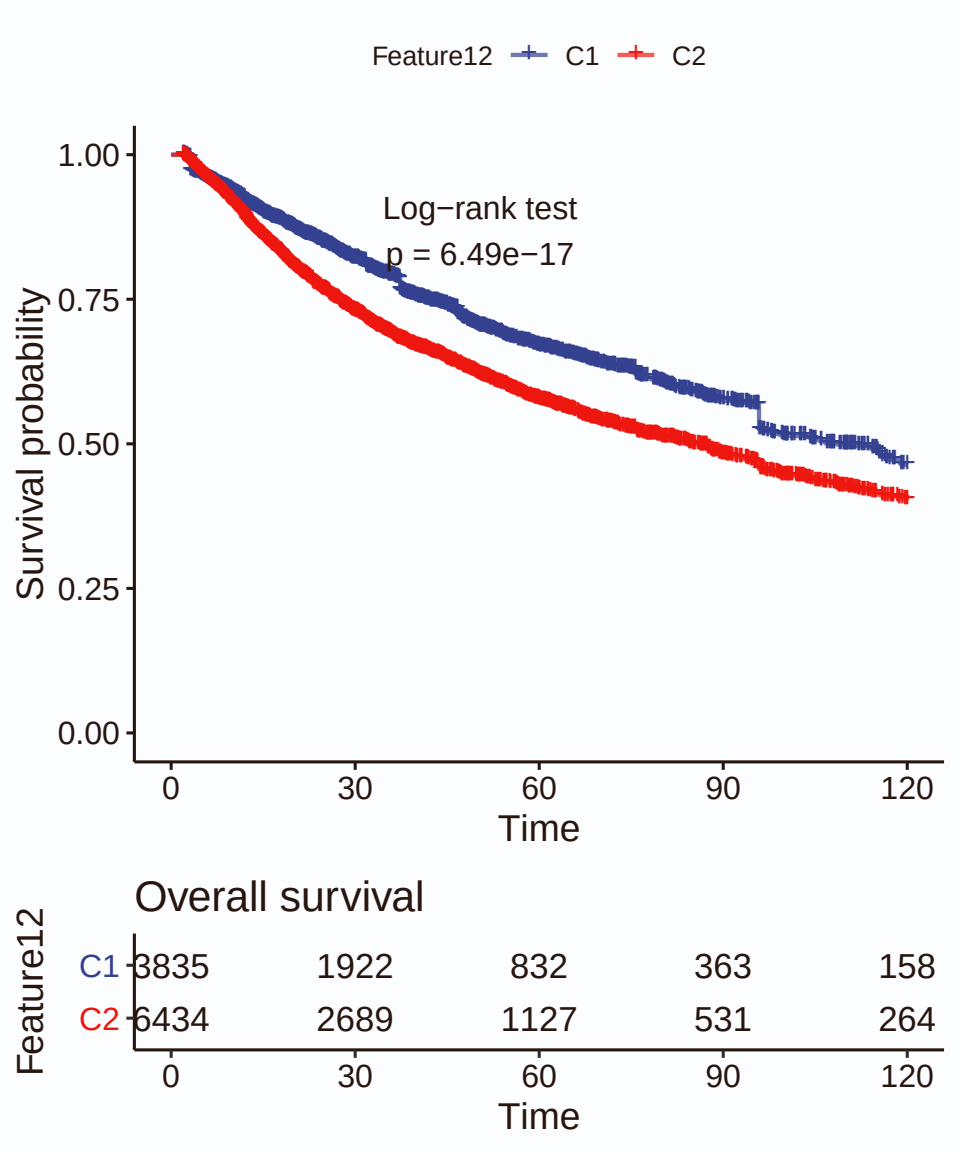

Feature54
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 1.15e−35
0
30
60
90
120
Time
Overall survival
Feature54
C1 6140 2970 1283 598 285
C2 4129 1641 676 296 137
0
30
60
90
120
Time

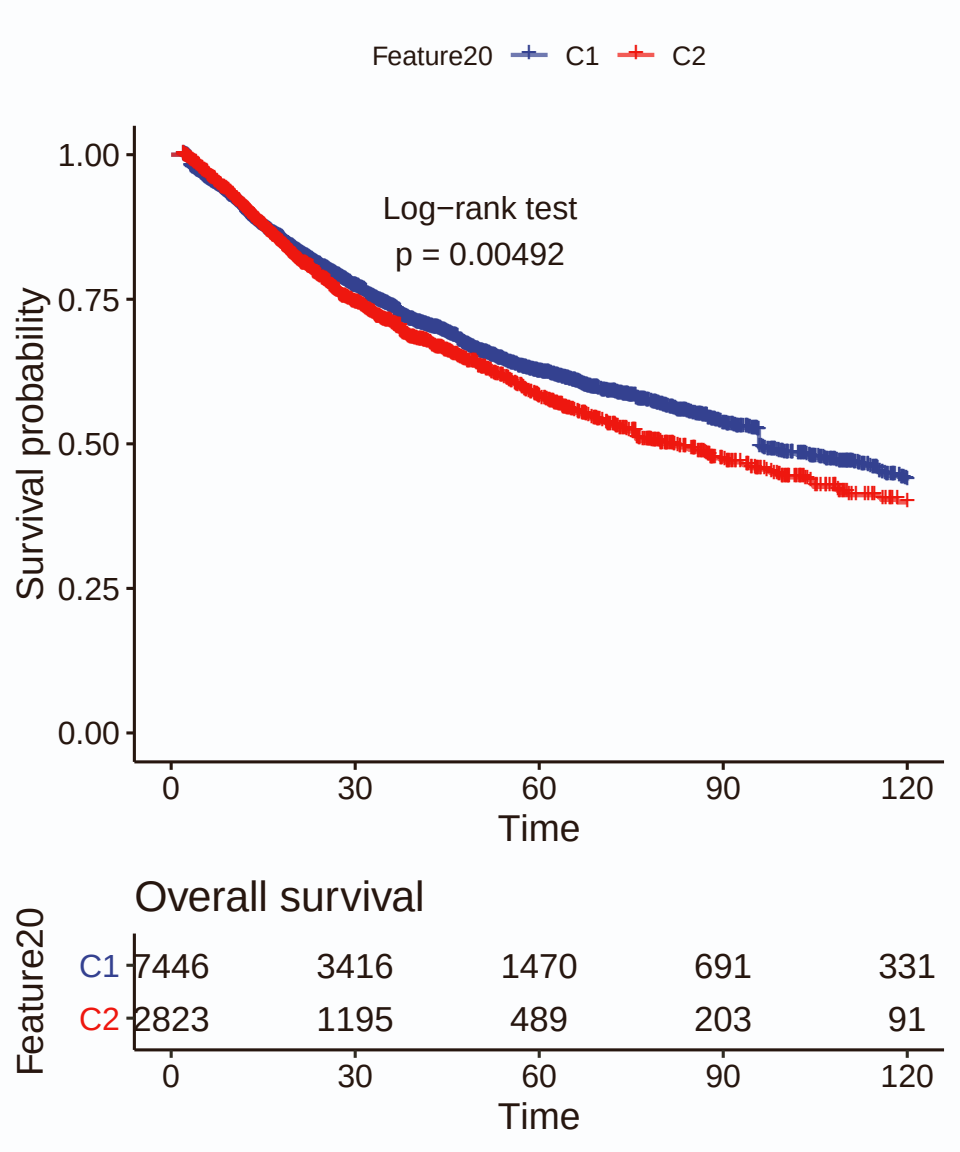

Feature6
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.00218
0
30
60
90
120
Time
Overall survival
Feature6
C1 4298 1966 882 362 156
C2 5971 2645 1077 532 266
0
30
60
90
120
Time

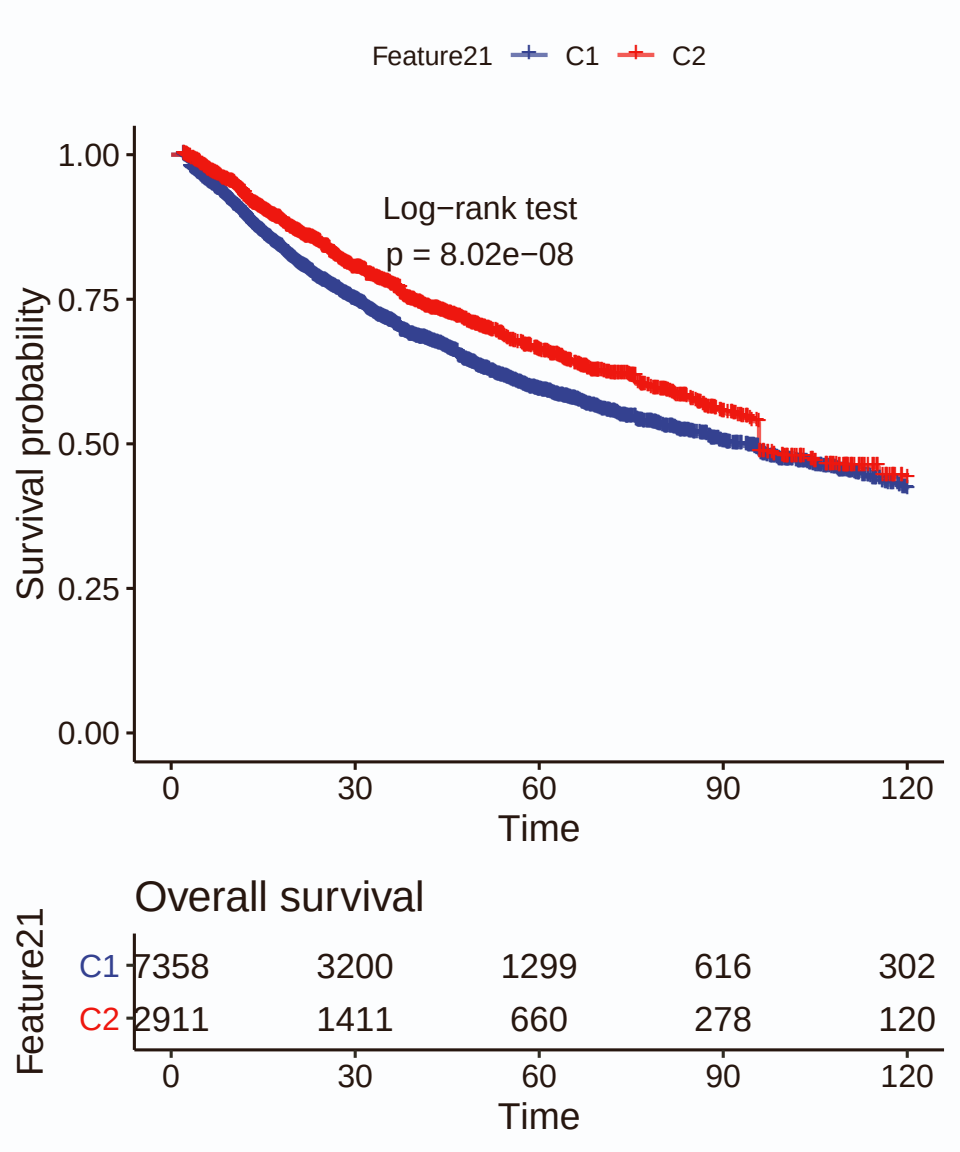

Feature27
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.344
0
30
60
90
120
Time
Overall survival
Feature27
C1 6091 2704 1112 532 255
C2 4178 1907 847 362 167
0
30
60
90
120
Time

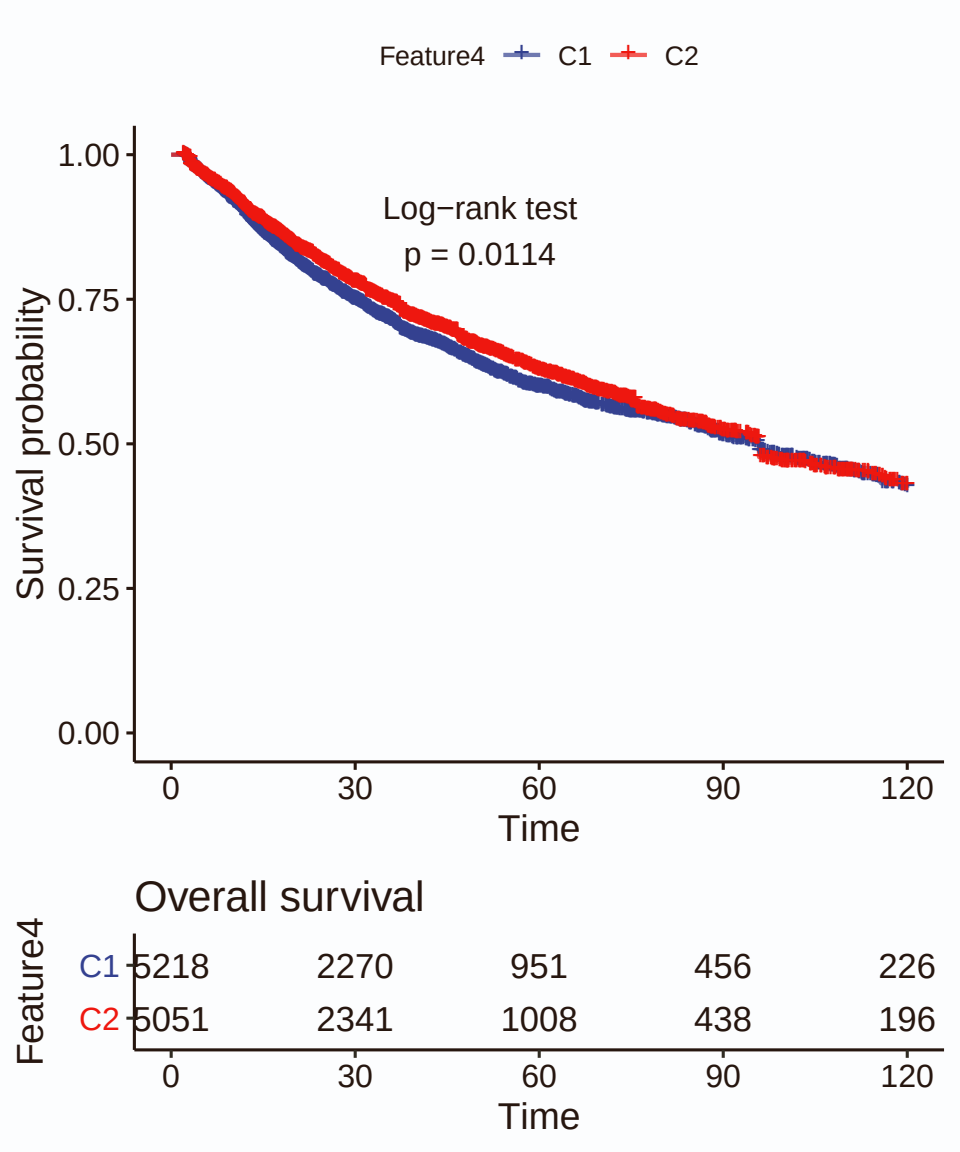

Feature7
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 3.12e−15
0
30
60
90
120
Time
Overall survival
Feature7
C1 8000 3417 1411 628 311
C2 2269 1194 548 266 111
0
30
60
90
120
Time

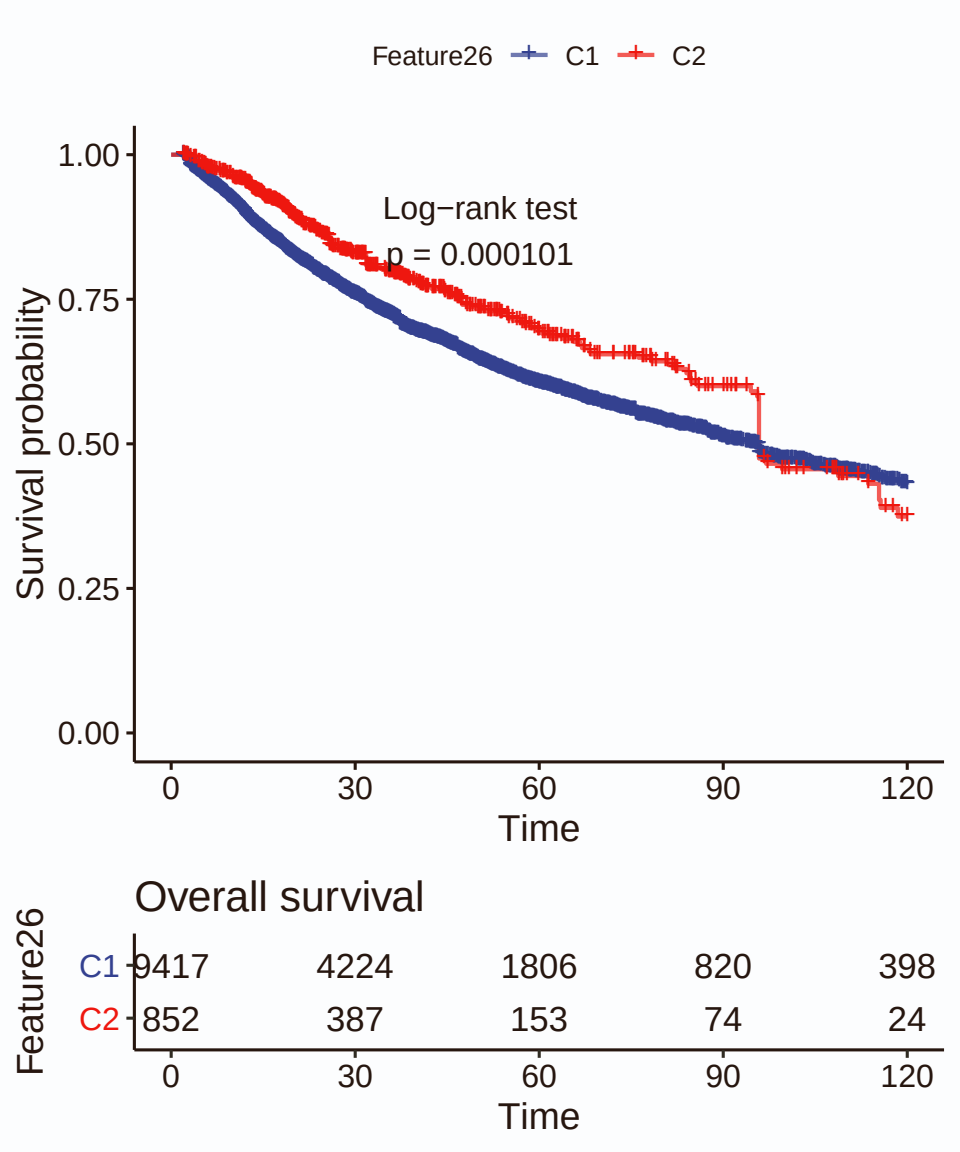

Feature10
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 1.15e−26
0
30
60
90
120
Time
Overall survival
Feature10
C1 6482 2778 1202 557 261
C2 3787 1833 757 337 161
0
30
60
90
120
Time

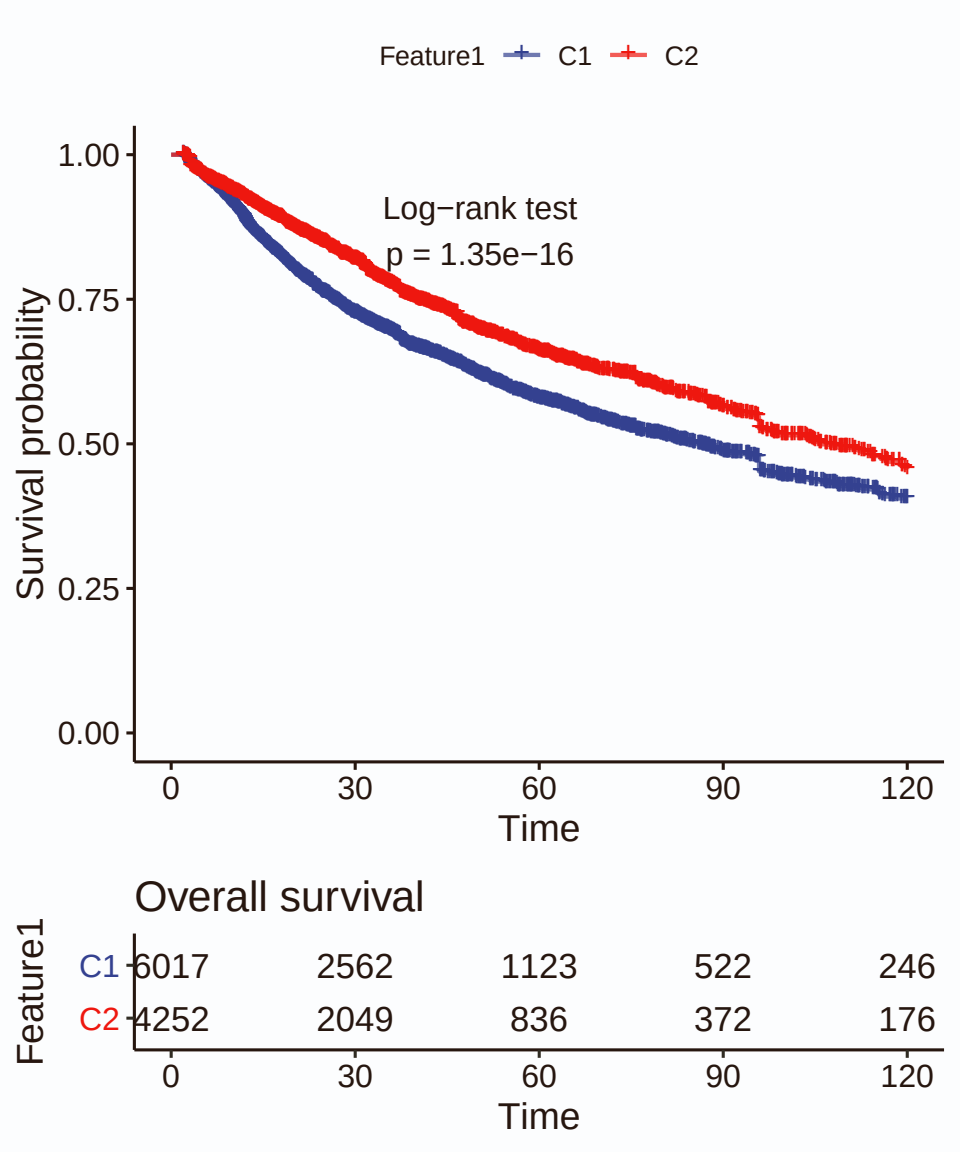

Feature12
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 6.49e−17
0
30
60
90
120
Time
Overall survival
Feature12
C1 3835 1922 832 363 158
C2 6434 2689 1127 531 264
0
30
60
90
120
Time

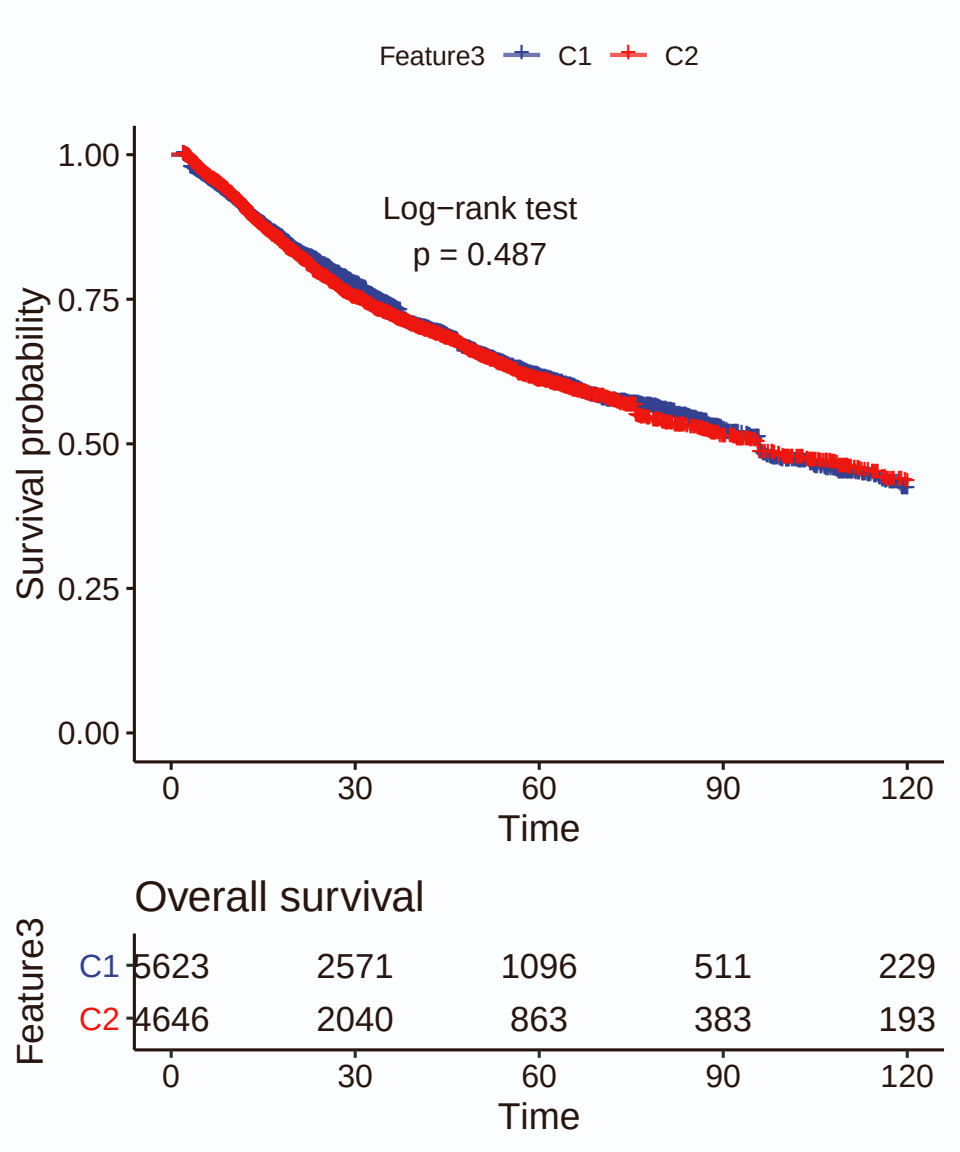

Feature20
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.00492
0
30
60
90
120
Time
Overall survival
Feature20
C1 7446 3416 1470 691 331
C2 2823 1195 489 203 91
0
30
60
90
120
Time

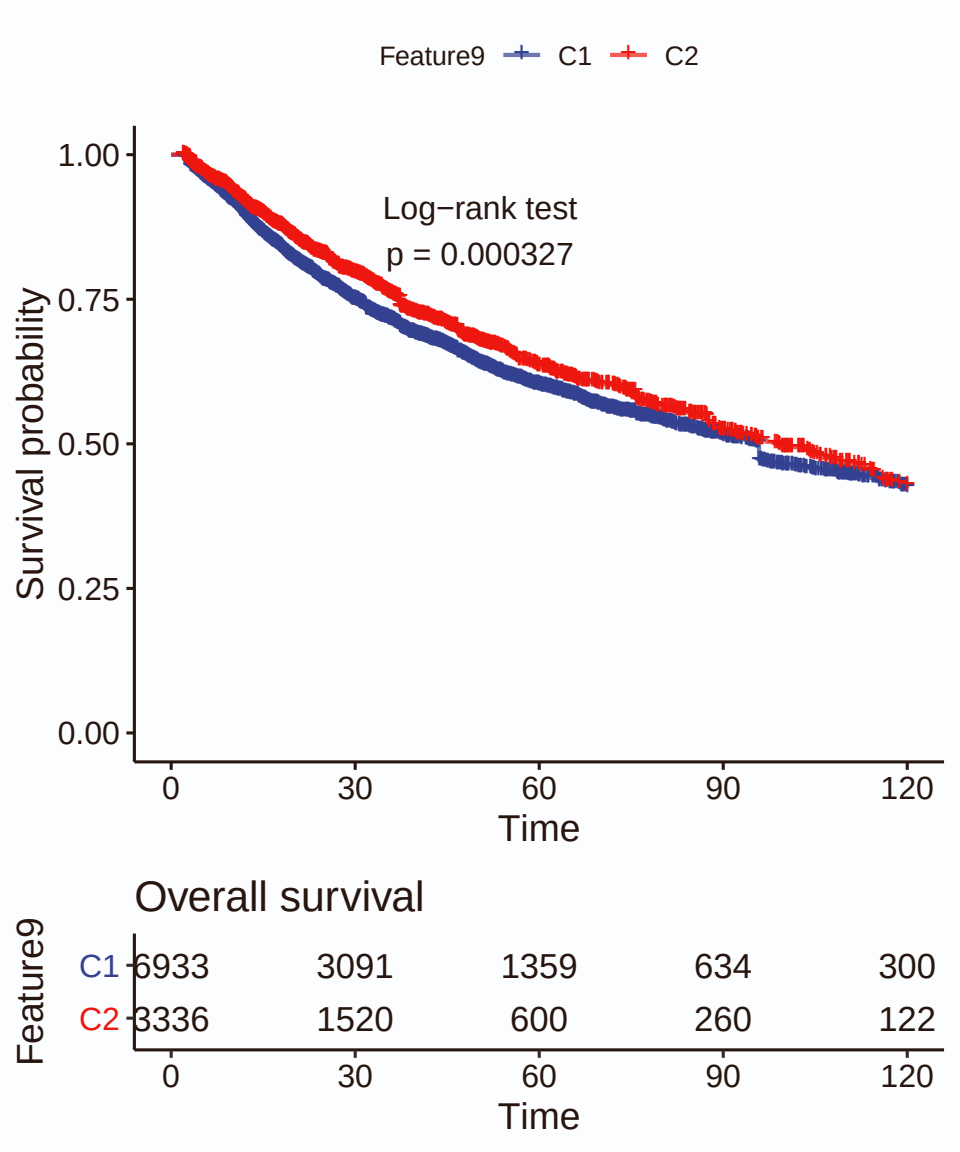

Feature21
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 8.02e−08
0
30
60
90
120
Time
Overall survival
Feature21
C1 7358 3200 1299 616 302
C2 2911 1411 660 278 120
0
30
60
90
120
Time

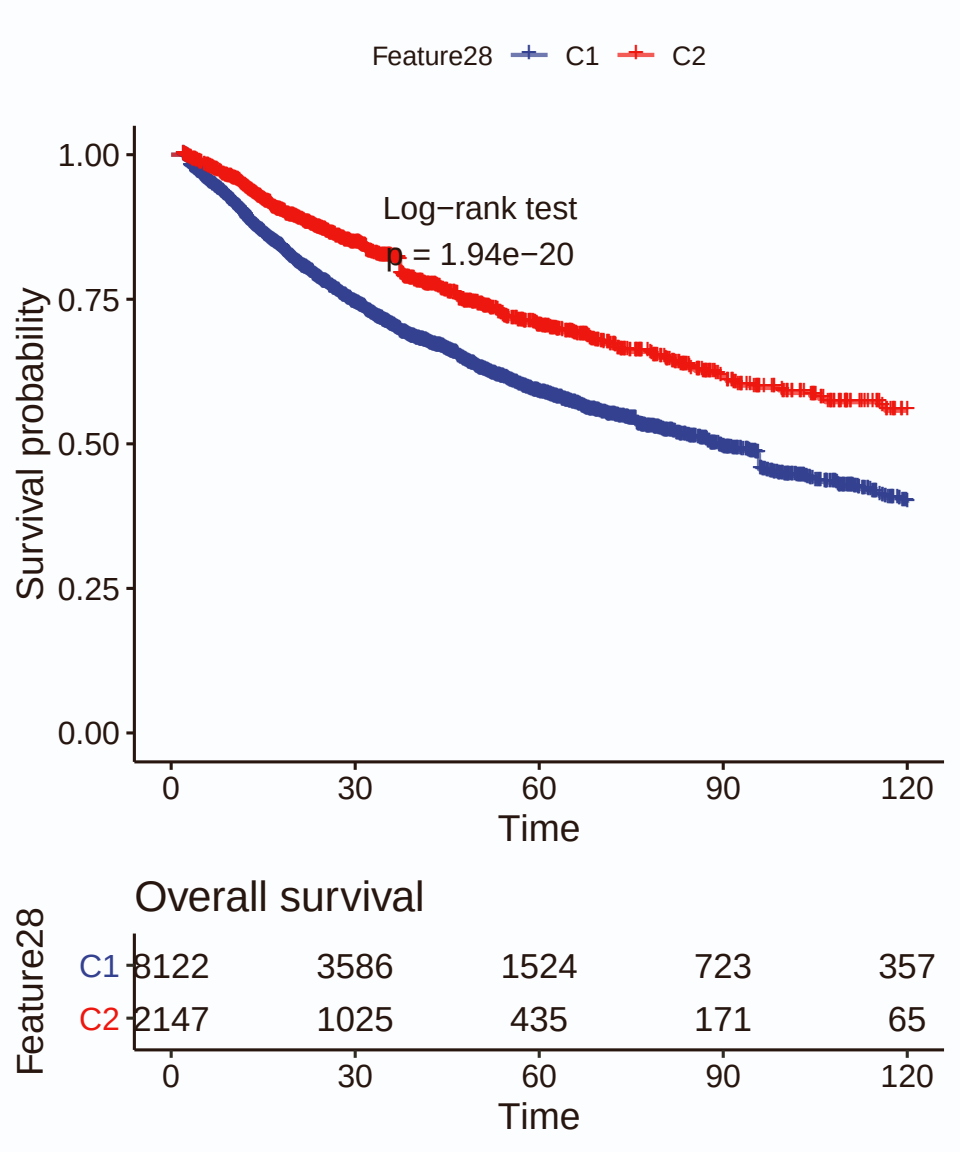

Feature4
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.0114
0
30
60
90
120
Time
Overall survival
Feature4
C1 5218 2270 951 456 226
C2 5051 2341 1008 438 196
0
30
60
90
120
Time

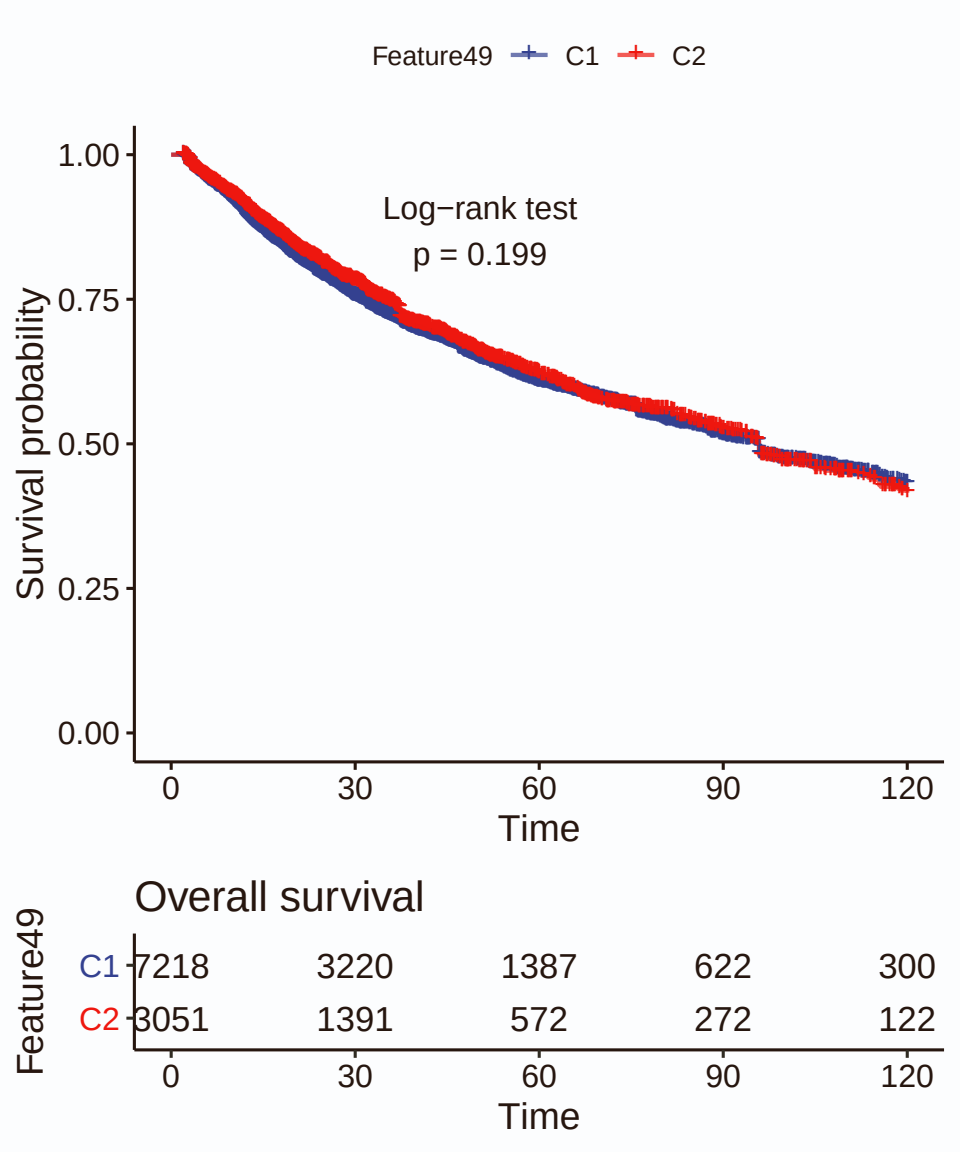

Feature26
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.000101
0
30
60
90
120
Time
Overall survival
Feature26
C1 9417 4224 1806 820 398
C2 852 387 153 74 24
0
30
60
90
120
Time

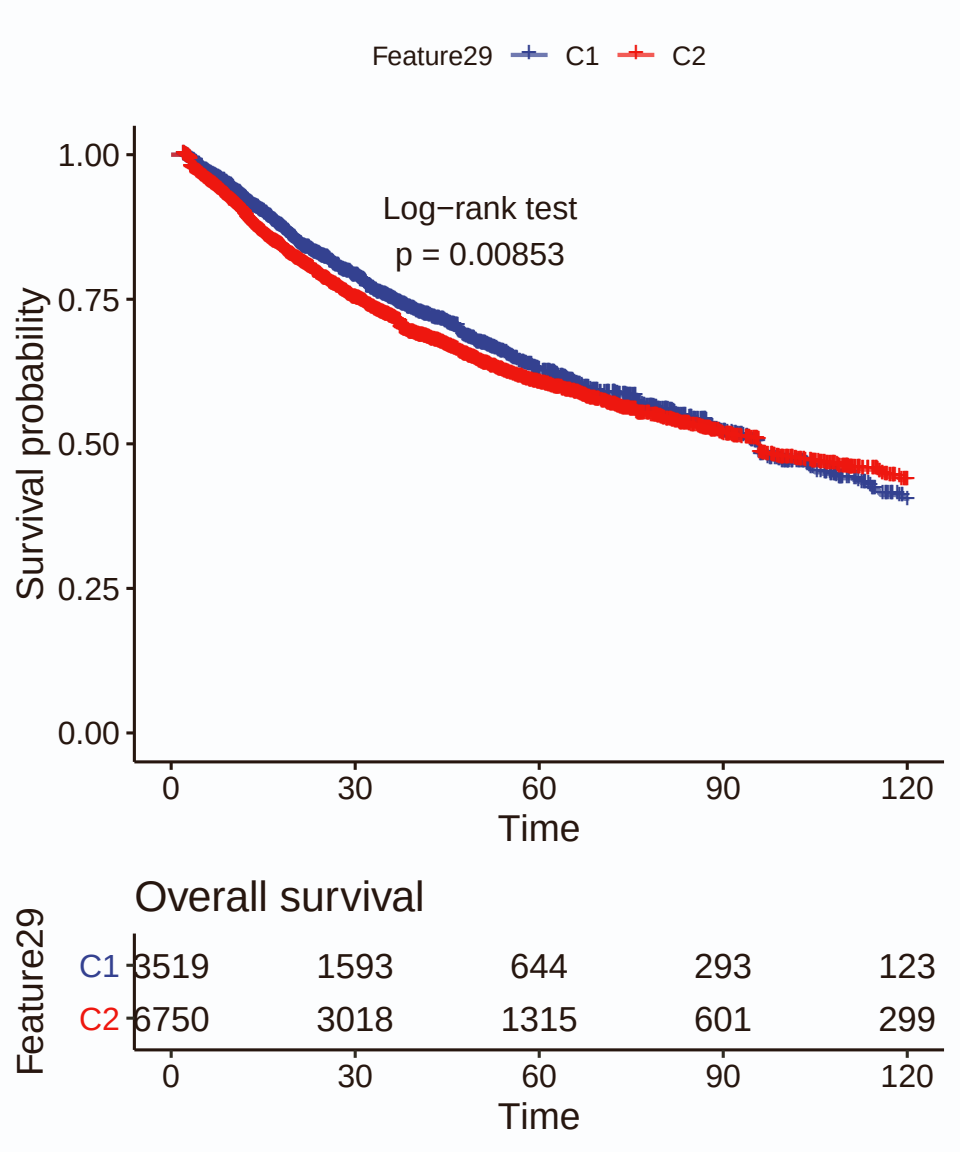

Feature1
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 1.35e−16
0
30
60
90
120
Time
Overall survival
Feature1
C1 6017 2562 1123 522 246
C2 4252 2049 836 372 176
0
30
60
90
120
Time


Feature3
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.487
0
30
60
90
120
Time
Overall survival
Feature3
C1 5623 2571 1096 511 229
C2 4646 2040 863 383 193
0
30
60
90
120
Time


Feature9
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.000327
0
30
60
90
120
Time
Overall survival
Feature9
C1 6933 3091 1359 634 300
C2 3336 1520 600 260 122
0
30
60
90
120
Time


Feature28
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 1.94e−20
0
30
60
90
120
Time
Overall survival
Feature28
C1 8122 3586 1524 723 357
C2 2147 1025 435 171 65
0
30
60
90
120
Time


Feature49
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.199
0
30
60
90
120
Time
Overall survival
Feature49
C1 7218 3220 1387 622 300
C2 3051 1391 572 272 122
0
30
60
90
120
Time


Feature29
C1
C2
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank test
p = 0.00853
0
30
60
90
120
Time
Overall survival
Feature29
C1 3519 1593 644 293 123
C2 6750 3018 1315 601 299
0
30
60
90
120
Time