

A Natural Separation of GRBs using Autoencoders on Light Curves

JE QIN CHOOI¹

¹*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA*

ABSTRACT

Gamma-ray bursts (GRBs) are traditionally separated by their duration of 90% of significant flux at $T_{90} = 2\text{s}$ into short GRBs (SGRBs) and long GRBs (LGRBs). Short GRBs have been associated to compact merger events which are sources of r -process nucleosynthesis that is of great academic interest. Compact mergers can also be cross-validated with multi-messenger astronomy through gravitational wave observations. Long GRBs are instead associated with core-collapse Type Ic supernovae. However, recent studies suggest that there is a population of long GRBs that does not have a corresponding visual supernovae. Furthermore, some are confirmed to be compact merger events. This suggests for a new classification of GRBs that is more sophisticated than inspecting T_{90} . Recent work saw machine learning being applied to extract local structures from GRB light curves through t-SNE. However, t-SNE is designed for data visualization of high dimensional data sets and should not be used for clustering. Here we applied a proper dimensionality reduction technique using a type of neural network known as autoencoders. We successfully reduced high-resolution minute-long light curves into 2, 3, and 5 dimensions. A visual separation between SGRBs and LGRBs is observed at $T_{90} = 2\text{s}$, supporting previous literature of a natural separation. However, a clustering technique known as K-means failed to objectively extract this cluster, mainly due to the imbalanced sample with a much higher number of LGRBs than SGRBs, and the Euclidean distance metric and centre-focused approach used with K-means.

1. INTRODUCTION

Gamma-ray bursts (GRBs) have at least two astrophysical origins: core-collapse and compact mergers (Kouveliotou et al. 1993). Compact mergers are of interest as they are the only confirmed sources of r -process nucleosynthesis responsible for the generation of heavy elements (Tanvir et al. 2017). Previous work partitions GRBs by the duration containing 90% of the statistically significant flux at $T_{90} \approx 2\text{s}$ into short and long GRBs (Paciesas et al. 1999). Long GRBs are associated with core-collapse (Stanek et al. 2003) while short GRBs are associated with compact mergers (Berger et al. 2013). However, the separation is not clear, with some overlap across classes. Some long GRBs have no observed core-collapse supernovae (Valle et al. 2006) and there is a long GRB with a compact merger progenitor (Rastinejad et al. 2022). As compact merger events are of great interest, it is desirable if they can be detected from among the core-collapse long GRBs. There are suggestions of a third class of GRBs known as extended emission GRBs (EE GRBs) (Norris & Bonnell 2006) which has soft X-ray and gamma-ray emission after the main peak. The duration of this proposed class overlaps with the long GRBs while their spectral properties are similar to short

GRBs. There has not been a successful attempt at a complete catalogue of this class.

Recent works attempted to cluster the GRBs by their light curves using the “t-distributed stochastic neighbor embedding” (t-SNE) (Jespersen et al. 2020). The light curve of each GRB is reduced into a 2-dimensional data point, and a visual separation is observed between the two clusters of these data points. Furthermore, the two clusters correspond to short and long duration of significant flux at approximately $T_{90} = 2\text{s}$ as previously used by the community. However, t-SNE is designed for data visualization and a different choice of its perplexity parameter results in a different structure. The latent space of the dimensionally-reduced light curves is also hard to interpret, as the space merely provides a space for local structures to form and not describing any particular physical features.

In this paper, a different dimensionality reduction technique known as autoencoders is applied to the GRB light curves. The latent space of the autoencoders reflects the quantitative variation of physical features of the light curves. Furthermore, K-means clustering technique is applied to the reduced data set, where a separation of GRBs into long and short duration is found, con-

sistent with previous work. Extended Emission (TALK ABOUT RESULTS)

This project uses the GRB dataset by (Garcia-Cifuentes et al. 2023) and employs autoencoders for dimensionality reduction and K-Means for clustering. We find that our results largely agree with the t-SNE approach, but we have a predicted a larger size for the smaller cluster that is associated with shorter T_{90} . Several outlier GRBs are identified and compared to known GRBs of interest, including one with $T_{90} = 31$ s that is classified as part of the short GRBs.

2. METHODS

The light curves of 1,589 GRBs are obtained from the SWIFT Burst Alert Telescope (BAT) through the ClassiPyGRB package (Garcia-Cifuentes et al. 2023). The earliest GRB was on 17th Decemeber 2004 and the most recent GRB was on 15th December 2023. Each light curve is available in 5 bands: 15-25 keV, 25-50 keV, 50-100 keV, 100-350 keV and 15-350 keV. Bins available are 2ms, 8ms, 16ms, 64ms, 256ms, 1s and 10s. Since short GRBs can be less than 2 seconds long, and long GRBs can last more than a minute, the 64 ms bin are chosen for this analysis to provide enough data points for each light curve within reasonable computation.

The preprocessing procedures on the ClassiPyGRB package are followed. First, we apply the nonparametric noise reduction technique know as Fully Adaptive Bayesian Algorithm for Data Analysis (FABADA) (Sanchez-Alarcon & Sequeiros 2023). Implementation is provided through the ClassiPyGRB package.

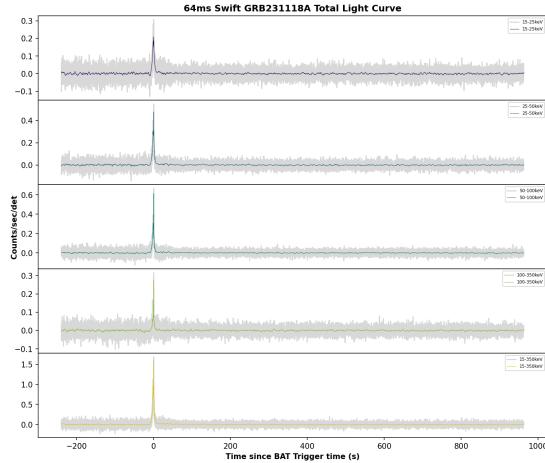


Figure 1. The original Swift/BAT data is very noisy, as shown in gray. The colored lines show the results after applying FABADA to the light curves.

The resulting noise-reduced light curve after applying FABADA can be seen through an example of

GRB231118A in Figure 1. Next, we limit the light curves to the duration of the significant flux of the GRB, denoted as T_{100} . An example of this step is in Figure 2. The light curves are then normalized by divid-

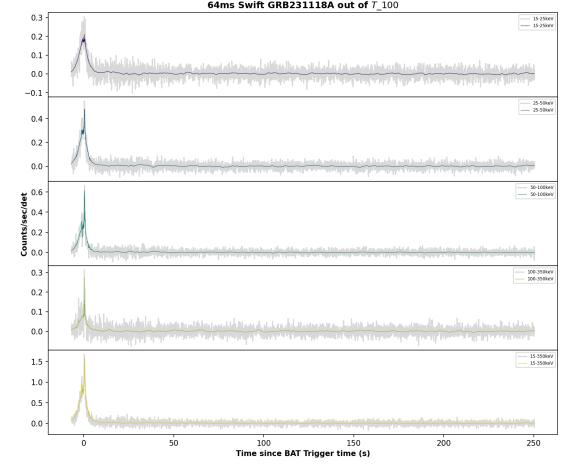


Figure 2. The light curve is limited to T_{100} which is the duration of the significant flux of the GRB. Note that the main peak is near the start of the limited light curve.

ing over the time-integrated total flux obtained through the trapezoid rule. Next, the light curves are all zero-padded so that every light curve has the same dimension. The final dimension is 19,200 bins which corresponds to 1,228.8 seconds. Discrete Fourier Transform (DFT) is then applied to reduce the dataset into half of its dimension, and it will be used for dimensionality reduction along with the plain dataset to compare the effect of DFT. The data below the Nyquist frequency of the light curve is kept, resulting in a dimension of 48,000 data points per light curve. An example of a DFT is in Figure 3.

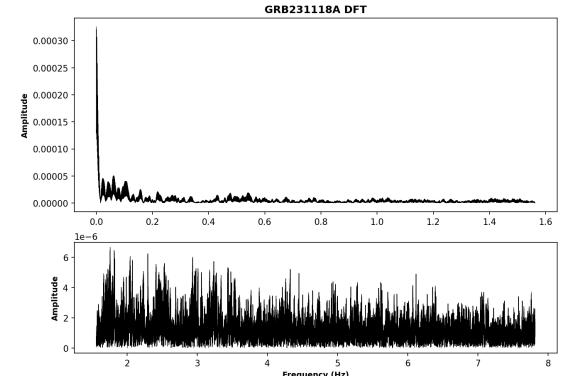


Figure 3. Frequency spectrum of GRB231118A upon applying DFT

Some light curves have missing data, resulting in an effective dataset of 1,513 GRBs after preprocessing.

A set of autoencoders is trained for dimensionality reduction on both the DFT and plain light curves. For the DFT data, since we have only 1,513 GRBs, there is an upper bound on the number of hidden variables even though we have a rich set of features. Features are standardized to have mean of 0 and standard deviation of 1 before being used for training and testing. Autoencoders using choices of latent size, hidden size and input size are trained and tested with their losses compared. Each autoencoder has two hidden layers with the latent layer in between. For a given input size D , we select the D highest amplitude frequencies from the DFT. All models are trained for 200 epochs with an Adam optimizer and a learning rate of 10^{-3} . Layers are initialized with a Gaussian distribution with mean of 0 and standard deviation of 0.01. Layers are connected with leaky ReLU.

A list of compact merger GRBs is compiled at Table 1 to be compared in the dimensionally-reduced dataset for any clustering.

GRB	Year of Study	Source
GRB130603B	2014	(Postigo et al. 2014)
GRB060614	2015	(Jin et al. 2015)
GRB080503	2015	(Gao et al. 2015)
GRB050709	2016	(Jin et al. 2016)
GRB050724	2017	(Gao et al. 2017)
GRB061006	2017	(Gao et al. 2017)
GRB070714B	2017	(Gao et al. 2017)
GRB160821B	2017	(Kasliwal et al. 2017)
GRB111005A	2018	(Michałowski et al. 2018)
GRB150101B	2018	(Troja et al. 2018)
GRB070809	2020	(Jin et al. 2020)
GRB211211A	2022	(Mei et al. 2022)
GRB211227A	2022	(Lü et al. 2022)
GRB191019A	2023	(Levan et al. 2023)
GRB230307A	2024	(Levan et al. 2024)

Table 1. List of compact merger GRBs

K-means will then be applied as an attempt to systematically cluster the GRBs.

3. RESULTS

3.1. Dimensionality Reduction using Autoencoders

First we present the training and testing losses for each choice of latent size at epoch 200. The minimum test loss is bolded for each latent size choice.

Hidden Size	Input Size	Training Loss	Test Loss
100	1000	0.087869	0.079932
100	2000	0.065483	0.077554
200	2000	0.064954	0.075277
200	4000	0.070439	0.072166
200	8000	0.066869	0.074685

Table 2. Training and test loss for latent size of 2

Hidden Size	Input Size	Training Loss	Test Loss
100	1000	0.063662	0.065302
100	2000	0.063720	0.062473
200	2000	0.056500	0.061706
200	8000	0.050907	0.058759

Table 3. Training and test loss for latent size of 3

Hidden Size	Input Size	Training Loss	Test Loss
100	1000	0.044205	0.050466
200	2000	0.041912	0.045728
200	8000	0.040719	0.042493

Table 4. Training and test loss for latent size of 5

For each latent size, the combination of parameters that achieves minimum test loss is selected for further analysis. Their corresponding train/test loss curves are at Figure 4, 5, 6.

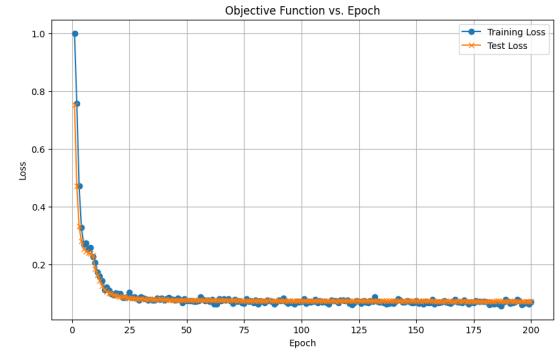


Figure 4. Training and test loss for latent size 2, hidden size 200, input size 4000

These three autoencoders are then used to reduce the entire dataset into their respective latent space. Here we present their latent space representation color mapped by $\log_{10} T_{90}$. The conventional separation around the $T_{90} = 2\text{s}$ is displayed, categorising into short GRBs (SGRB) and long GRBs (LGRB). In each latent space, there is a clear separation that coincides with the conventional separation of $T_{90} = 2\text{s}$.

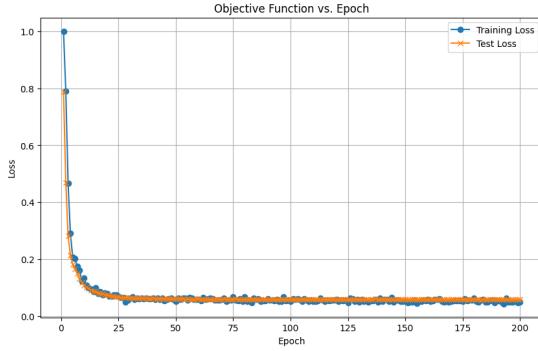


Figure 5. Training and test loss for latent size 3, hidden size 200, input size 8000

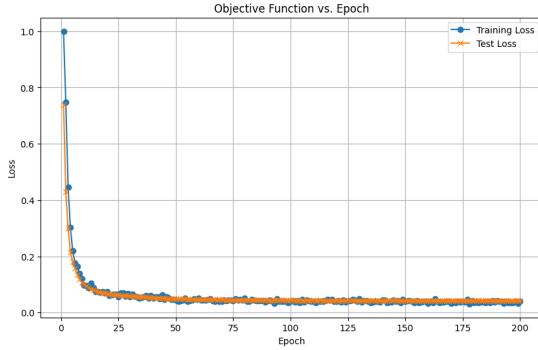


Figure 6. Training and test loss for latent size 5, hidden size 200, input size 8000

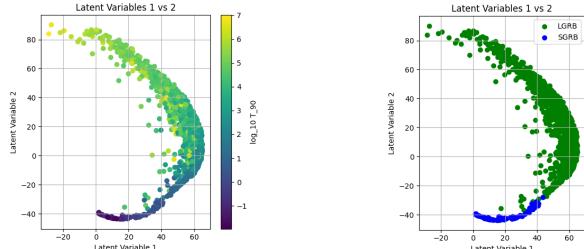


Figure 7. Latent space representation of GRB light curves using a latent size of 2 color mapped by $\log_{10} T_{90}$.

Figure 8. Latent space representation of GRB light curves using a latent size of 2 color coded by SGRB vs LGRB.

The results for the autoencoder with latent size 3 reflects a higher dimensional analogue of the autoencoder with latent size 2. In each plot, there is a smooth continuum from short to long duration that spans out spatially. The conventional separation of $T_{90} = 2$ s is also visible and well-motivated through Figure 10.

The 3D representation at Figure 11 shows clearly that there is a clearly distinct cluster of GRBs with short duration. The conventional separation correctly identifies this region through Figure 10.

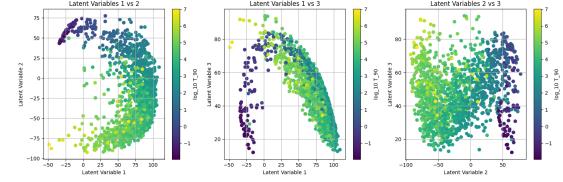


Figure 9. Latent space representation of GRB light curves using a latent size of 3 color mapped by $\log_{10} T_{90}$.

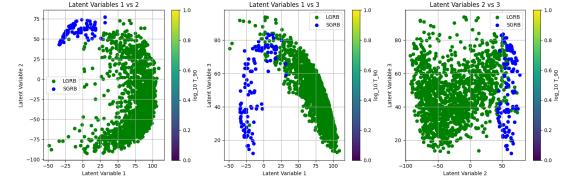


Figure 10. Latent space representation of GRB light curves using a latent size of 3 color coded by SGRB vs LGRB.

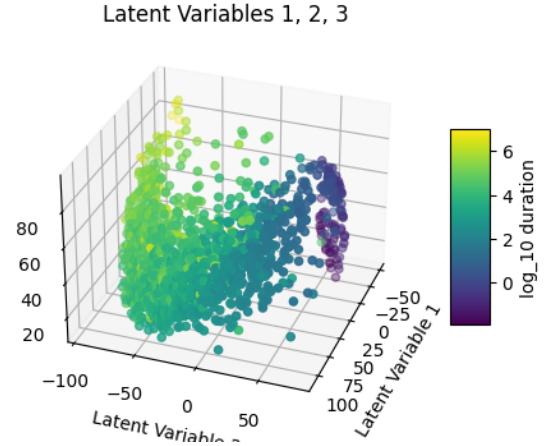


Figure 11. 3D latent space representation of GRB light curves using a latent size of 3 color mapped by $\log_{10} T_{90}$.

For latent size of 5, the graph for all pairs of latent nodes are plotted. Structures that are smooth both spatially and in T_{90} are observed in most pairs. This suggests that the relationship between latent space and T_{90} is robust across different parameters and is an important feature when reconstructing GRBs. This shouldn't be surprising as the network is rewarded when it correctly predicts a flat line of zeroes when the duration is complete.

The 5-dimensional dataset can be visualized in 3D by picking three latent variables for visualization. One such triple is visualized at Figure 14. Three clear trends can be seen in this representation. The small “appendix” has a cluster of short GRBs, extending laterally to a ladder that bridges up towards a horizontal branch of longer GRBs.

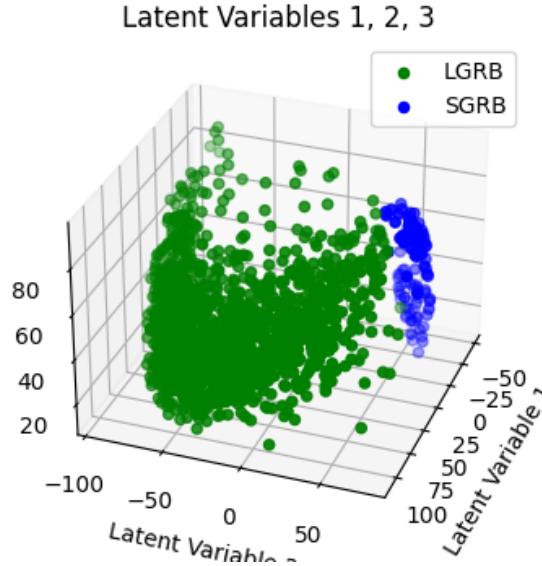


Figure 12. 3D latent space representation of GRB light curves using a latent size of 3 color coded by SGRB vs LGRB.

The conventional separation is again well-defined in the 3D representation of latent size 5 at Figure 16.

3.2. K-means Clustering

First, K-means with $K = 2$ is applied to the dimensionally reduced datasets as motivated by the SGRB/LGRB conventional separation at $T_{90} = 2$ s. Comparing Figure 8 to Figure 17, the K-means separation did not coincide with the 2s mark. This is reasonable since K-means attempts to fit two clusters through Euclidean distances. As there are many more LGRBs than SGRBs, the cluster center that is supposed to fit SGRBs is actually closer to some subset of LGRBs compared to the actual cluster center of LGRBs. It is likely that linkage based clustering methods like DB-scan and Hierarchical Agglomerative Clustering can provide better performance considering the current structure of the GRBs in the latent space.

Inspecting the duration of the GRBs of each cluster at Figure 18, there are two distributions characterized by different means of $\log_{10} T_{90}$, but it did not coincide with the $T_{90} = 2$ s mark.

K-means with $K = 2$ is also applied to the latent space of size 3. Again, no natural separation is observed at Figure 19. The K-mean separation meets abruptly between a large mass of GRBs, and did not coincide with $T_{90} = 2$ s. No clustering of the compact mergers is observed.

The duration histogram using latent size 3 at Figure 20 is similar to that of latent size 2 at Figure 18 with two distributions characterized by two different means,

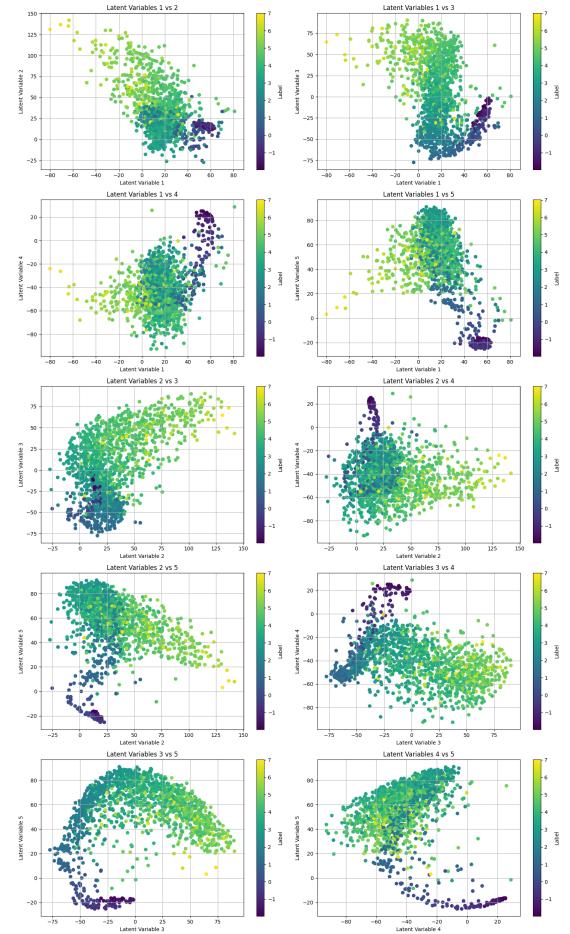


Figure 13. Latent space representation of GRB light curves using a latent size of 5 color mapped by $\log_{10} T_{90}$.

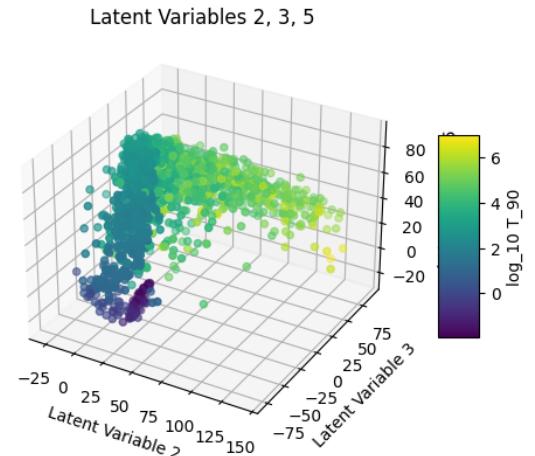


Figure 14. 3D latent space representation of GRB light curves using a latent size of 5 color mapped by $\log_{10} T_{90}$.

with a single distribution dominating across the $T_{90} = 2$ s separation.

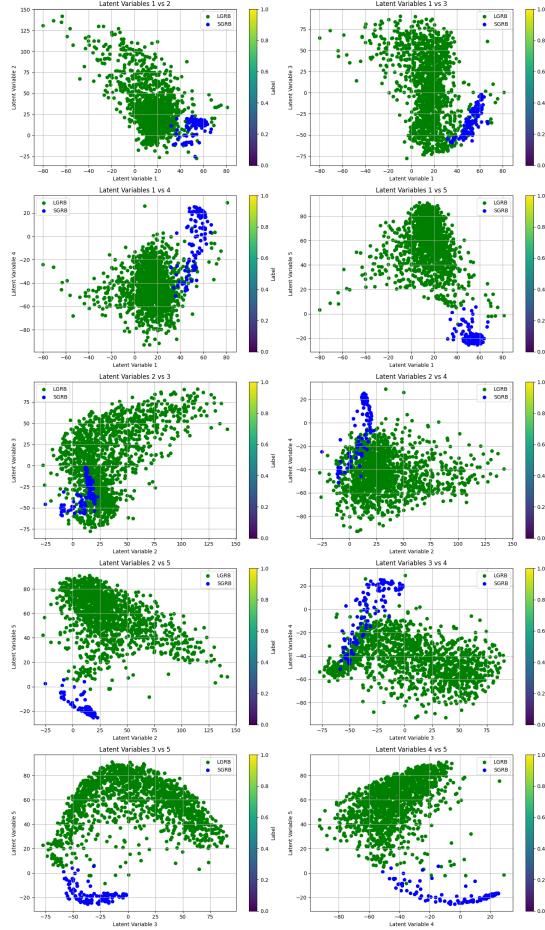


Figure 15. Latent space representation of GRB light curves using a latent size of 5 color coded by SGRB vs LGRB.

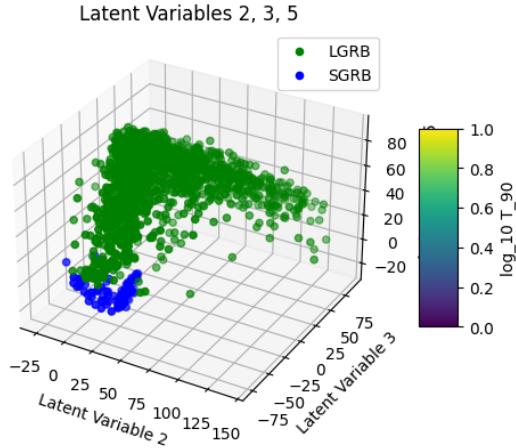


Figure 16. 3D latent space representation of GRB light curves using a latent size of 5 color coded by SGRB vs LGRB.

Similar results are also obtained for latent size 5 through Figure 21 and Figure 22.

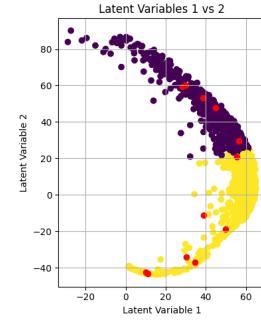


Figure 17. Color coded clusters assigned by K-means with $K = 2$. Red dots are compact mergers.

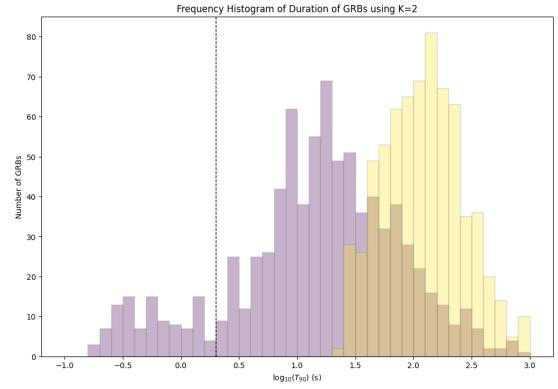


Figure 18. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 2$ on latent size 2. Colors correspond to colors previously assigned using K-means. The dotted line corresponds to $T_{90} = 2$ s.

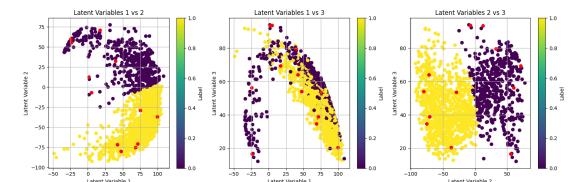


Figure 19. Color coded clusters assigned by K-means with $K = 2$. Red dots are compact mergers.

To inspect whether there is a natural number of clusters of GRBs, the inertia of the K-means clustering is plotted for each latent space through Figure 23, 24, 25. No knee is observed where the gradient of inertia changes discontinuously. It is concluded that there is no natural number of clusters under the K-means paradigm using the Euclidean metric. However, there can still be a natural cluster of GRBs under a different measure of distance between clusters.

The clusters using $K = 3$ is plotted for all latent spaces through Figures 26, 28, 30. Their corresponding duration histogram is shown through Figures 27, 29, 31.

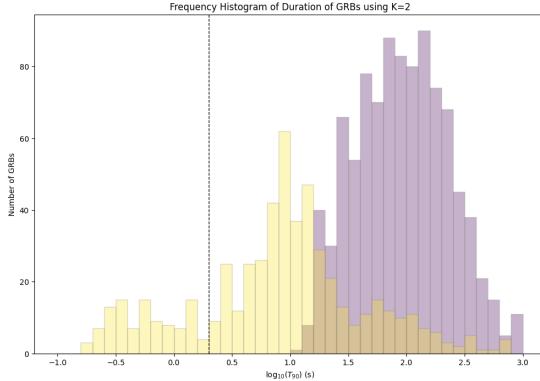


Figure 20. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 2$ on latent size 3. Colors correspond to colors previously assigned using K-means.

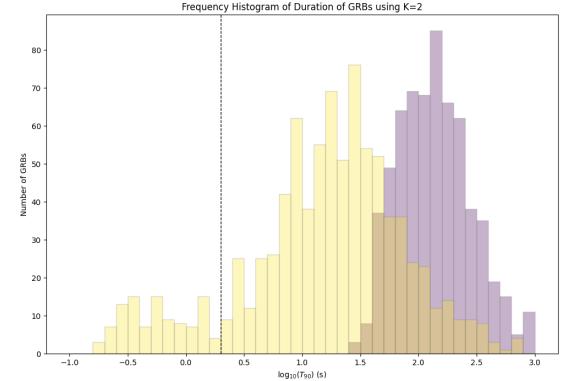


Figure 22. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 2$ on latent size 5. Colors correspond to colors previously assigned using K-means.

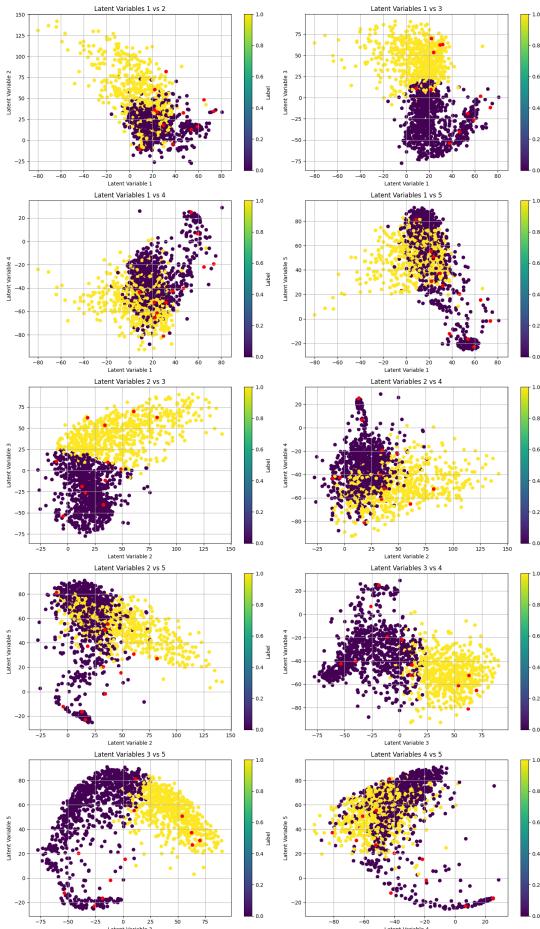


Figure 21. Color coded clusters assigned by K-means with $K = 2$ on latent size 5. Red dots are compact mergers.

4. CONCLUSION

Dimensionality reduction using autoencoders successfully results in the emergence of the visual separation of GRBs by duration of significant flux at $T_{90} = 2$ s. How-

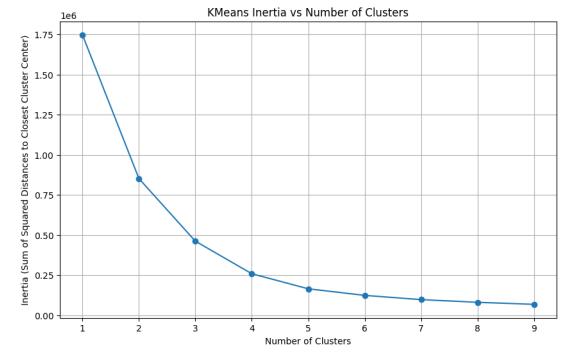


Figure 23. K-means inertia using latent size 2

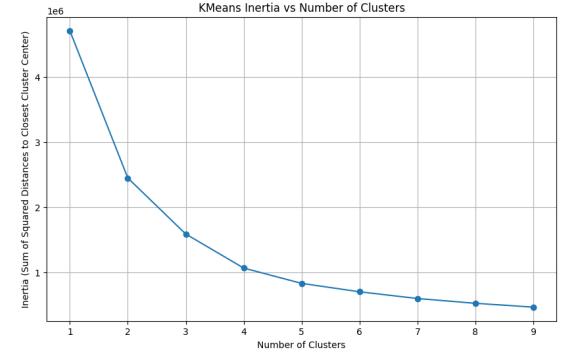


Figure 24. K-means inertia using latent size 3

ever, K-means failed to cluster these GRBs by the visual separation, as the latent space causes GRBs to wrap around so the spherical Euclidean metric fails to identify elongated or curved clusters. However, the clusters identified by K-means reflect an association with duration of significant flux. Future work in clustering can be pursued using alternative distance metric that does not depend solely on the centre of the cluster and techniques that are agnostic to the number of clusters, including Hi-

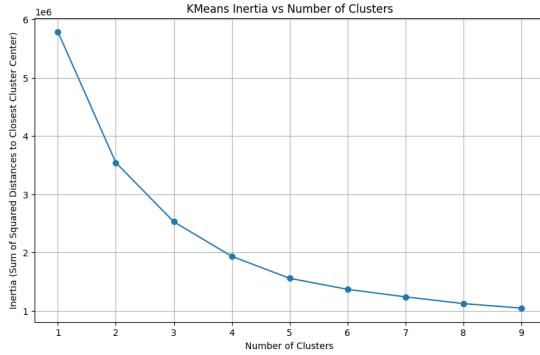


Figure 25. K-means inertia using latent size 5

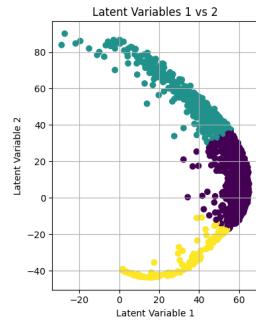


Figure 26. Color coded clusters assigned by K-means with $K = 3$ with latent size 2

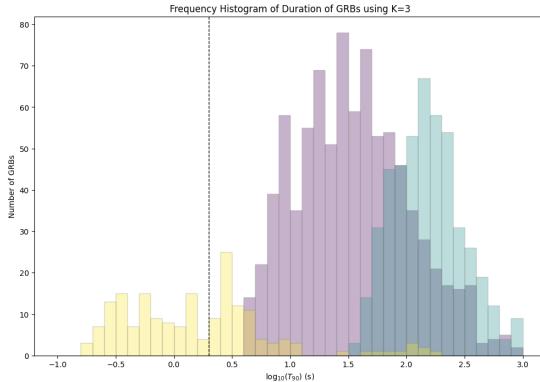


Figure 27. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 3$ on latent size 2. Colors correspond to colors previously assigned using K-means.

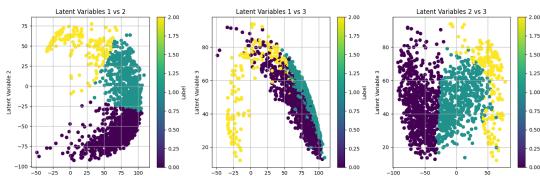


Figure 28. Color coded clusters assigned by K-means with $K = 3$ with latent size 3

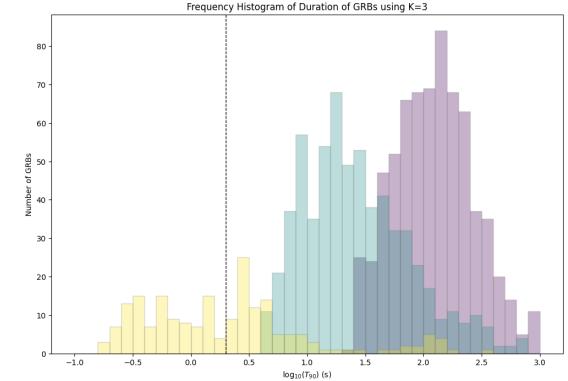


Figure 29. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 3$ on latent size 3. Colors correspond to colors previously assigned using K-means.

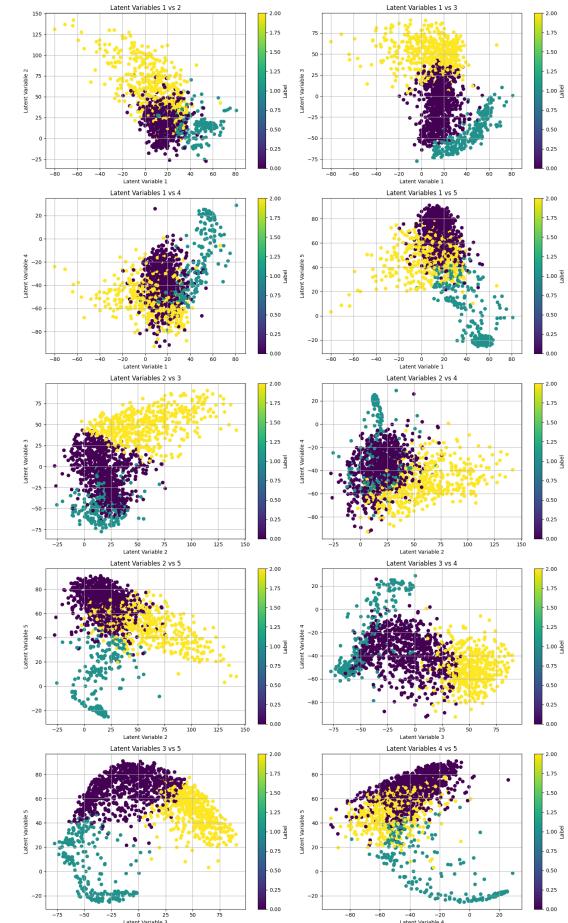


Figure 30. Color coded clusters assigned by K-means with $K = 3$ with latent size 5

erarchical Agglomerative Clustering (HAC). Under this reduction, DB-scan can also be used to identify outlier GRBs for closer inspection.

Another approach would be to counter the data imbalance by sampling the dataset according to duration as to

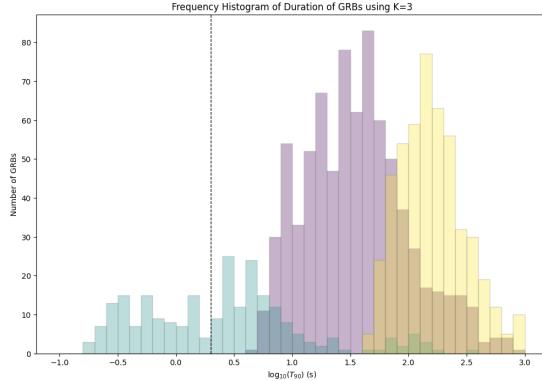


Figure 31. Frequency histogram of $\log_{10} T_{90}$ using clusters assigned by K-means using $K = 3$ on latent size 5. Colors correspond to colors previously assigned using K-means.

REFERENCES

- Berger, E., Fong, W., & Chornock, R. 2013, The Astrophysical Journal Letters, 774, L23, doi: [10.1088/2041-8205/774/2/L23](https://doi.org/10.1088/2041-8205/774/2/L23)
- Gao, H., Ding, X., Wu, X.-F., Dai, Z.-G., & Zhang, B. 2015, The Astrophysical Journal, 807, 163, doi: [10.1088/0004-637X/807/2/163](https://doi.org/10.1088/0004-637X/807/2/163)
- Gao, H., Zhang, B., Lü, H.-J., & Li, Y. 2017, The Astrophysical Journal, 837, 50, doi: [10.3847/1538-4357/aa5be3](https://doi.org/10.3847/1538-4357/aa5be3)
- Garcia-Cifuentes, K., Becerra, R. L., Colle, F. D., Cabrera, J. I., & Burgo, C. D. 2023, The Astrophysical Journal, 951, 4, doi: [10.3847/1538-4357/acd176](https://doi.org/10.3847/1538-4357/acd176)
- Jespersen, C. K., Severin, J. B., Steinhardt, C. L., et al. 2020, The Astrophysical Journal Letters, 896, L20, doi: [10.3847/2041-8213/ab964d](https://doi.org/10.3847/2041-8213/ab964d)
- Jin, Z.-P., Covino, S., Liao, N.-H., et al. 2020, A kilonova associated with GRB 070809, arXiv, doi: [10.48550/arXiv.1901.06269](https://doi.org/10.48550/arXiv.1901.06269)
- Jin, Z.-P., Li, X., Cano, Z., et al. 2015, The Astrophysical Journal Letters, 811, L22, doi: [10.1088/2041-8205/811/2/L22](https://doi.org/10.1088/2041-8205/811/2/L22)
- Jin, Z.-P., Hotokezaka, K., Li, X., et al. 2016, Nature Communications, 7, 12898, doi: [10.1038/ncomms12898](https://doi.org/10.1038/ncomms12898)
- Kasliwal, M. M., Korobkin, O., Lau, R. M., Wollaeger, R., & Fryer, C. L. 2017, The Astrophysical Journal Letters, 843, L34, doi: [10.3847/2041-8213/aa799d](https://doi.org/10.3847/2041-8213/aa799d)
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, The Astrophysical journal, 413, L101, doi: [10.1086/186969](https://doi.org/10.1086/186969)
- Levan, A. J., Malesani, D. B., Gompertz, B. P., et al. 2023, Nature Astronomy, 7, 976, doi: [10.1038/s41550-023-01998-8](https://doi.org/10.1038/s41550-023-01998-8)
- Levan, A. J., Gompertz, B. P., Salafia, O. S., et al. 2024, Nature, 626, 737, doi: [10.1038/s41586-023-06759-1](https://doi.org/10.1038/s41586-023-06759-1)
- Lü, H.-J., Yuan, H.-Y., Yi, T.-F., et al. 2022, The Astrophysical Journal Letters, 931, L23, doi: [10.3847/2041-8213/ac6e3a](https://doi.org/10.3847/2041-8213/ac6e3a)
- Mei, A., Banerjee, B., Oganesyan, G., et al. 2022, Nature, 612, 236, doi: [10.1038/s41586-022-05404-7](https://doi.org/10.1038/s41586-022-05404-7)
- Michałowski, M. J., Xu, D., Stevens, J., et al. 2018, Astronomy & Astrophysics, 616, A169, doi: [10.1051/0004-6361/201629942](https://doi.org/10.1051/0004-6361/201629942)
- Norris, J. P., & Bonnell, J. T. 2006, The Astrophysical Journal, 643, 266, doi: [10.1086/502796](https://doi.org/10.1086/502796)
- Paciesas, W. S., Meegan, C. A., Pendleton, G. N., et al. 1999, The Astrophysical Journal Supplement Series, 122, 465, doi: [10.1086/313224](https://doi.org/10.1086/313224)
- Postigo, A. d. U., Thöne, C. C., Rowlinson, A., et al. 2014, Astronomy & Astrophysics, 563, A62, doi: [10.1051/0004-6361/201322985](https://doi.org/10.1051/0004-6361/201322985)
- Rastinejad, J. C., Gompertz, B. P., Levan, A. J., et al. 2022, Nature, 612, 223, doi: [10.1038/s41586-022-05390-w](https://doi.org/10.1038/s41586-022-05390-w)
- Sanchez-Alarcon, P. M., & Sequeiros, Y. A. 2023, RAS Techniques and Instruments, 2, 129, doi: [10.1093/rasti/rzad006](https://doi.org/10.1093/rasti/rzad006)
- Stanek, K. Z., Matheson, T., Garnavich, P. M., et al. 2003, The Astrophysical Journal, 591, L17, doi: [10.1086/376976](https://doi.org/10.1086/376976)
- Tanvir, N. R., Levan, A. J., González-Fernández, C., et al. 2017, Astrophysical journal. Letters, 848, L27, doi: [10.3847/2041-8213/aa90b6](https://doi.org/10.3847/2041-8213/aa90b6)
- Troja, E., Ryan, G., Piro, L., et al. 2018, Nature Communications, 9, 4089, doi: [10.1038/s41467-018-06558-7](https://doi.org/10.1038/s41467-018-06558-7)
- Valle, M. D., Chincarini, G., Panagia, N., et al. 2006, Nature, 444, 1050, doi: [10.1038/nature05374](https://doi.org/10.1038/nature05374)

arrive at a uniform or log-uniform distribution of duration before applying K-means on the latent space. This will prevent the problem of one populous cluster from over-spilling into the other smaller but nearby cluster.

No natural clustering of compact merger GRBs are found. However, a few compact mergers cluster together in locally dense regions. Globally, the compact mergers are scattered across the latent space.