TRANSIENT LUMINOUS EVENT DETECTION WITH UNSUPERVISED CONVOLUTIONAL VARIATIONAL AUTO-ENCODER

Rasmus Christian Jørgensen (s164044)

Technical University of Denmark
DTU Space
Elektrovej Bygning 327,
2800 Kongens Lyngby

Søren Christian Winther Engell (s165518)

Technical University of Denmark
DTU Electrotechnology
Ørsteds Plads Bygning 348,
2800 Kongens Lyngby

ABSTRACT

The Modular Multispectral Imaging Array (MMIA) is an instrument onboard the Atmosphere Space Interaction Monitor (ASIM) on the International Space Station (ISS), that measures Transient Luminous Events (TLEs) on top of thunderstorms. The event classification is a difficult and time consuming task. In order to automate this process, a Convolutional Variational Autoencoder (CVAE) has been developed, with the intention of utilizing Cluster Analysis on the latent space to detect groupings in the dataset. The reconstruction of the events appear to correspond to the original data based on visual comparisons, however, a boundary issue with the reconstruction is currently present. Additionally, a fine-tuning of the hyperparameters and a score to determine how well the CVAE performs, is still missing. The project repository can be found on https://bit.ly/3rQhQHJ

1. INTRODUCTION

The Atmosphere Space Interaction Monitor (ASIM) is an instrument onboard the International Space Station (ISS). It consists of two instruments, the Modular Multispectral Imaging Array (MMIA) and the Modular X- and Gamma-ray Sensor (MXGS). The focus of the report is on investigating data from the MMIA instrument, which measures Transient Luminous Events (TLEs) in three different wavelengths; 337 nm (Nitrogen second positive, N_2P2 emission), 180-230 nm (UV) and 777.4 nm (Atmospheric Oxygen, O) [1]. TLEs cover a range of different kinds of events in the atmosphere, all of which the physical processes behind them are still being investigated. An overview of the different kind of known TLEs can be seen on fig. 1. The signal measured in each wavelengths is not yet well understood for each type of TLE, which means that currently it is a manual process to investigate and interpret the data received from the instrument. This is a slow and tedious process, which would benefit from a more systematic approach [2]. Early analysis of different types of TLEs show clear distingushing features in each

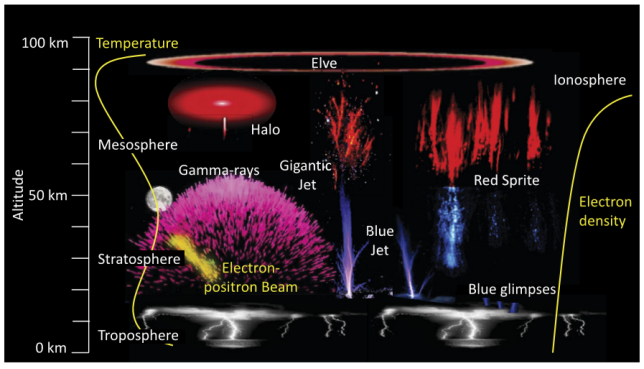

Fig. 1: Overview of the different kind of known TLEs, along with the temperature and elcetron density profile of the atmosphere, [1]

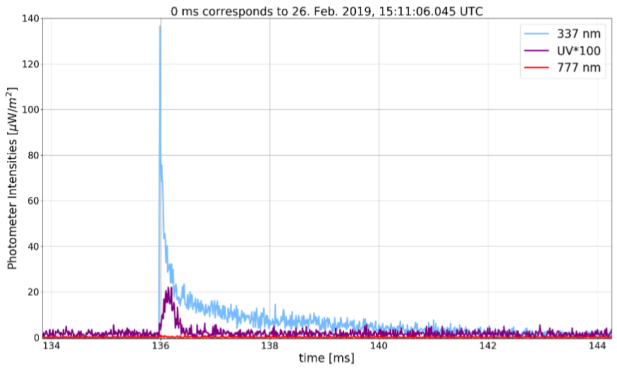

Fig. 2: The 337nm, UV, and 777.4nm spectrum of a Blue Jet event, classified and presented in [3]. For easy comparison, the three signals are stacked in the same plot here, which displays the very strong 337nm signal compared to the other two signals. The UV signal has been boosted to show the structure of the signal.

measured wavelength. For instance Blue Jets have shown extreamly fast risetimes, and high amplitude in the 337 nm band [3], see fig. 2, whereas Elves show strong signal in the UV spectrum [4]. Sprites are assumed to be caused by strong positive cloud-to-ground (CG) lightning, which means that we assume a strong 777.4nm signal, in combination with possibly a UV peak [5]. These observations are still based on limited observations, but at this point, the assumption is that each type of TLE has its own defining signal, whether it is dependent on risetime, amplitude, or perhaps correlation between spectres.

In order to achieve a more systematic approach of searching the data for events, the goal is to try and stack the three spectra (337nm, UV, 777.4nm) utilizing a 1D convolutional structure and applying variational autoencoders to the stacked spectra. The hope is that this will reveal clusters in the latent space of the encoding, which corresponds to different types of TLEs. A similar neural net structure has been developed with the purpose of detecting anomalies in flight data [6]. The structure of the neural network presented by [6] will be the basis of the neural net developed and discussed in this report.

2. THEORY

A variational autoencoder (VAE) consist of an encoder which maps the original data to a lower dimensional latent space and a decoder which reconstruct the original space from latent samples. It can be defined as an autoencoder whose training is regularized to avoid overfitting and ensure the latent space has good properties that enable generative process [7]. The model is trained in the following way [7]:

1. The input is encoded as a distribution over the latent space.
2. A point from the latent space is sampled from that distribution.
3. The sampled point is decoded, and the reconstruction error can be computed.
4. The reconstruction error is backpropagated through the network.

If $x \in X$ is the original data and $z \in Z$ is the latent variable then the posterior $p(z|x)$ can be determined from Bayes rule

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \tag{1}$$

The denominator can be calculated by computing an integral, however this is usually intractable [7]. Instead, variational inference (VI) is used to approximate the posterior. VI relies on optimising the evidence lower bound (ELBO) which can be expressed using a Kullback-Leibler (KL) divergence.

$$\mathcal{L}(\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})|p(\mathbf{z})\right) \tag{2}$$

Where $q_\phi(\mathbf{z}|\mathbf{x})$ is the approximate posterior and D_{KL} is the KL divergence. The first term in equation 2 is called the reconstruction error and the second term is called the regularisation term. Optimizing ELBO is a tradeoff between these two terms. It is possible to insert a β value in front of the regularisation term as a regularisation hyperparameter. It has been shown that a β-VAE with appropriately tuned $\beta > 1$ qualitatively outperforms a VAE ($\beta = 1$) [8]. The expectation term in equation 2 require generating samples $\mathbf{z} \sim \mathbf{q}_\phi(\mathbf{z}|\mathbf{x})$, which is a stochastic process and cannot be backpropagated. To be able to train the network the reparameterization trick is used. This makes it possible to express the random variable $\mathbf{z}$ as a deterministic variable $\mathbf{z} = \mathbf{T}_\phi(\mathbf{x}, \epsilon)$, where ϵ is an independent random variable and $\mathbf{T}_\phi$ is the transformation function that converts ϵ to $\mathbf{z}$ [9].

3. METHOD

The dataset is constructed from measurements taken every 7th day over the period of one year. These measurements have been randomly split into a training and validation set. The training set contains 90% of the measurements while the validation set contains 10%. However only 30% of the training and validation sets has been used since jupitor notebook crashed when using too large datasets. This gives 1277 detected events used for validation and 11496 events used for training.

The architecture of the Convolutional Variational Auto-Encoder (CVAE) used for the obtained results can be seen in figure 3. Before the network can be trained then the raw input data needs some prepossessing. The raw input data varies in length, which is difficult to handle in a convolution. There are two easy ways to manage this. The first is to use zeropadding and make all signals equal in length to the longest detected signal. The second is to slice all signals into smaller parts of equal length. We have tried both approaches and found that slicing gave the best reconstruction of the three spectra. When the total length of the input signal is not divisible by the slicing length then the end of the signal can be discarded without losing much information, as long as the slicing length is small.

The chosen observation model $p(x|z)$ is a beta distribution. This is due to the random behaviour of noise signals in large parts of the dataset. The beta distribution is defined on the interval [0, 1] and parameterised by two positive shape parameters α and β. For this reason the dataset has been normalised before training. The prior is a standard Gaussian, $N(0, I)$, and the approximate posterior follows a gaussian distribution as well, $N(g(x), h(x))$. ELBO from equation 2 becomes

$$\mathcal{L}(\mathbf{x}) = \frac{1}{L}\sum_{l=1}^{L} \log p_\theta(x|z^{(l)}) + \sum_{i=1}^{n} \left(\sigma_i^2 + \mu_i^2 - 2\log(\sigma_i) - 1\right)$$

Where μ and σ are the mean and standard variation parameters of the approximate posterior $q_\phi(\mathbf{z}|\mathbf{x})$.

Figure 3 shows that the CVAE has a convolutional layer followed by a fully connected layer in the encoder and that the decoder uses a similar architecture. The difference is an extra transposed convolution in the decoder. The reason for the extra transposed convolutional layer is that the beta distribution requires both an α and β parameter, that both need to be learned. Therefore, it is necessary to have double the length of the original signal, which can then be chunked into two parts, corresponding to the α and β parameter respectively.

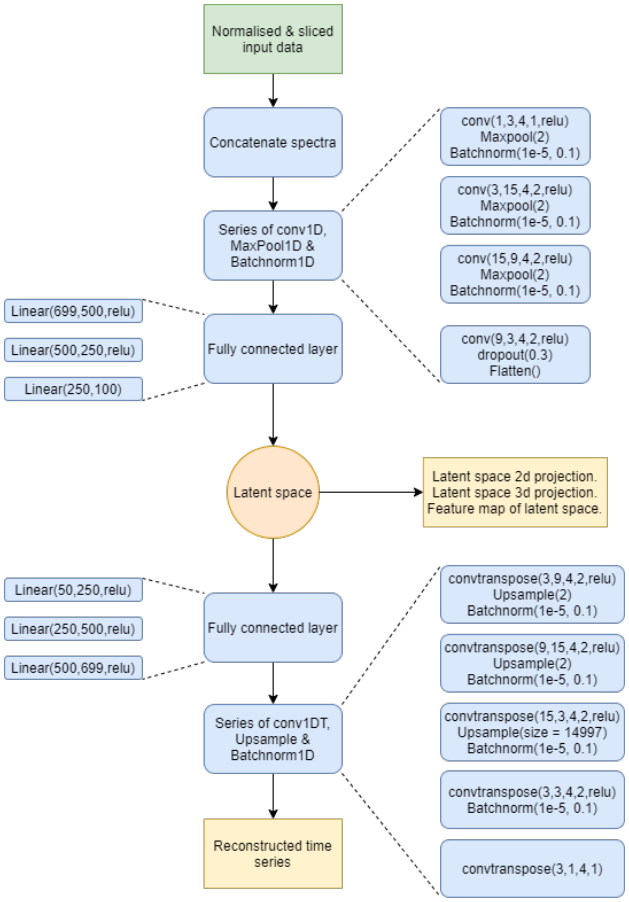

Fig. 3: Convolutional Variational Auto-Encoder (CVAE) network architecture. To the right the structure of the convolutional layer is shown. Conv and convtranspose stands for 1D convolution and transposed convolution. The inputs are: (number of input channels, number of output channels, kernel size, stride, activation function). To the left the structure of the linear layer is shown. The inputs are: (input size, output size, activation function)

This is similar to the last layer in the encoder, which has double the size of the expected latent layer, which is chunked into a σ and μ parameter. In the decoder the third upsample has been hardcoded to size 14997, to get the correct dimensions. Since the latent space has 50 dimensions, we make a 2D and 3D projection of the latent space and a feature map of the actual 50 dimensional latent space, which can all be used for evaluation.

4. RESULTS & DISCUSSION

Now it is possible to look at an example of a part of a time series and its reconstruction, to try and validate if the CVAE is working as intended. On fig. 4 an observation and its reconstruction is shown. Some things are very apparent; the CVAE appear to pick up on the features in the 337nm and 777.4nm channels, and correctly reconstruct them as correlated signals. The UV spectrum has a peak at the very end, of which the amplitude is equal to that of the observations peak in the UV spectrum. However, the location is not correct, and appears uncorrelated with the two other signals, which is not the case as seen from the observation. Additionally, there appear to be a problem with the 337nm channel in the boundary, with an initial peak of amplitude 1, which is not seen in the observation. This makes it hard to properly compare the features in the channels. For this reason, a similar plot without the boundaries of the signal is made, as shown on fig. 5. Now it is much clearer that the features in 337nm and 777.4nm channels are accurately reconstructed, even capturing the faint signal after the peak itself. However, the steep

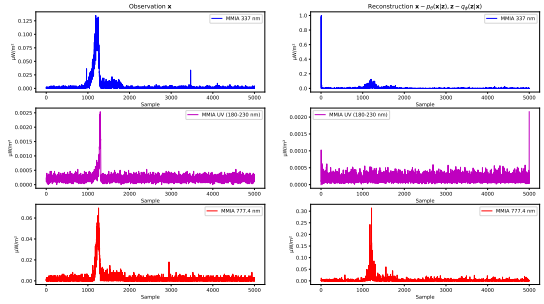

Fig. 4: Example of an observed signal and the reconstruction. On the left is the observation, **x**. This particular case have not been classified, but shows a clear peak in all three channels, with a sudden drop off after reaching its peak. On the right is the reconstruction, which shows that the CVAE appear to pick up the features present in 337nm and 777.4nm, but not in the UV. The y-axis corresponds to the photonflux, μWm^{-2}, and the x-axis is the samples, unitless. The upper figure (blue) is the 337nm band, the middle (magenta) is the UV band, and the bottom (red) is the 777.4nm band.

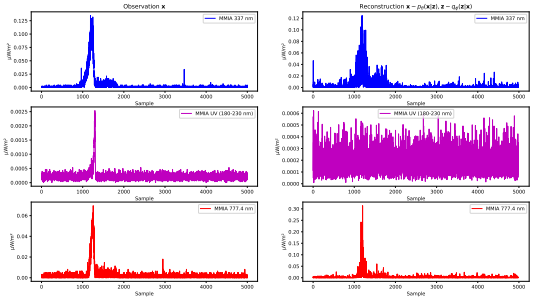

Fig. 5: Same event as shown on fig. 4, however the first, and last, samples of the signal has not been plotted, to show the features in the reconstruction that were overshadowed by the boundary error. The y-axis corresponds to the photonflux, μWm^{-2}, and the x-axis is the samples, unitless. The upper figure (blue) is the 337nm band, the middle (magenta) is the UV band, and the bottom (red) is the 777.4nm band.

drop-off right after the observed peak is not present in the reconstruction, which appear more smooth. Additionally, it is apparant that the UV spectrum unfortunetly does not pick up any features correlated with the 337nm and 777.4nm, as would be expected in this case.

Next up, we can look at the latent space of the encoding, as shown on fig. 6. From the 2D projection, it is hard to see any structure, except for perhaps small local clusters, for instance around (2.1, 1.75). The same is true for the 3D projection, however one feature appear to be located quite far away from the remaining features. In the feature map, some structure does appear to take form. For instace, feature 11 appear to be mostly negative, with small outliers. However, at this point, it is still hard to pinpoint any defining features in the feature map.

Additionally, a lot of optimization with regards to the hyperparameters is still needed. As of now, the CVAE is trained for 300 epochs, with batch size 100, and no regularization on the KL parameter in the ELBO. Perhaps one of the more interesting hyperparemeters to look into would be the size of the latent space. Since we are dealing with very long time series, with a lot of samples, increasing the size of the latent space could perhaps reduce the need for slicing the series. Likewise, increasing the channels of the convolutional layers could also be expected to increase the learning capabilities of the CVAE. Of course this is mostly speculation at this point, but would perhaps be a good place to start the optimization.
Finally, a better way of evaluating how well the CVAE performs is necessary. Determining a concrete parameter is difficult, as the data is unlabelled. This means that conventional methods, such as comparing the predicted label to the ground truth is not an options. Perhaps one approach could be to use the few catagorized events that currently exist, such as those presented in [3], [4], [5], and encode these events. By inspecting the latent space corresponding to these events, it would be possible to assign the clusters these events fall into, and from that determine the intra-cluster distance to determine how well the CVAE clusters the few known events we have, inspired by the approch in [10], and also used in [11]. We recognize that the proposed idea here is not identical, but the idea of determining the location of a few known labels in the latent space, and using these as 'pointers' for potential groups appears to be intuitively easy to understand. However, if the CVAE is not able to correctly distinguish the finer details in the measured time series, this could cause problems. For instance, normal lightning events could be grouped with Sprites, as these are expected to origin from strong positive CG ligthning, which means that both signals would be present in the data, with only a slight time delay. At this point in time, this is also still speculation, and an actual test would be necesary in order to evaluate the approach.

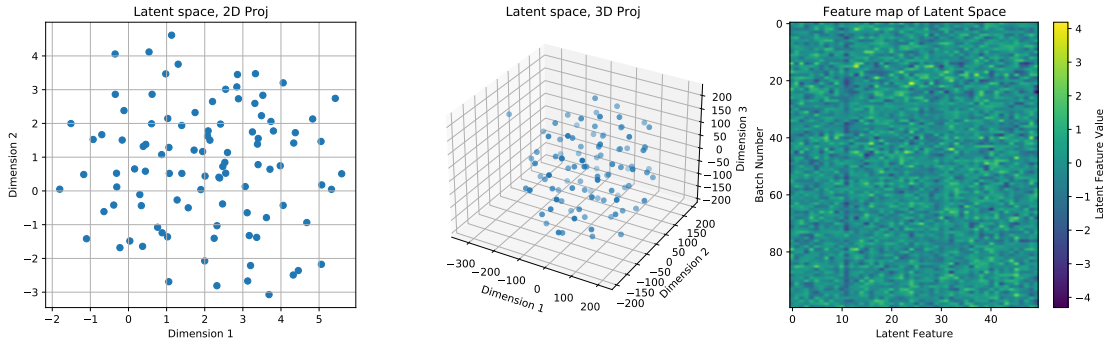

Fig. 6: The latent space of the CVAE. Since the latent space is 50 features, a 2D projection is shown on the left, a 3D projection in the middle, and a feature map of all 50 on the right. t-SNE is used in order to create the 2D and 3D projections.

5. CONCLUSION

The goal of the report has been to develop a CVAE in order to encode and decode time series provided by the MMIA instrument, onboard the ASIM mission. The reason for developing the CVAE is to conduct a cluster analysis on the latent space, with the intentions of finding outliers in the data, which would correspond to interesting events in the data. Looking at the reconstruction of one of the time series, fig. 5, the CVAE appears to be able to accurately reconstruct the details in the 337nm and 777.4nm band, but is unable to determine the details in the UV band. This is possibly because the UV band is much weaker than the other two bands. Looking at fig. 4, there also appears to be a boundary issues in the reconstruction. Looking at the latent space, fig. 6, the feature map begins to reveal faint features, these are however very hard to locate in the 2D and 3D projections. The presence of faint features in the latent space could indicate that the approach is worth pursuing, and with fine-tuning the CVAE better results could hopefully be achieved.

Finally, a proper optimization of the hyperparameters of the CVAE is necessary. In order to do this, a score of how well the CVAE performs is necessary, opposed to simply visually comparing the reconstruction to the original data. An approach similar to that presented in [10] and [11] is suggested, but has not been implemented yet. Since a few classified events exist, it is possible to encode these and determine their latent space as a baseline. However, due to the very limited amount of classified events, this is still a naive approach.

6. REFERENCES

[1] T. Neubert, N. Østergaard, V. Reglero, E Blanc, O. Chanrion, C. A. Oxborrow, A. Orr, M. Tacconi, O. Hartnack, and D. D. V. Bhanderi, "The asim mission on the international space station," *Space Sci Rev*, 2019.

[2] T. Neubert and O. Chanrion, "Personal communication," 2020.

[3] T. Neubert, N. Østgaard, O. Chanrion, M. Heumesser, K. Dimitriadou, L. Husbjerg, and V. Reglero, "The genesis of blue lightning into the stratosphere," *American Geophysical Union*, 2019.

[4] T. Neubert, N. Østgaard, V. Reglero, O. Chanrion, M. Heumesser, K. Dimitriadou, F. Christiansen, C. Budtz-Jørgensen, I. Kuvvetli, I. L. Rasmussen, A. Mezentsev, M. Marisaldi, K. Ullaland, G. Genov, S. Yang, P. Kochkin, J. Navarro-Gonzalez, P. H. Connell, and C. J. Eyles, "A terrestrial gamma-ray flash and ionospheric ultraviolet emissions powered by lightning," *Science*, vol. 367, no. 6474, pp. 183–186, 2020.

[5] Toru Adachi, Mitsuteru Sato, Tomoo Ushio, Atsushi Yamazaki, Makoto Suzuki, Masayuki Kikuchi, Yukihiro Takahashi, Umran S. Inan, Ivan Linscott, Yasuhide Hobara, Harald U. Frey, Stephen B. Mende, Alfred B. Chen, Rue-Ron Hsu, and Kenichi Kusunoki, "Identifying the occurrence of lightning and transient luminous events by nadir spectrophotometric observation," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 145, pp. 85 – 97, 2016.

[6] M. Memarzadeh, B. Matthews, and I. Avrekh, "Unsupervised anomaly detection in flight data using convolutional variational auto-encoder," *Aerospace*, 2020.

[7] "Understanding variational autoencoders," https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73.

[8] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohammed, and A. Lerchner, "β-vae: Learning basic visual concepts with a constrained variational framework," *ICLR*, 2017.

[9] "From autoencoder to beta-vae," https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.html.

[10] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian J. Goodfellow, "Adversarial autoencoders," *CoRR*, vol. abs/1511.05644, 2015.

[11] Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, and Taylor Berg-Kirkpatrick, "Improved variational autoencoders for text modeling using dilated convolutions," *CoRR*, vol. abs/1702.08139, 2017.