The Dr. Bessie F. Lawrence 53rd International Summer Science Institute 2022

Journal of Scientific Reports

Dear graduates of the 2022 ISSI,

A few months ago, we stood in front of a difficult decision to hold the ISSI in as a digital program for a second year. Covid-19 has changed our social and cultural behaviors, affected our habits, and our means to interact, study and socialize. Following the successful program last summer, we had a drive to further improve the program this year.

It was important to us to preserve the spirit of the ISSI, to bring the atmosphere of the Weizmann, the open doors and open minds approach which are at the heart of the Weizmann Institute. Through diverse scientific talks, discussing the harmony and boundaries of arts and science, asking questions and raising theories- all these provide the foundations to enable the freedom to think differently and the courage to dare, to test, to go where no one has gone before and to pave your own road with partners, collaborators and friends from any field and discipline, even out of scientific fields.

We have put together the structure, but you students made it your own meaningful experience. You have shown your commitment and overcame the obstacles along the way: The different time zones, maximizing your interactions with your mentors and peers, and came forward to assist where needed. The outcomes of your hard work were highly commended by your mentors and creatively crafted into your final presentations, demonstrating your dedication and enthusiasm in your scientific tasks. The mini-papers presented ahead are yet another part of your high achievements in this short period of research, and an introductory step into the endless world of scientific discoveries.

We hope that by participating in this program you were able to acquire new sets of skills and tools for your future academic studies, career, and as a general life habit no less. That the broad outlook on science as a collaborative field has encouraged you to create a community of shared interests and passion, and we wish for you that these friendships will last for life as you are now members of the ISSI community.

It was our great pleasure to experience the ISSI program of July 2022 with every one of you. We wish you all the best and continued success in your future academic studies and career.

Sincerely,

ISSI coordinators

Dr. Dorit Granot, Dr. Aya Shkedy & Ms. Nirit Alon

Table of Contents

Making the Invisible Visible: Regression and Artificial Neural Networks in Low Light Signal Detection

Avraham Balsam[4], Jonas Kuehne[3], Anna Noyvert[3], Jansen Wong[4], Xinyue Yu[1]

China (Hong Kong)[1], Switzerland[2], United Kingdom[3], United States[4]

Mentored by Dr. Vyacheslav Kalchenko
Department of Veterinary Resources
Weizmann Institute of Science, Rehovot, Israel

Abstract

Data analysis is crucial for biological experiments, and often involves substantial manpower investments, high time cost, and strict requirements of accuracy. Moreover, manual detection of low-light signals is sometimes complicated and tedious. Consequently, finding efficient solutions for data generation and analysis is critical for the continued progress of scientific research. Many researchers have used machine learning, supervised and unsupervised, to detect low-light signals in biological experiments. In our study, we used some of those methods on kinetic frame sequences to recognize anomalous patterns which may not be discernible from single frame data. We used a variety of supervised and unsupervised machine learning algorithms, and eventually concluded that Truncated Singular Value Decomposition and Non-Negative Matrix Factorization are the most effective. Additionally, we explored a novel use case of the autoencoder as an effective tool for feature analysis and denoising. To make our research available to the general public, we developed a Flask application with the capacity to enhance user-submitted images. We hope that our research will open the world of low-light image analysis to a wider audience, and thereby increase the efficiency of cross-disciplinary research intersecting with biophotonics.

Motivations

Since scientists began conducting biological experiments on mice, ethical controversies were raised regarding animal experimentation. According to a new study, more than 111 million mice and rats are experimented on annually in United States biomedical research, and the number of mice and rats used for experiments all over the world may be much larger [1]. Additionally, machine learning data generation and analysis algorithms are often more cost effective than traditional methods. Mainly based on these two factors, machine learning may be useful as a complement to traditional low-light image analysis solutions.

Machine learning

Machine learning (ML) is the subset of artificial intelligence (AI) that focuses on building systems that learn from input data [2]. In short, machine learning is performed by networks or systems that adapt to rules or regulations when given input data to match the output data. The process of machine learning involves discovering patterns in the data, making predictions, and then offering suggestions or solutions for future improvement. Moreover, once the networks or systems are built, they can be used to improve the efficiency of data analysis. Machine learning algorithms can be categorized into two types: supervised learning and unsupervised learning. In essence, supervised learning takes labeled data, where in unsupervised learning the data is unlabeled.

Materials and Methods

The data used in this study were taken from existing physical bioluminescence images, and this limited dataset was then augmented with artificially generated mouse bioluminescent images generated in Blender that accurately simulate the behavior of different luciferin-luciferase kinetic reactions that occur in test subjects.

For the supervised learning training, which requires both an underexposed input image and a well-exposed target image, we used three frames from the time sequence of both the artificial and real bioluminescent data: the first frame, the 30th frame which was determined to consistently have the highest bioluminescent signal, and the last frame. Using these frames, we set the first frame as the red channel of the resulting image, the 30th frame as the green channel, and the last frame as the blue channel. To create the underexposed input image, we multiplied the green channel with the brightest signal by 0.4 to dim the signal, and to create the well-exposed target image, we multiplied the green channel with the brightest signal by 2 to enhance the signal, see Fig. 1. Then, using these image pairs, we trained the supervised learning model.

Supervised Learning Process

In supervised learning, labeled datasets are used to train models to classify data or predict possible outcomes accurately. After weights of input data in neural networks are adjusted according to the cross-validation process, they can be applied to execute further data classification or analysis [3]. In this study, we used an existing neural network called MIRNet [4, Fig. 2], which was pre-trained to enhance low-light images, followed by transfer learning to transfer the knowledge learned from the source dataset to our target dataset. MIRNet uses the LOL dataset, composed of 500 low-light and normal-light image pairs and divided into 485 training pairs and 15 testing pairs. Each image pair in the dataset consists of a low-light input image and its corresponding, well-exposed reference image, which acts as a label and allows the network to learn how to recover high-quality image content from its degraded content [5]. During the fine-tuning process, we froze some layers in the network to ensure that the weights in these layers were not changed when the model was partially re-trained on our data. We then used our RGB image pairs to re-train the neural network, which updates the weights in specific unfrozen layers, to make it better at enhancing the light sources in our bioluminescent images. However, training of MIRNet with our dataset was accomplished with an unsatisfactory outcome (MIRNet contains over 12 million parameters and over 1000 layers). We were not able to fully finish the fine-tuning process to get the desired results, but with some more adjustments, it should be able to work as expected.

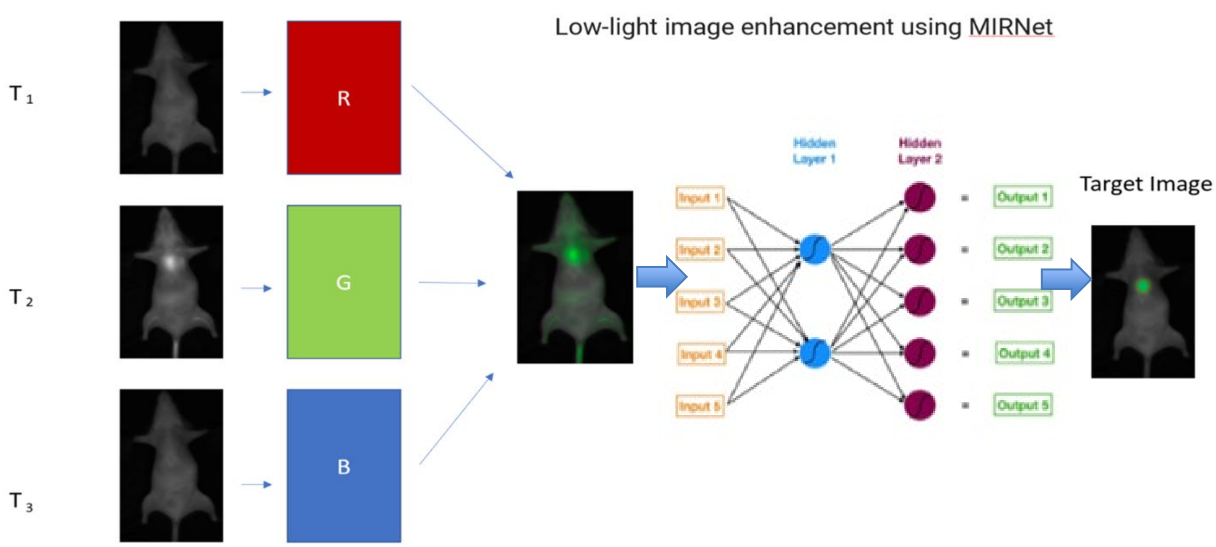

Figure 1. Process of supervised learning in this study.

Figure 2. MIRNet enhanced image examples.

Unsupervised Learning Methods

There exist many methods for unsupervised learning. One of them is dimensionality reduction. One can imagine this process like expressing a data set with much fewer data. The context of images one could imagine an object from multiple perspectives. We can, for example, imagine an object with just one isometric view. The object is not represented accurately with, for example, just the view from top. Our goal was to represent the whole sequence of images, picturing the luciferin decay, with one single image.

This, in essence, is the idea of dimensionality reduction. There exist multiple algorithms to do so. We tried a number of them. Those included principle component analysis (PCA), sparse pda, factor analysis, independent component analysis, truncated singular value decomposition (TSVD) and non-negative matrix factorization (NMF)[6]. To quantify our results, we defined an error function, which shows the distance between the right and the detected location of the brightest spot in the image. This was done for multiple noise strengths.

Autoencoder

Another approach used an autoencoder to detect anomalies in individual images. Autoencoders are a specialized type of artificial neural network, sometimes called "semi-supervised" since they lie on the border between supervised and unsupervised learning. Autoencoders are a composite of two neural networks: the encoder, which compresses the input data into a lower dimensionality vector, and the decoder, which decodes that vector to reconstruct the original image [7]. This approach has been used for noise reduction [8] and anomaly detection [9], both of which have fundamental relevance to low-light image anomaly detection.

To minimize the risk of overfitting and streamline the data analysis process, we chose not to combine the training data we generated into a single dataset. Instead, each individual frame sequence was segmented into one hundred individual images which were passed to an autoencoder as training data. This approach allowed our model to make full use of the kinetics incorporated into the training data, as it could more accurately analyze patterns in blob activation and noise intensity. The network was trained for two hundred epochs, after which it was able to successfully denoise and detect anomalies in the frame sequence it was passed. Because the autoencoder is trained on an individual frame sequence, training is relatively short (55 seconds) when compared to other machine learning algorithms.

Results

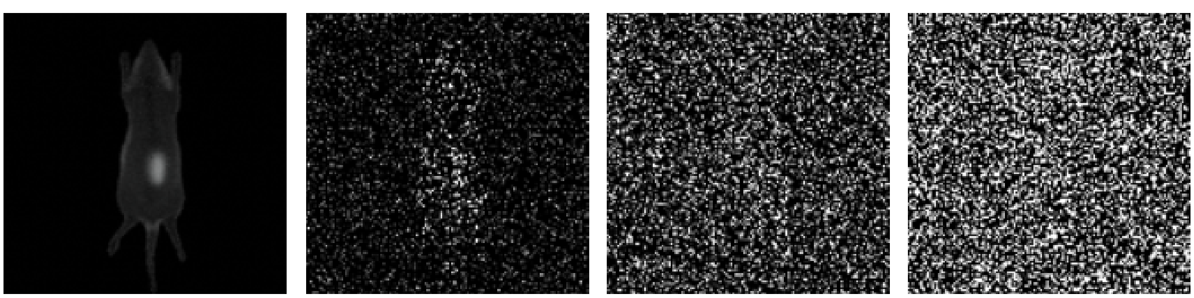

Figure 3. Combined noise of strength 0.0, 0.1, 0.9, 1.9 (left to right) at peak intensity frame

To compare performance, we required images with different noise levels. The noise in our data is a combination of Poisson- and Gaussian noise [10, Fig. 3].

Unfortunately, we did not have the time to get results with the approach using supervised learning, thus it will not be mentioned in this section.

With the unsupervised methods we used, we were able to compute the above-mentioned error measure and averaged it over all 500 samples, see Fig. 4.

It becomes clear that the TSVD and NMF algorithms outperform the rest. If computation time is taken into account, we suggest the use of TSVD, as it takes on average 0.9 seconds per sample while NMF takes 2.2 seconds. Thus, only results generated with this algorithm will be shown, see Fig. 5.

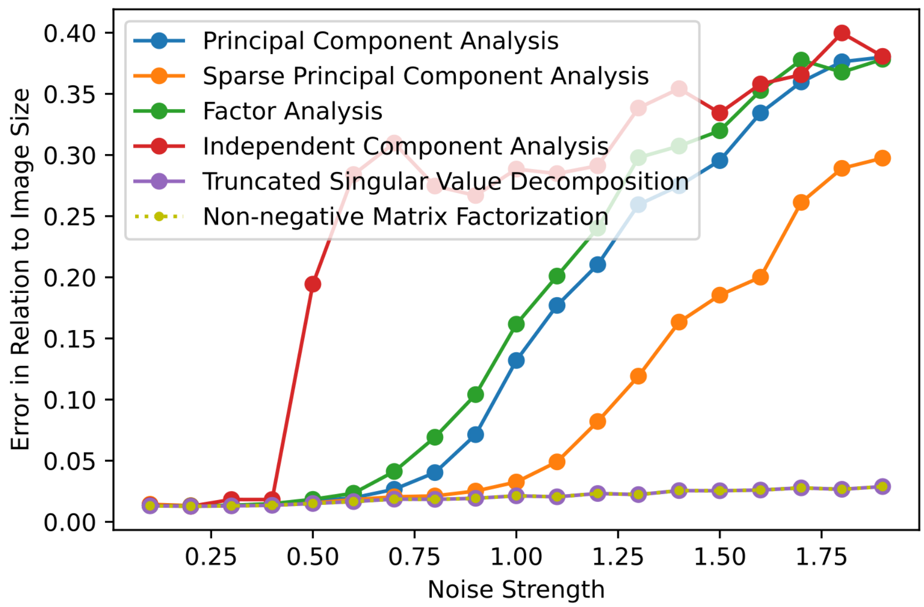

Figure 4. Performance of the different dimensionality reduction / regression algorithms.

For the auto-encoder, we were also able to create a working structure, see Fig. 5.

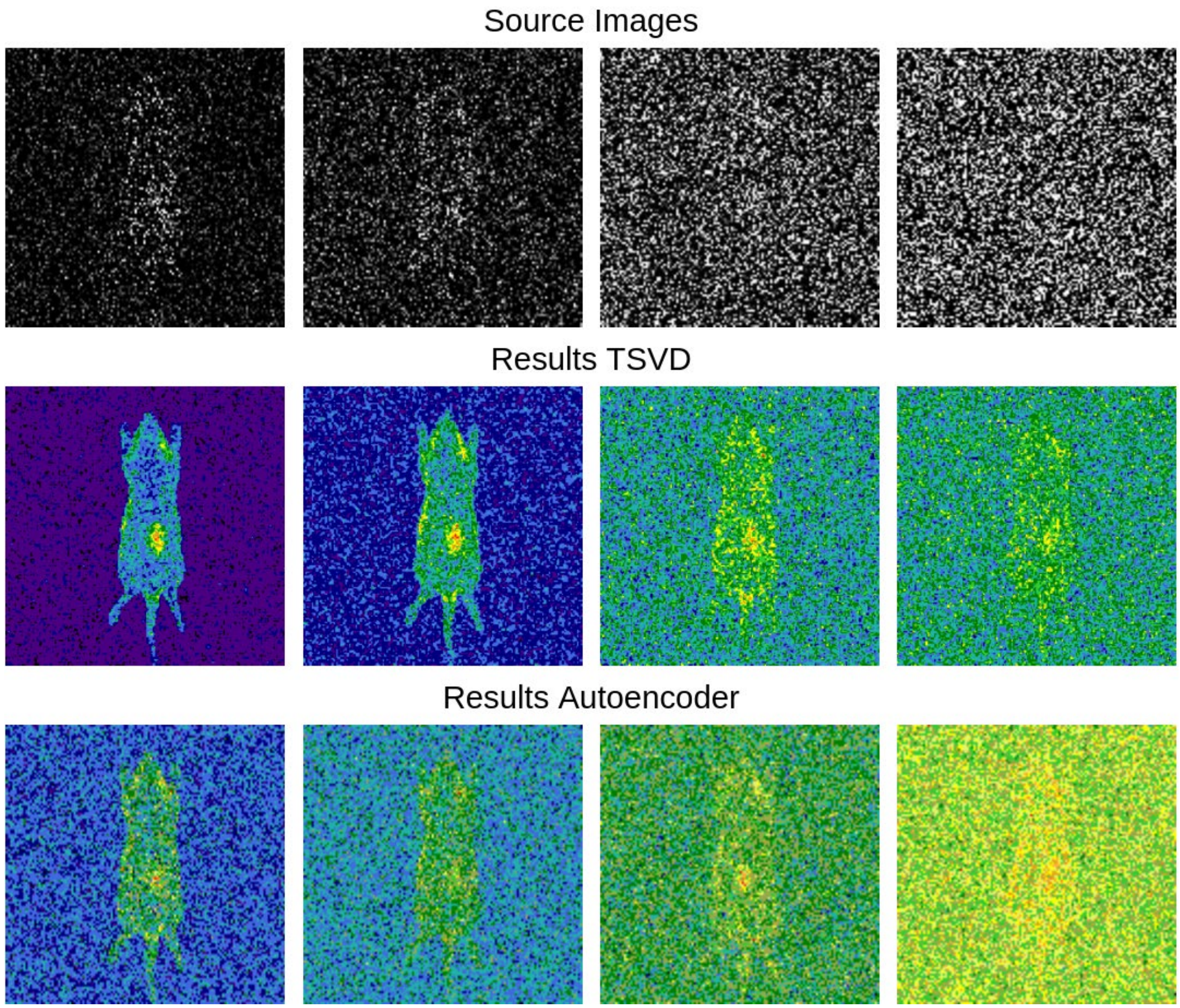

Figure 5. Results of TSVD and autoencoder on noise strengths 0.0, 0.1, 0.9, 1.9 (left to right).

To further enhance the images, we applied some minor denoising to the results generated with TSVD, see Fig. 6.

Comparison

As shown in the previous section, dimensionality reduction is marginally more effective at denoising and feature analysis than the autoencoder, although both were able to detect low-signal images at high noise intensities. With further refinement, we believe that both approaches can be improved upon. These results may be useful in the field of bioluminescence imaging as well as any other study which involves kinetic data analysis.

With the additional denoising, which could also be applied to the results of the autoencoder, the signal can be detected and shown clearly, even at high noise strength.

If we compare computation time, then the processing of each sample takes with the autoencoder (55s) about 55 times longer than with TSVD (0.9s).

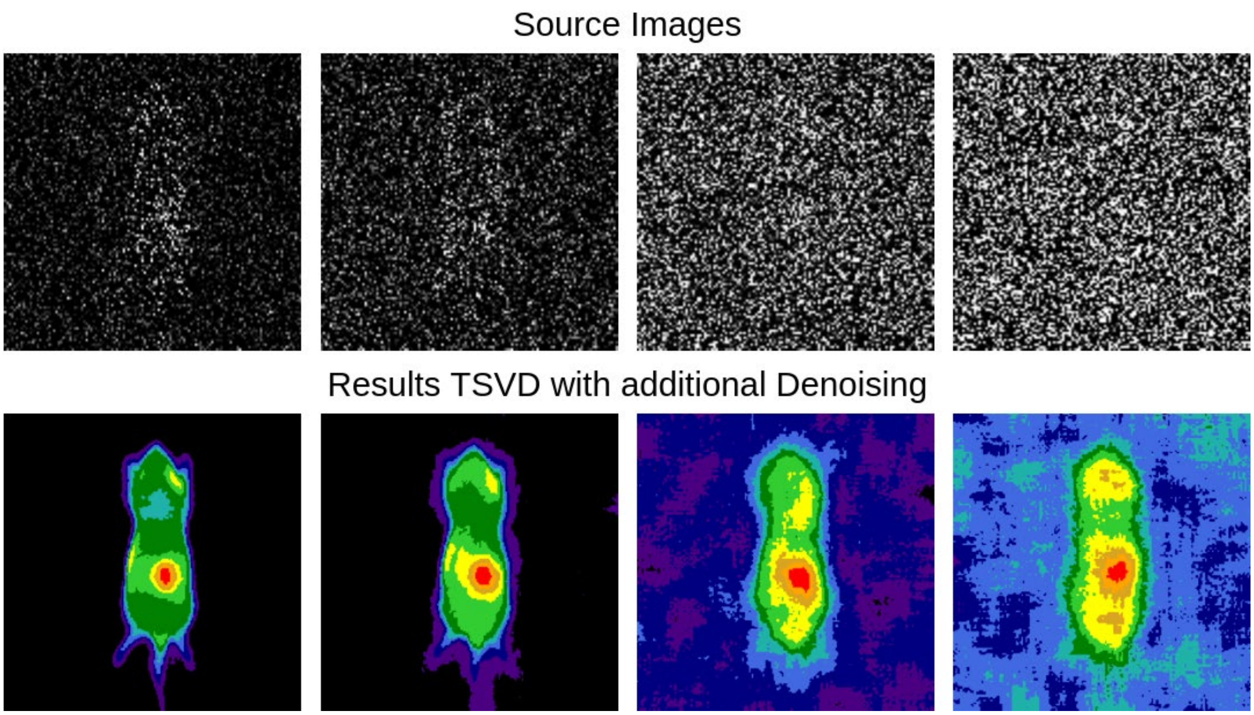

Figure 6. Results of TSVD with additional denoising on noise strengths 0.0, 0.1, 0.9, 1.9 (left to right).

Flask

To make the machine learning algorithms we used more accessible, we developed a Flask application which incorporates all of the unsupervised learning algorithms discussed in this paper, see Fig. 7. The user uploads a noisy image frame sequence, and the server enhances it and detects any anomalies that may be present in the image. We implemented a customizable form template to allow the end user to adapt our software to their own needs; the user can toggle the colormap in which images will be displayed, decide which algorithms to use, and set some low-level parameters of the autoencoder. The application was deployed using Heroku and can be accessed by anybody with an internet connection. We hope that this Flask application will open our methods to the broader scientific world and make clear the benefits of kinetic data analysis.

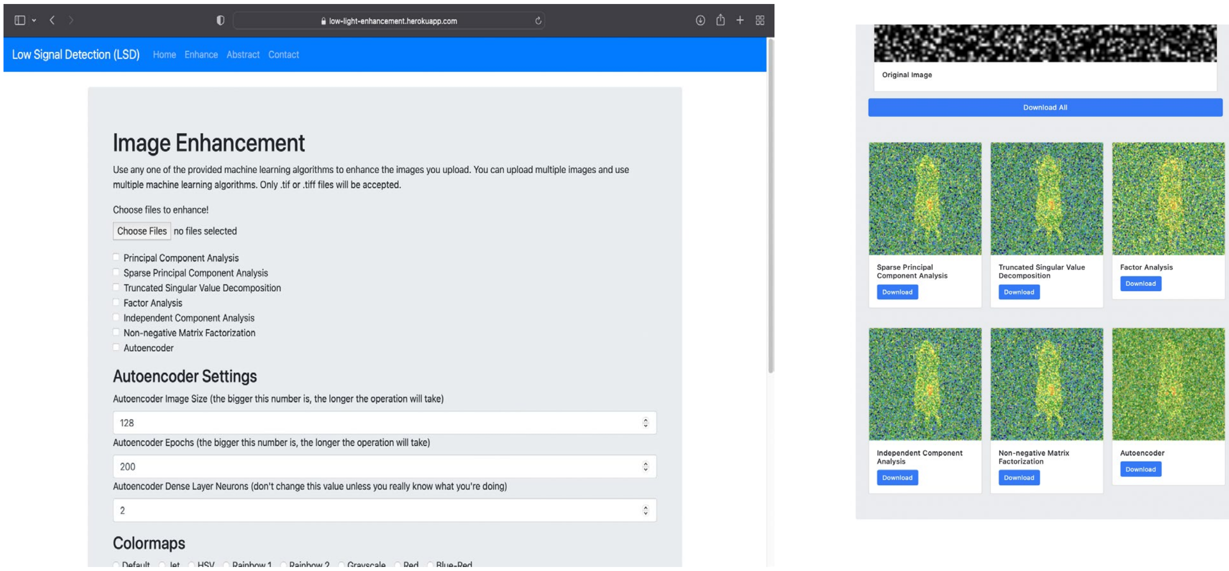

Fig. 7. User interface of Flask application

Discussion

We believe we have conclusively demonstrated the effectiveness of unsupervised machine learning techniques in analyzing noisy and low-quality images using kinetics. Using dimensionality reduction techniques, we were able to detect anomalies in an image which, to the naked eye, were indiscernible. We also have shown that for our examined data, these techniques have achieved improved results in comparison to supervised and semi-supervised machine learning techniques, both of which require long training periods and a wealth of training data. To an extent, the fact that these images behaved in a similar fashion to real-world data when subjected to analysis indicates that our synthetic data generation system accurately mimics low-light imaging conditions. However, it is still important to recognize the limitations of some of the approaches we have used in this research. In order to conserve computational resources and demonstrate that data generation is possible even without high-end software, we used Blender to generate our data. In order to rigorously test the efficacy of the algorithms we used, we must test them on real bioluminescence imaging data. Additionally, although we accounted for both gaussian noise and poisson noise, we have not tested the noise our system generates against natural noise. In the future, we hope to quantify the similarity of our data to random noise and continue to improve the accuracy of our synthetic data. Also, we will incorporate more sophisticated Monte Carlo simulations such as Blender Photonics and MCX into our model and use them to generate more accurate data.

As discussed earlier, the systems we have developed can aid the scientific community in quickening the pace and humanity of preclinical studies. Our models require no specialized knowledge of Monte Carlo systems nor a deep understanding of computer programming.

Lowering the threshold for participation and engagement with the biophotonics community will allow for more efficient and productive cross-disciplinary research, and for greater exposure of the biophotonics community to the greater scientific world.

However, while our methods are effective as a tool for kinetic data analysis, the tools we developed will not work on still images. Thus, researchers without access to multi-frame images cannot use our software. Additionally, training the autoencoder is much more time-consuming than implementing one of the other unsupervised methods, which perform marginally better. Hyperparameter optimization of the autoencoder is also more difficult than optimization of the regression methods. But these weaknesses are also strengths. The autoencoder is a complicated and powerful instrument which, when applied in the correct circumstances, can solve complicated problems. We hope that the novel approach we have developed will be used by other scientists to increase the power and effectiveness of the autoencoder in kinetic imaging. Some possible avenues of exploration include convolutional [11], variational [12], and LSTM autoencoders [13].

Acknowledgements

We would like to express our gratitude to the research group of Dr. Slava Kalchenko and the Department of Veterinary Resources for providing us with the necessary resources and mentoring through the course of this project as well as Aya Shkedy, Dorit Granot and Nirit Alon for coordinating the Virtual 2022 ISSI, allowing us the opportunity to participate in this mind-opening program.

References

[1] Grimm, D., 2021. *How many mice and rats are used in U.S. labs? Controversial study says more than 100 million*. [online] Science.org. Available at: https://www.science.org/content/article/how-many-mice-and-rats-are-used-us-labs-controversial-study-says-more-100-million.

[2] Oracle.com. n.d. *What is Machine Learning?*. [online] Available at: https://www.oracle.com/hk/data-science/machine-learning/what-is-machine-learning.

[3] Ibm.com. 2020. *What is Supervised Learning?*. [online] Available at: https://www.ibm.com/cloud/learn/supervised-learning.

[4] Zamir, S., Arora, A., Khan, S., Munawar, H., Khan, F., Yang, M. and Shao, L., 2022. Learning Enriched Features for Fast Image Restoration and Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1-1.

[5] Wei, C., Wang, W., Yang, W. and Liu, J., 2018. *Deep Retinex Decomposition for Low-Light Enhancement*. [online] Paperswithcode.com. Available at: https://paperswithcode.com/paper/deep-retinex-decomposition-for-low-light.

[6] scikit-learn. n.d. *Decomposing signals in components (matrix factorization problems)*. [online] Available at: https://scikit-learn.org/stable/modules/decomposition.html#decompositions.

[7] Tschannen, M., Bachem, O. and Lucic, M., 2018. Recent Advances in Autoencoder-Based Representation Learning. [online] Available at: https://arxiv.org/abs/1812.05069.

[8] L. Yasenko, Y. Klyatchenko and O. Tarasenko-Klyatchenko, "Image noise reduction by denoising autoencoder," 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT), 2020, pp. 351-355, doi: 10.1109/DESSERT50317.2020.9125027.

[9] Z. Chen, C. K. Yeo, B. S. Lee and C. T. Lau, "Autoencoder-based network anomaly detection," 2018 Wireless Telecommunications Symposium (WTS), 2018, pp. 1-5, doi: 10.1109/WTS.2018.8363930.

[13] Balsam, A., Kuehne, J., Rodríguez de León, V., Sánchez Vargas S. and Kalchenko V., 2022. Bioimaging Data Generation for Machine Learning in Blender: an Unorthodox Approach.

[10] Guo, X., Liu, X., Zhu, E. and Yin, J., 2017. Deep Clustering with Convolutional Autoencoders. *Neural Information Processing*, pp.373-382.

[11] An, J. and Cho, S., 2015. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. [online] Available at: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf.

[12] H.D. Nguyen, K.P. Tran, S. Thomassey, M. Hamad,
Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management, International Journal of Information Management, Volume 57, 2021, 102282, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2020.102282.