PROSTHETIC VISION ENHANCEMENT USING DEEP REINFORCEMENT LEARNING

N. Kavitha
Assistant Professor (Sr.Gr.), Department of Computer Science and Engineering
Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India

K. Pooja & C. Logeshwari
Department of Computer Science and Engineering
Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India

DOI: doi.org/10.34293/shanlax.9789361631474.ch002

Abstract

A novel end-to-end vision processing pipeline is proposed that directly addresses the need for task-based, context-aware scene visualizations for human-centric computer mediated displays such as prosthetic vision. The proposed model aims to detect objects in a scene through a viable and scalable method for learning image filters that are adaptable to new tasks and environments. Our proposed approach is evaluated quantitatively and qualitatively using simulated prosthetic visuals and deployed on any computer-mediated visual guidance, including remote control of vehicles such as drones, mining equipment, rescue robots, etc.), over bandwidth constrained display.

I. Introduction

This section provides an overview of the proposed model for human-centric vision processing. A model for human-centric vision processing refers to a computational framework designed to mimic the human visual system and process visual information in a way that is centered around human perception and understanding. This model aims to bridge the gap between raw visual data and human-level interpretation. The model typically consists of multiple interconnected modules, each responsible for a specific aspect of vision processing. These modules may include image preprocessing, feature extraction, object recognition, scene understanding, and semantic segmentation, among others. The model's core concept revolves around feature extraction, where relevant visual features such as edges, textures, colors, and shapes are extracted from the input images. This process is inspired by the early stages of visual processing in the human brain. Once the features are extracted, the model moves on to higher-level tasks such as object recognition and scene understanding. Object recognition involves identifying and classifying objects present in the image, while scene understanding focuses on interpreting the overall context and relationships between objects. To achieve human-centric vision processing, the model takes into account various factors that influence human perception. These factors include attention mechanisms, which prioritize certain regions of the image based on their saliency or relevance. By simulating human attention, the model can focus on important details and discard irrelevant information. Additionally, the model may incorporate knowledge about human cognition and semantics. This enables it to reason about the content of the image and make inferences based on prior knowledge or common sense. Overall, a model for human-centric vision processing aims to go beyond mere image analysis and strives to understand visual data in a way that aligns with human perception and cognition. By doing so, it opens up possibilities for applications in fields such as autonomous driving, augmented reality, robotics, and healthcare. It also highlights the advantages of this model for the development of task-targeted and customizable vision processing pipelines, as well as its scalable deployment on prosthetic vision devices.

Vision processing for prosthetic vision is a field that focuses on developing computational methods to restore or enhance visual perception in individuals with visual impairments. It is based on the understanding of how the visual system works and aims to replicate or mimic its functionality through artificial means.

The human visual system consists of several interconnected stages, starting from the capture of visual information by the eyes to the interpretation of that information by the brain. In individuals with visual impairments, these stages may be disrupted or impaired, leading to a loss of vision. Vision processing for prosthetic vision aims to compensate for these impairments by processing captured visual information and presenting it to the individual in a way that they can perceive and interpret.

One of the key aspects of vision processing for prosthetic vision is image preprocessing. This involves techniques such as denoising, image filtering, and color correction to improve the quality of captured images. By reducing noise and enhancing image clarity, preprocessing helps to provide a better input for subsequent processing stages.

Feature extraction is another important component of vision processing. It involves identifying and extracting relevant visual features from the preprocessed images. These features can include edges, textures, colors, and shapes. By extracting these features, the processing system can focus on the most important aspects of the visual scene.

Image enhancement techniques are then applied to enhance the visibility of key visual elements. This can involve adjusting contrast, sharpening edges, or adapting the image to optimize visibility. The goal is to make important details more distinguishable and easier to perceive.

Object recognition algorithms play a crucial role in vision processing for prosthetic vision. These algorithms are trained to identify and classify objects present in the scene. By recognizing objects, the system can provide meaningful information to the individual about their surroundings, enabling them to navigate and interact with their environment more effectively.

The processed visual information is then mapped onto a spatial representation that can be interpreted by the user. This can involve converting the processed image into a tactile display or generating auditory cues that convey information about the scene. The output is tailored to the specific needs and capabilities of the individual and the prosthetic vision device being used.

Overall, vision processing for prosthetic vision is a multidisciplinary field that combines knowledge from computer vision, machine learning, and neuroscience. It aims to develop computational models and algorithms that can replicate or enhance the functionality of the human visual system, providing individuals with visual impairments with functional vision and improving their quality of life.

II. Background and Related Work

The proposed vision processing pipeline aims to enhance the structural features of captured photosensor data to generate a functional representation for prosthetic vision devices. This involves modifying the raw input data to improve the visibility of key surface boundaries in real images. Deep reinforcement learning can be utilized as one of the methods to achieve this enhancement.

The first step in producing a phosphene visualization is to find out the best feature among many available features. As discussed in [1], the wrapper approach is used to evaluate best feature but it evaluates only the subsets of feature it is based on the performance of specific algorithm. Under wrapper approach, the recursive feature elimination technology is used which eliminates the least important features. The challenge in recursive feature elimination method is it is not suitable for the high dimensional datasets. The next step is to produce the phosphene visualization using the best feature map. In [2], The proposed method classifies the spectral images using convolutional neural network it captures the relevant spatial and spectral features, the accurate predictions are made. For Hyperspectral images the network called convolutional neural network is more preferred. It considers both spectral and spatial features at different scales and classifies the most accurate. The multiscale feature is less interpretable it limits the ability to gain insights.

In [3], The Psychological sensors captures the human activity like muscle activity, heart rate, skin conductance etc. Which acts as a raw data. The transfer Learning is employed to automatically capture the relevant features from the raw data and the deep learning neural network LSTM and CNN on combination identifies the complex patterns and recognize accurate predictions and improves the performance. Due to the integration of sensor, transfer learning and deep neural network the complexity and the cost is not efficient and the user feels discomfort due to the continuous monitoring and the sensors need to be calibrated properly. In [4], the image exposure is enhanced using the neural network and loss function. Some images may be less exposure which means it might be too bright or it might be too dark so it affects the visualization of users so those images exposure need to be enhanced correctly it is done by the neural network and appropriate loss function. During the enhancement process the unintended distortions are generated it affects the fidelity of original data.

In [5], Binary threshold is measured and binary image is converted to phosphene with respect to binary threshold. In Otsu's method optimal threshold is taken and with respect to optimal threshold, the binary image is converted to phosphene image. Otsu's method is efficient compared to other methods as in other's the nature of the image is first is considered but here it can be applied to all types of images. In [6], It involves the integration of transfer learning approaches over the deep reinforcement learning. The Deep Reinforcement learning make the sequential correct decisions based on the feedback. The transfer learning is the effective learning method from the pre-trained model and it classifies automatically. On integrating the both, the classification is done accurately and appropriately the relevant features are predicted and classified, the performance is improved.

In [7], the YOLO-v5 is used as an object detection algorithm which detects the objects in the images, it extracts the features that is the high dimensional features. The feature map of underground water image acts as a state s, visual enhancement algorithm acts as an action a, based on the object detection score the reward r is evaluated. The reward guides the agent to select a visual enhancement algorithm. The evaluation metrics like precision, recall, accuracy, F-score are used to measure the performance and to improve the performance. In [8], the information is gathered from to sources one from human gaze and other from machine vision. The Human gaze involves tracking, monitoring and analyzing the direction of person gaze. Machine vision it is a computer vision technology it processes the visual information using machine learning process. The feature is extracted from these two sources and the relevant fusion algorithm is employed and the model is trained to predict the intended locomotion mode.

In [9], the features may be discrete or continuous, for the continuous data the feature space is infinite so the infinite number of features is selected using graph-based feature filtering approach. Graphs can represent the relationships or dependencies between entities. The feature filtering is proposed it is a type of feature selection in which features are assessed and filtered based on certain criteria without involving a learning algorithm, it is likely to derive from graph-based structure. In [10], The FTIR spectral resolution enhancement technology is proposed in order to enhance the resolution and to reveal the fine spectral features it is done by employing the minimization algorithm. Here, the enhancement leads to make the spectrum more sensitive to noise.

In [11], Ultrasound imaging technique is used to visualize the internal structures it is applicable to medical field in order to visualize the internal features of our body for generation of phosphene image the visualization of internal feature is not as mandatory. In [12], The Deep Reinforcement Learning is applied directly to the image and the best feature is selected based on utility by applying utility function. The challenge we face in utility function is it could provide only the appropriate feature that has highest utility whereas reward function enables the learning of agent through the rewards and feedback.

In [13], Here Reinforcement Learning is applied using locomotion skills, based on the movement here the agent gets trained and learned. Based on the locomotion activity, the RL algorithm is chosen. As it requires a large number of interactions with environment in order to learn effective locomotion skills, the sample inefficiency is problematic. In [14], Batch RL is applied to train the agent and to make informative decisions to attain a desirable outcome but it is applicable only to fixed datasets, it undergoes distributional shift which does not cover the entire state-action space whereas a RL is applicable to all datasets it covers all the state-action spaces.

In [15], the features are extracted using filter and wrapper methods, the filter methods use statistical measures to evaluate the relevant features whereas wrapper methods use specific machine learning model to evaluate the subset of features. In Filter method, the features are evaluated and ranked using certain metrics like correlation, mutual information and statistical tests. This method is useful for high dimensional data.In[16], The Feature Selection selects the strongly relevant features and eliminate the weekly relevant features, in this, the extra classifier is applied to the irrelevant feature which again process the strongly relevant features and weekly relevant features.

In [17], the instances involved belonged to multiple classes so it involves multi-labelling it is done by employing correlation measure. It involves analyzing the correlation structure between different labels it is done using statistical measure to capture dependencies between labels. The performance is evaluated using evaluation metrics includes precision, recall, f-score or Hamming loss. The selected feature is used in conjunction with multi-label classifiers. In [18], The relevant features are automatically selected from the dataset using unsupervised learning. The technique Metric Fusion is applied to the relevant features which fuses different metrics according to the data. Following this, the novel low-rank approximation is applied in order to retain the important information in the data. As it involves combining the multiple metrics it makes the computational cost high. In ref 19, Novel Feature selection method is imposed to deal with uncertainty in relevant features. This method quantifies the uncertainty during the selection process. This method leads to overfitting and the computational cost is also not efficient. In [20], The multiple features are ranked using various ranking techniques like using statistical measures, recursive feature elimination, Lasso etc…Then the ranked features are clustered together using various clustering techniques like correlation-based clustering, hierarchical clustering and dimensionality reduction techniques. To asses the quality of clusters the cluster evaluation metrics like silhouette score, Davies Bouldin is used, then the most important feature is identified. In [21], The unsupervised dual Learning Feature it captures the relevant features and instances without using labeled data. This method is useful for the data with high dimensions, its objective is to enhance the interpretability.

Section III discusses about proposed method by overcoming all the limitations faced in the above said methods.

III. Proposed Work

The process of producing phosphene visualization of an image is done using VGG16 (Fig.1) to extract various features. Then it the set of features is passed as input to reinforcement learning model to select the best feature based on cumulative rewards.

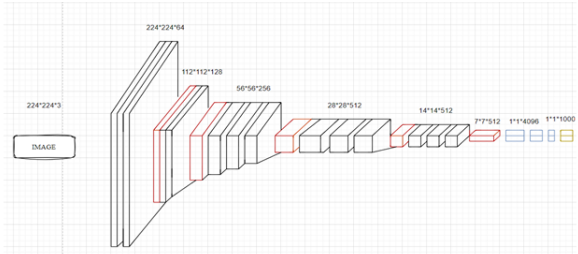

Fig.1 The architecture of VGG16, whose input is an image of dimension 224 x 224 x 3

The system flow involves, passing an input image to VGG16 (Section III-A) to extract features and passing those features to the reinforcement learning model to select the best feature. Then we apply binary thresholding to convert it to a phosphene visualization.

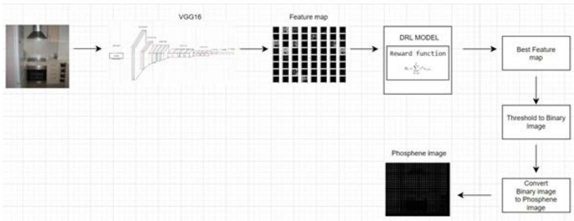

Fig.2 System flow diagram showing the entire process involved in generating the phosphene visualization. The input is an image of size 224 x 224 x 3. VGG16 extacts various features and the DRL model extracts the best feature out of it. Then thresholding is applied to produce binary image, which is then converted to phosphene visualization.

A. *VGG16 Architecture*

The VGG16 model is a popular convolutional neural network architecture that is commonly used for image classification tasks. It consists of 16 layers, including convolutional layers, pooling layers, and fully connected layers. Here are the features (layers) of the VGG16 model:

The input layer represents the input image. The first convolutional layer with 64 filters. The second convolutional layer with 64 filters. Max pooling layer after the first two convolutional layers. The third convolutional layer with 128 filters.

The fourth convolutional layer with 128 filters. Max pooling layer after the second two convolutional layers. The fifth convolutional layer with 256 filters. The sixth convolutional layer with 256 filters.

The seventh convolutional layer with 256 filters. Max pooling layer after the third three convolutional layers. The eighth convolutional layer with 512 filters. The ninth convolutional layer with 512 filters.

The tenth convolutional layer with 512 filters. Max pooling layer after the fourth three convolutional layers. The eleventh convolutional layer with 512 filters. The twelfth convolutional layer with 512 filters. The thirteenth convolutional layer with 512 filters. Max pooling layer after the fifth three convolutional layers. The first fully connected layer with 4096 units. The second fully connected layer with 4096 units. The final fully connected layer with the number of units equal to the number of classes in the classification task.

These are the features (layers) of the VGG16 model, which collectively enable it to learn hierarchical representations of images and perform feature extraction. The above said architecture is shown in fig.1.

B. Flow of the proposed system

1. Preprocessing: The raw photosensor data is preprocessed to remove noise and artifacts. This may involve techniques such as denoising, image filtering, and normalization. The goal is to obtain a clean input for subsequent processing stages.
2. Feature Extraction: In this stage, deep learning techniques can be employed to extract relevant features from the preprocessed data. VGG16 are commonly used for this purpose. These networks are trained on large datasets to learn discriminative features that are important for object recognition and boundary detection.
3. Deep Reinforcement Learning: Deep reinforcement learning can be integrated into the pipeline to enhance the visibility of key surface boundaries. This involves training a reinforcement learning agent to make decisions on modifying the input data based on a reward signal. The agent learns through trial and error to apply modifications that result in improved boundary visibility. The reward function given below is used to calculate the cumulative reward and select the best feature from the obtained features.

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k}.$$

R_t- total discounted reward

γ: The discount factor.

k: Time period from 0 to T

4. Boundary Enhancement: The modifications suggested by the reinforcement learning agent are applied to the preprocessed data to enhance the visibility of key surface boundaries. This can be achieved through various techniques such as contrast enhancement, edge sharpening, and adaptive filtering. The goal is to make the boundaries more prominent and distinguishable from the surrounding areas. For our case, contrast enhancement gives clear difference between the surfaces.
5. Post-processing: After boundary enhancement, additional post-processing steps can be performed to further refine the functional representation. This may include noise reduction, smoothing, and contour extraction. The final output is a processed image that highlights the structural features and boundaries of interest.

By incorporating deep reinforcement learning into the vision processing pipeline, the system can learn to adaptively modify the input data to enhance key surface boundaries. This approach allows for personalized adjustments based on individual preferences and visual needs, leading to improved functionality and usability of prosthetic vision devices.

C. Training the Deep Reinforcement Learning model

Training the DRL model involves the process of building a reinforcement learning model to train the Deep Learning model. Training the reinforcement model results in selecting the best feature map from a list of feature maps to generate the target phosphene visualization. The reinforcement learning agent gains reward in each episode. As observed, the cumulative reward keeps on increasing.

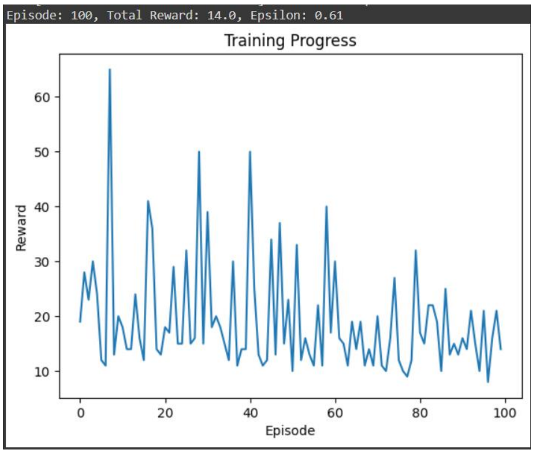

Fig. 3 Rewards gained in each episode while training the DRL model

Results

This process of producing phosphene visualization of an image to enhance prosthetic vision was first done using built-in Image package as shown in Fig.4 which produced a clear differentiation between the boundaries. This was more effective in detecting indoor objects.

Fig. 4 Showing results obtained by using built-in Image package, differentiating the boundaries clearly between the objects in the scene. Left is the input and to the right is the corresponding phosphene visualization.

References

1. Recursive Feature Elimination Based Feature Selection in Modulation Classification for Mimo Systems, Zhou Shuai, LI Tao, and LI Yongzhao, *Chinese Journal of Electronics Vol.32, No.4, July 2023.*
2. Multiscale Spatial-Spectral Feature Extraction Network for Hyperspectral Image ClassificationZhen Ye; Cuiling Li; Qingxin Liu; Lin Bai; James E. Fowler, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing Vol.15, June 2022*
3. Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting Physiological Sensors and Feature Learning, Nurul Amin Choudhury; Badal Soni, *IEEE Sensors Letters Vol.7, Issue:11, November 2023*
4. Deep Perceptual Image Enhancement Network for Exposure Restoration, Karen Panetta; Shreyas Kamath K. M.; Shishir Paramathma Rao; Sos S. Agaian, *IEEE Transactions on Cybernetics Vol.53, Issue: 7, July 2023*
5. Evaluating Current Density Modeling of Non-Invasive Eye and Brain Electrical Stimulation Using Phosphene Thresholds, B. A. Sabel, A. Kresinsky, L. Cárdenas-Morales, J. Haueisen, Member, IEEE, A. Hunold, M. Dannhauer, and A. Antal, *IEEE transactions on Neural Systems and Rehabilitation Engineering, vol. 29, 2021, IEEE Access*.
6. Transfer Learning in Deep Reinforcement Learning: A Survey, Zhuangdi Zhu; Kaixiang Lin; Anil K. Jain; Jiayu Zhou, *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.45, Issue: 11, 01 November 2023*
7. A Reinforcement Learning Paradigm of Configuring Visual Enhancement for Object Detection in Underwater Scenes, Hao Wang; Shixin Sun; Xiao Bai; Jian Wang; Peng Ren, *IEEE Journal of Oceanic Engineering Vol.48, Issue: 2, April 2023*
8. Fusion of Human Gaze and Machine Vision for Predicting Intended Locomotion Mode, Minhan Li; Boxuan Zhong; Edgar Lobaton; He Huang, *IEEE Transactions on Neural Systems and Rehabilitation Engineering Vol.30*
9. Infinite Feature Selection: A Graph-based Feature Filtering Approach, Giorgio Roffo, Simone Melzi, Member, IEEE, Umberto Castellani, Alessandro Vinciarelli, Member,

IEEE, and Marco Cristani, Member, IEEE, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 43, NO. 12, DECEMBER 2021.*

10. Flexible FTIR Spectral Imaging Enhancement for Industrial Robot Infrared Vision Sensing, Tingting Liu, Student Member, IEEE, Hai Liu, Member, IEEE, You-Fu Li, Senior Member, IEEE Zengzhao Chen, Zhaoli Zhang, Member, IEEE, and Sannyuya Liu, *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 16, NO. 1, JANUARY 2020.*
11. Image Quality Enhancement Using a Deep Neural Network for Plane Wave Medical Ultrasound Imaging, Yanxing Qi, Yi Guo and Yuanyuan Wang, Senior Member, IEEE, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2020.*
12. Deep reinforcement learning for task-based feature learning in prosthetic vision, Jack White1, Tatiana Kameneva1,2 and Chris McCarthy1. *Benchmarking deep reinforcement learning for continuous control, International Conference on Machine Learning, vol. 48, pp. 1329- 1338, 2020.*
13. DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning, Xue Bin Peng and Glen Berseth, University of British Columbia Kangkang Yin, Michiel Van de Panne. *ACM Transactions on Graphics, Vol. 36, No. 4, Article 41. Publication date: July 2021*
14. Efficient Batch-Mode Reinforcement Learning Using Extreme Learning Machines, Jiahang Liu, Lei Zuo, Xin Xu, Senior Member, IEEE, Xinglong Zhang, Junkai Ren, Qiang Fang, and Xinwang Liu, *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 51, NO. 6, JUNE 2021.*
15. Metric and Accuracy Ranked Feature Inclusion: Hybrids of Filter and Wrapper Feature Selection Approaches, G. S. THEJAS 1, (Member, IEEE), RAMESHWAR GARG2, S. S. IYENGAR 3, (Life Fellow, IEEE), N. R. SUNITHA4, (Member, IEEE), PRAJWAL BADRINATH3, AND SHASANK CHENNUPATI 5, *MultiDsicilplinary, Open Access Journal, IEEE Access, vol. 9,2021, Publication date: September 13, 2021.*
16. Feature Selection Boosted by Unselected Features, Wei Zheng , Shuo Chen , Zhenyong Fu , Fa Zhu , Hui Yan , and Jian Yang , Member, IEEE, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 33, NO. 9, SEPTEMBER 2022.*
17. Online Multi-Label Streaming Feature Selection With Label Correlation, Dianlong You , Member, IEEE, Yang Wang, Jiawei Xiao, Yaojin Lin, Maosheng Pan, Zhen Chen, Limin Shen, Member, IEEE, and Xindong Wu, Fellow, IEEE, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 35, NO. 3, MARCH 2023.*

18. Unsupervised Feature Selection via Metric Fusion and Novel Low-Rank Approximation, Yin Long 1, Liang Chen1, Linfeng Li1, And Rong Shi 2, *MultiDsicilplinary, Open Access Journal, IEEE Access, vol. 10,2022.*
19. A Novel Feature Selection Method for Uncertain Features: An Application to the Prediction of Pro-Anti-Longevity Genes, Pablo Nascimento da Silva, Alexandre Plastino, Fabio Fabris, and Alex A. Freitas, *IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 18, NO. 6, NOVEMBER/DECEMBER 2021.*
20. Combining Multiple Feature-Ranking Techniques and Clustering of Variables for Feature Selection, ANWAR UL HAQ 1, Defu Zhang 1, (Member, IEEE), He Peng1, And Sami Ur Rahman, *MultiDsicilplinary, Open Access Journal, IEEE Access, vol. 7,2020.*
21. Unsupervised Dual Learning for Feature and Instance Selection, LIANG DU 1,3, (Member, IEEE), Xin Ren1, Peng Zhou (Member, IEEE), AND Zhiguo Hu1, *MultiDsicilplinary, Open Access Journal, IEEE Access, vol. 7,2020, Publication date: September 18, 2020.*
22. A Hybrid Feature Method for Handling Redundant Features in a Sentinel-2 Multidate Image for Mapping Parthenium Weed, Zolo Kiala, Onisimo Mutanga, John Odindi, Serestina Viriri, Member, IEEE, and Mbulisi Sibanda, *IEEE journal of selected topics in applied earth observations and remote sensing, vol. 13, 2020.*