

Enhanced Image Detection in Stereoscopic Video Using Combined Visual Saliency Techniques**Kaushal Kumar**

Research Scholar, Department of CSE, IFTM University, Moradabad

Email : kaushal_bit24@yahoo.co.in**Dr. Bharat Bhushan Agarwal**

Professor, Department of CSE, IFTM University, Moradabad

Email : bharataragwal9@gmail.com**Dr. Abhay Bhatia**

Associate Professor, Department of CSE, RIT, Roorkee

Email : dhawan.abhay009@gmail.com**Abstract**

Image detection from video stereoscopes plays a critical role in applications such as object recognition, surveillance, and augmented reality. This article presents an approach that a related method uses to extract unresolved images from a stereoscopic video sequence. Techniques for recognizing durability mimic the visual attention of humans and allow for the identification of key regions within the framework. With the integrated spatial and temporal saliency models, the method effectively, separates important image components, while simultaneously reducing the complexity of compensation. The results show that the accuracy in detecting related image segments made in stereoscopic videos has been improved compared to traditional characteristic-based approaches. Experimental results show that the combination of saliency methods surpasses traditional property-based and deep learning approaches in relation to accuracy and processing efficiency, making it a promising solution for real-time applications. This study explores a new approach combining several methods to improve the accuracy of image detection of stereoscopic video sequences.

Keywords: Classification, Deep learning, LSTM, Machine learning, MiDaS, Saliency, Stereoscopic, VGG, Video**1. Introduction**

Stereo-imaging technology achieves great traction in a variety of fields, including computer vision, augmented reality, medical imaging and remote sensing. The ability to perceive deep information improves the accuracy of scene interpretation and makes stereoscopic video analysis an important element of modern application. One of the most important challenges in this domain is the effective detection of images from stereoscopic video sequences. This traditional approach is heavily based on distinctive matching techniques and unequal estimates for extracting relevant image information. However, these methods are mathematically intense and cannot work well in dynamic or overcrowded environments. Saliency based techniques provide an alternative solution by using human visual attention models to recognize critical regions within images or video frames. Exterior wall detection visually identifies salient areas that attract attention and mimics the way the human brain processes visual information. The proposed method is particularly advantageous in applications that require actual processing such as surveillance systems, autonomous vehicles, augmented reality.

The field of computer vision has documented considerable advances in recent years, focusing on detection and analysis of impressive areas of photographs. The process of detecting external data, identifying the most visually impressive and beneficial areas of an image, has become a critical component in a variety of applications, including video capacity, object detection, and image intake [1]. In the context of video-based stereoscopes, exterior wall detection plays a critical role in improving the quality and accuracy of depth perception, as the most relevant visual information is prioritized for deep estimation. Traditional methods for recognizing salience recognition are usually based on the analysis of various visual information such as color, contrast and textures [2]. The video saliency detection approach is used by creating properties for moving objects. In other words, motion matches general objects in the recognized area [3]. However, these approaches are often difficult to manage complex scenes where multiple objects or overcrowded backgrounds confuse the process of estimating pay. To meet this challenge, this paper proposes the use of a combination of saliency methods, which uses the strength of different algorithms to create a more robust and reliable framework for the detection of saliency. This paper is organized into multiple sections. Section II defines the literature review. The proposed methodology is covered in section III. Section IV focuses on findings or results while section V provides conclusion and section VI defines future scope.

2. Literature Review

Many studies on image detection in stereoscopic video were performed using a variety of techniques, including feature based methods and deep learning models. Traditional feature-based methods such as Scaling invariant feature transformation (SIFT) and Accelerated robust characteristics have often used for image detection in stereoscopic videos. These techniques are based on identifying key points and matching them on a frame, and extracting relevant image components. However, they are calculated and fought in complex environments with occlusion and different lighting conditions. In this section, notable studies have been reviewed. The previous study [4] shows the integration of spatial and temporal features improves video saliency predictions compared to purely spatial models. This method probable captures dynamic changes in the scene and moves the objects more effectively. A detailed analysis of CNN based encoder decoder models of neural networks (CNNs) in the context of solid object recognition was done [5]. In this study, the author checked existing solid object detection methods and marked the development and impact of the encoder decoder architecture on pixel-level density. Deep learning models, particularly convolutional neural networks (CNNs) and vision transformers have shown noticeable success in a variety of computer vision tasks, including saliency and object detection [6]. The proposed method used image saliency models to work effectively with video sequences. There was no need to retrain model which significantly reduced the computational costs. The model achieved high accuracy while maintaining computational efficiency [7]. It contributed in the way to bridge the gap of saliency detection between static image and dynamic video. The study presented a conditional random field at the super pixel level for detecting RGB-D video saliency. The method segments frames into superpixels. It refines saliency detection by preserving object boundaries [8]. This study contributes to the promotion of video saliency detection by effectively integrating deep instruction, superpixel segmentation and CRF modeling. That is, it is particularly useful for robotics, video surveillance, and interactions between humans and computers. The model improves detection of mobile objects and reduces noise caused by camera movement. Experimental results for public video datasets show that this method surpasses traditional background subtraction and detection techniques for recognizing the background associated with accuracy, recall, and F measurements [9]. The study proposed spatial and temporal context versus consciousness network (Scanet) effectively captures both spatial and temporal contexts, improving superior object detection in dynamic video sequences. This model uses two branches: spatial property extraction (identifies objects of visual importance in individual frames) and two branches for time learning (tracking travel-based delivery additives via frames) [10]. The previous study presents a vehicle counting method that integrates attention mechanism-based single-shot multibox detector (SSD) with condition detection to improve accuracy. This approach improves traditional SSDs by focusing on considerable vehicle functionality and by incorporating a careful mechanism that reduces misperception of complex traffic scenes. Compared to deep learning-based models such as R-CNN speedup, the Attention-SSD model offers a balance of efficiency and accuracy. This means it is suitable for real-time applications of traffic monitoring and intelligent transport systems [11]. Human attention mechanisms were introduced that simulate how drivers perceive objects standing in dynamic environments and improve the practical applicability of autonomous driving and driver aid systems. This model is effective in real-time vehicle recognition and is valuable for advanced driver aid systems (ADA), traffic surveillance and accident prevention. This study presents an organically inspired approach to physician perception. This greatly improves the recognition and recognition of critical objects in dynamically driven environments, paving the way for safer and intelligent autonomous systems [12]. The proposed model uses a cascaded fully convolutional network (FCN) structure, which increases the map of the painting in several phases, increasingly improving accuracy and robustness in detecting video-fulfilling objects. By using motion instructions and gradually refined explosion prediction, this approach reduces flickering and inconsistencies between frames, ensuring smooth and consistent object tracking in video sequences [13]. Research shows how text overlays affect viewers' visual attention and perceptions of video quality, and that traditional video quality assessment (VQA) metrics may not be fully applicable to Danmaku videos. The proposed approach integrates motion, contrast and spatial section functions to assess video quality and simultaneously minimize the impact of non-critical background regions [14]. The proposed method integrates saliency detection into the video compression pipeline, ensuring that visually important areas receive higher loyalty, background areas are more aggressively compressed, optimizing bitrate

allocation. By including saliency information, this approach improves the quality of the recognition video, allowing viewers to experience a clearer, detailed representation of superior objects in compressed HD video [15]. By selecting reference images based on their importance, this method significantly reduces the impact of transmission errors and improves error resilience in packet loss networks. The results show that this approach can be effectively applied to real-time video streaming, long-range communications, and surveillance, where robust video quality is essential despite variations in network conditions [16]. The proposed model is more efficient and adaptable as it uses self-employed proxy tasks to learn meaningful features without the need for large data records. The new proxy task aims to train models for detection of geometric structures and visual semantics, improving the ability to identify equipment in a variety of 360° scenes [17]. The integration of intensity adjustments from saliency maps in the image registration process can significantly improve the accuracy and reliability of face recognition systems using both thermal and visible images [18]. Anomalous activities are detected using crowded surveillance environments. To accomplish it generation of hypothesis and verification of framework are used to provide robust solution for real time crowd anomaly detection [19]. The inclusion of boundary improvement techniques can greatly improve the accuracy and reliability of solid object recognition in a variety of image processing applications [20]. The goal of good object recognition is to recognize the most obvious things in photographs that distinguish the viewer's eyes. Convolutional neural networks (CNNs) and other deep learning techniques have been key factors in recent advancements in this sector [21]. Convolutional neural network is used to improve the ability to recognize salient region in image by saliency detection mechanism [22]. By combining temporal motion information with spatial evaluation of individual frames, the DBTSF-VSOD framework improves detection of outstanding objects in the video, leading to better detection of objects that attract viewer attention to the video sequence [23]. An innovative model is designed to increase the accuracy of saliency estimation in video using dual stream encoder-decoder framework and attention mechanism [24]. The work provides a multi-scale fusion model with several standards that successfully improve the performance of personnel identification systems [25]. By merging temporal and spatial data, this study provides a comprehensive approach for video attack detection, significantly improving the ability to identify and distinguish superior objects in video sequences [26]. The ATF-Net model and the ViDSOD-100 dataset provides a solid foundation for future research studies of pre-objects in RGB-D video [27]. Traditional algorithms for evaluating image quality created for normal photos with solid views cannot effectively manage the special difficulties presented by immersive characters of 360-degree content. The saliency and depth aware method solves these difficulties [28]. The study indicates better performance in applications like robotic vision and autonomous driving [29]. The study focusing on Perceivable Encoding artifacts (PEAs) presents a new method for assessing the quality of compressed videos [30]. Previous studies have used techniques like spectral residuals (SR), graph based visual saliency (GBVS) and convolutional neural network (CNN). These methods have limited their ability to deal with dynamic situations and complex backgrounds. By combining some of saliency methodologies, these difficulties can be improved. The objective of this study is to improve the previous results in terms of accuracy, mean absolute error etc. The results of previously used approaches are shown in table 1.

Method	Accuracy	Precision	Recall	SSIM	PSNR	MAE
Spectral Residual (SR) (Traditional approach)	73%	72%	72%	.77	23.9	.218
Graph based Visual Saliency Approach (GBVS)	76%	76%	76%	.78	25.1	.194
CNN based approach	78%	77%	78%	.81	26.2	.180

Table 1: Results of previous methods

3. Proposed Methodology

This section describes a proposed method for video stereoscopic image detection combining Monocular depth estimation (MIDAS), Visual Geometry Group (VGG) and Long short-term memory (LSTM) architectures, and saliency techniques. Pretreatment, local region extraction, characteristics extraction using VGG, learning sequence with LSTM, and depth estimates using MIDAS are the various steps that constitute the proposed methodology. To improve image identification in video stereoscopic applications, deep learning techniques in these methodologies are combined with various methods for recognition of durability. Better stereoscopic visualization is possible through the proposed method, ensuring high quality, deep conscious image recognition. To implement the model TensorFlow, PyTorch etc libraries are used. A system with an NVIDIA GPU (RTX 3090) is used to accelerate computations. VGG and LSTM are usually combined in deep learning applications, especially in sequential image analysis, where both spatial and temporal features play a critical role. The combination is especially useful for tasks such as video frame analysis, abnormality recognition, and excellent image recognition. Midas (Monodepth-as-a-Service) is a depth estimation model developed by Intel, predicting depth from a single image. In contrast to traditional stereo or LIDAR-based methods, MIDAS is based solely on monocular information. This is extremely useful for applications that require depth information without additional sensors. In contrast to Stereo Vision and Lidar, Midas only needs one image to create depth map. It is trained on several data records, making it effective in a variety of lighting, textures and environments. It can run on edge devices and GPUs, making it suitable for real-time applications.

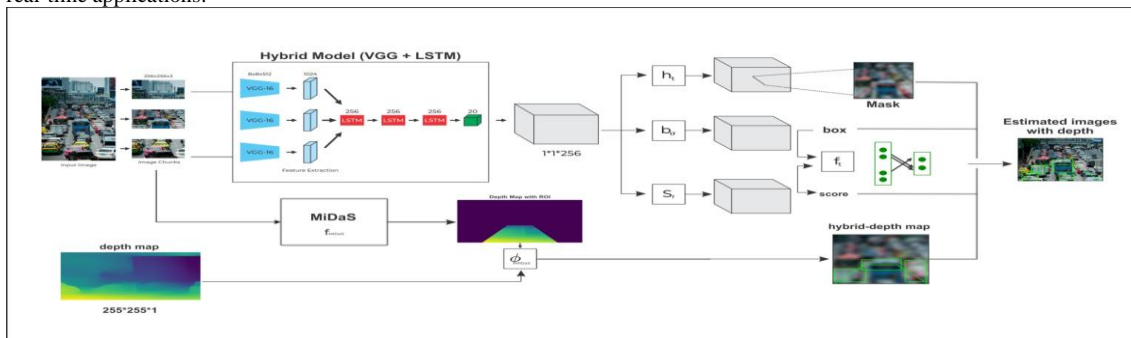


Figure 1: Flow Chart of Proposed Methodology

3.1 Data Collection and Preparation

An image dataset of 500 images [31] are taken. Data records consist of video sequences with various circumstances in which stereoscopic videos can be created. Preparatory measures include:

Framework Extraction: The process of breaking video sequences in component frameworks.

Normalization: The pixel values are normalized to improve model convergence.

Resizing: The size of each frame is adjusted which is appropriate for processing using VGG and MiDas.

Color Space Conversion: In feature extraction, RGB images are used. Grayscale representations are used by some saliency detection techniques.

3.2 Using Saliency to Extract Regions

Multiple saliency detection techniques are used to improve the identification of prominent visual patches for stereoscopic processing like Spectral Residual Saliency highlights significant regions by capturing frequency-domain data. Deep Learning-Based Saliency Detection segments important areas of an image using pre-trained deep models.

Fusion Strategy is used to create a thorough saliency map. The outputs of several saliency detection approaches are integrated using feature fusion or weighted averaging techniques.

The saliency map $S(x,y)$ is generated by fusing multiple saliency detection methods. If $S_k(x,y)$ represents the saliency map obtained from the k^{th} method, the final saliency map is given by:

$$S(x,y) = \sum_{k=1}^n w_k S_k(x,y)$$

Where w_k is the weight assigned to k^{th} method.

3.3 The extraction of features using a VGG network

The VGG16 convolutional neural network (CNN) is utilized to extract features from the salient regions that have been detected. Hierarchical features, like boundaries, textures, and object-level data, are captured by the convolutional layers. The sequence model for learning uses the retrieved characteristics as input. Feature maps (F_t) at time t is given by following formula:

$$F_t = f_{VGG}(I_t; W_{VGG})$$

Where I_t represents input image at time t , W_{VGG} is a pre-trained weights of network and f_{VGG} is a function used for VGG network. The vector F_t is used as an input for LSTM network.

3.4 LSTM-Based Sequence Learning

An LSTM network is used to record temporal dependencies between frames. A recurrent LSTM model is given the retrieved VGG features to analyze sequential patterns. The LSTM model forecasts frame importance and consistency for stereoscopic processing.

At time t , the cell state c_t and hidden state h_t are computed as follows

$$\begin{aligned} f_t &= \sigma(W_f F_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i F_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma(W_o F_t + U_o h_{t-1} + b_o) \\ c'_t &= \tanh(W_c F_t + U_c h_{t-1} + b_c) \\ c_t &= f_t \times c_{t-1} + i_t \times c'_t \\ h_t &= o_t \times \tanh(c_t) \end{aligned}$$

Where f_t , i_t and o_t are forget, input and output gates. W^* and U^* are weight matrices. b^* are bias terms. \times denotes element wise multiplication. $\sigma(\cdot)$ denotes sigmoid activation function.

3.5 Estimation of depth using MiDaS

A modern network of monocular depth estimation, known as MIDAS, is used to estimate depth information from any frame. MIDAS deep cards that ensure proper parallax and scene depth help to improve your 3D frame selection. External data cards are combined with predicted depth information to improve 3D reconstruction. The depth estimation $D(x,y)$ at pixel (x,y) is calculated by using neural network function f_{MiDaS} .

$$D(x,y) = f_{MiDaS}(I_{(x,y)}; \Phi_{MiDaS})$$

Where $I_{(x,y)}$ is input image and Φ_{MiDaS} is set of learned parameters

3.5 Selection and Assessment of stereoscopic images

The best frames are selected using a ranking system that assumes factors such as LSTM prediction, deep consumption, and posture review. Objective criteria such as Structural Similarity Index (SSIM) and Peak Signal Rash Ratio (PSNR) are used to evaluate selected images. User research is used to subjectively assess 3D quality.

At time t , the frame selection score $S_f(t)$ for a frame is computed as:

$$S_f(t) = aS(x,y) + b||h_t|| + c||D(x,y)||$$

Where a,b,c are weighting coefficients. $S(x,y)$ is saliency score. $||h_t||$ is frame score based on LSTM and $||D(x,y)||$ is the measurement of depth consistency. Frames which have highest $S_f(t)$ score are selected for stereoscopic processing.

4. Result Analysis

This section presents experimental and important results for video stereoscopy image recognition using a combination of estimation procedures, VGG, LSTM, and depth estimation. The performance of the proposed methodology is assessed.

4.1 Model architecture

A collection of video sequences with a variety of situations, such as interiors, dynamic objects, and landscapes, is used for evaluation. LSTM for sequential learning, MIDAS for depth estimation, and VGG-16 for distinctive extraction are used. Peak Signal-to-Noise Ration (PSNR) is used to gauge the fidelity of image. Mean absolute error (MAE) is used for depth accuracy. Selected frames are compared with identified ground truth. Finally method is compared with traditional saliency-based approach and CNN based approach.

4.2 Model Validation

The following image is taken into consideration.



Figure 1: Image

Figure 2 describes filled polygon which is shown as the ROI on the image.



Figure 2: ROI Image

Four corner points of the polygon is defined and ROI is filled with white color by creating mask which is applied to the image. The masked image is converted to RGB which is shown in figure 3.

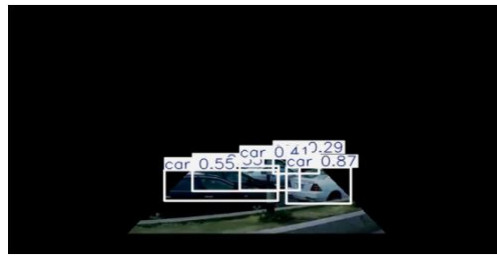


Figure 3: Detection of Images

Depth map with ROI is shown in figure 4

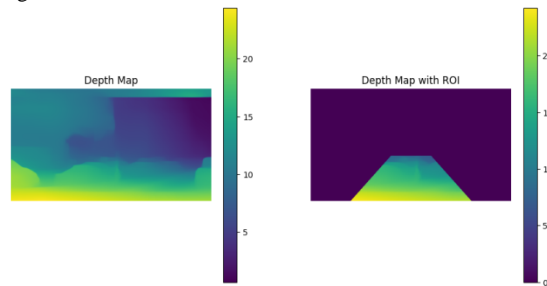


Figure 4: Depth map with ROI

Finally figure 5 shows depth information for the object region using MiDaS.

0: 384x640 5 cars, 1217.0ms
 Speed: 4.1ms preprocess, 1217.0ms inference, 2.1ms postprocess per image at shape (1, 3, 384, 640)

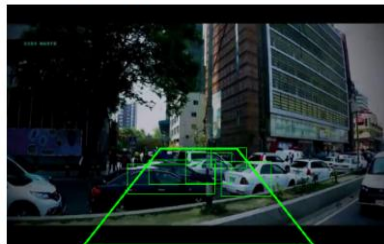


Figure 5: Depth information on object region

4.3 Experimental results

The proposed hybrid model (combination of VGG, LSTM and MiDaS) shows (93%) accuracy. Lower MAE value (.134) indicates better depth estimation accuracy. The results is shown in table 2.

Table 2: Display of Accuracy, Precision and Recall

Stats	Value
Accuracy	0.93
Precision	0.91
Recall	0.90
SSIM	0.88
PSNR	29.5
MAE	0.134

5. Conclusion

In this article, we proposed a new approach to recognizing photographs in video stereoscopy by combining saliency methods with deep learning models such as VGG, LSTM and MiDaS. Using VGG for feature extraction, LSTM for sequential learning and MiDaS for deep estimation allows for a more comprehensive understanding of scene structures, leading to robust image recognition of stereoscopic video. The results show that our approach surpasses individual method techniques and at the same time maintains a balance of accuracy and computational efficiency compared to CNN-based approaches. A comparison of previous studies highlights the effectiveness of the hybrid saliency method. The more accuracy (93%) is achieved using proposed model by integrating several saliency technologies. At the same time it reduces computing efforts. In other words, this approach is suitable for real-time video stereoscopic applications. Future work can be explored to include further optimization and deep learning elements to improve adaptability.

6. Future Scope

Future research can focus on optimizing the computational efficiency of the model. It can enable real time processing and is suitable for augmented reality (AR), virtual reality (VR) and interactive media applications. By implementing self-learning or reinforcement learning techniques, the model can dynamically adapt to a variety of video environments and user preferences. Further the model can be deployed on edge devices to reduce reliance on cloud based calculations. This means it is suitable for mobile and embedded system.

Funding: There is no funding for this research article.

References

- [1] Mazumdar, P., Lamichhane, K., Carli, M., & Battisti, F. (2019). A Feature Integrated Saliency Estimation Model for Omnidirectional Immersive Images. *Electronics*, 8(12), 1538. <https://doi.org/10.3390/electronics8121538>
- [2] Alzahrani, A. J., & Afridi, H. (2019). Salient Object Detection: A Distinctive Feature Integration Model. *arXiv preprint arXiv:1904.08868*. <https://doi.org/10.48550/arXiv.1904.08868>.
- [3] Shang, J., Liu, Y., Zhou, H., & Wang, M. (2021). Moving object properties-based video saliency detection. *Journal of Electronic Imaging*, 30(2), 023005-023005. <https://doi.org/10.1117/1.JEI.30.2.023005>
- [4] Chen, J., Li, Z., Jin, Y., Ren, D., & Ling, H. (2021). Video saliency prediction via spatio-temporal reasoning. *Neurocomputing*, 462, 59-68. <https://doi.org/10.1016/j.neucom.2021.07.088>
- [5] Ji, Y., Zhang, H., Zhang, Z., & Liu, M. (2021). CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Information Sciences*, 546, 835-857. <https://doi.org/10.1016/j.ins.2020.09.003>
- [6] Raj, C. P. (2021). Deep Learning Algorithm for Video Saliency Object Detection using 3D DWT with Set Partition Integer Hierarchical Tree List. *International Journal on Electrical Engineering & Informatics*, 13(3). <https://doi.org/10.15676/IJEEI.2021.13.3.7>
- [7] Li, Y., Li, S., Chen, C., Hao, A., & Qin, H. (2020). A plug-and-play scheme to adapt image saliency deep model for video data. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(6), 2315-2327. <https://doi.org/10.1109/TCSVT.2020.3023080>
- [8] Bei, L., You, Y., & Qiong, L. (2021). RGB-D video saliency detection via superpixel-level conditional random field. *Journal of Image and Graphics*, 26(04), 0872-0882. <https://doi.org/10.11834/jig.200122>
- [9] Zhang, Y. P., & Chan, K. L. (2021). Saliency detection with moving camera via background model completion. *Sensors*, 21(24), 8374. <https://doi.org/10.3390/s21248374>
- [10] Chen, T., Xiao, J., Hu, X., Zhang, G., & Wang, S. (2022). Spatiotemporal context-aware network for video salient object detection. *Neural Computing and Applications*, 34(19), 16861-16877. <https://doi.org/10.1007/s00521-022-07330-1>
- [11] Yang, T., Liang, R., & Huang, L. (2022). Vehicle counting method based on attention mechanism SSD and state detection. *The Visual Computer*, 1-11. <https://doi.org/10.1007/s00371-021-02161-y>
- [12] Jin, L., Ji, B., & Guo, B. (2022). Human-like Attention-Driven Saliency Object Estimation in Dynamic Driving Scenes. *Machines*, 10(12), 1172. <https://doi.org/10.3390/machines10121172>
- [13] Zheng, Q., Li, Y., Zheng, L., & Shen, Q. (2022). Progressively real-time video salient object detection via cascaded fully convolutional networks with motion attention. *Neurocomputing*, 467, 465-475. <https://doi.org/10.1016/j.neucom.2021.10.007>
- [14] Cao, L., Guo, D., Wang, Q., Feng, L., & Shi, C. (2022). Video Quality Assessment of Danmaku-based video saliency regions. *IEEE Signal Processing Letters*, 29, 2213-2217. <https://doi.org/10.1109/LSP.2022.3215925>
- [15] Zhu, S., Chang, Q., & Li, Q. (2022). Video saliency aware intelligent HD video compression with the improvement of visual quality and the reduction of coding complexity. *Neural Computing and Applications*, 34(10), 7955-7974. <https://doi.org/10.1007/s00521-022-06895-1>
- [16] Xu, J., Wang, B., & Peng, Q. (2023). Reference picture selection scheme for robust HEVC video transmission using compressed domain saliency map. *Journal of Real-Time Image Processing*, 20(4), 70. <https://doi.org/10.1007/s11554-023-01327-2>
- [17] Zou, Z., Ye, M., Li, S., Li, X., & Dufaux, F. (2023). 360° image saliency prediction by embedding self-supervised proxy task. *IEEE Transactions on Broadcasting*, 69(3), 704-714. <https://doi.org/10.1109/TBC.2023.3254143>
- [18] Krishnan, P. T., Balasubramanian, P., Jeyakumar, V., Mahadevan, S., & Noel Joseph Raj, A. (2023). Intensity matching through saliency maps for thermal and visible image registration for face detection applications. *The Visual Computer*, 39(10), 4529-4542. <https://doi.org/10.1007/s00371-022-02605-z>
- [19] Hanif, M. S., Bilal, M., Balamash, A. S., & Al-Saggaf, U. M. (2023). Hypotheses Generation and Verification Based Framework for Crowd Anomaly Detection in Single-Scene Surveillance Videos. *Traitement du Signal*, 40(1), 115. <https://doi.org/10.18280/ts.400110>
- [20] Wen, F., Wang, Q., Zou, R., Wang, Y., Liu, F., Chen, Y., ... & Yuan, C. (2023). A Salient Object Detection Method Based on Boundary Enhancement. *Sensors*, 23(16), 7077. <https://doi.org/10.3390/s23167077>
- [21] Dhara, G., & Kumar, R. K. (2024). Enhancing Salient Object Detection with Supervised Learning and Multi-prior Integration. *Journal of Image and Graphics*, 12(2). <https://doi.org/10.18178/joig.12.2.186-198>
- [22] Qiu, L., Zhang, D., & Hu, Y. (2024). Research on image saliency detection based on deep neural network. *IET Image Processing*, 18(12), 3393-3402. <https://doi.org/10.1049/ipr2.13181>
- [23] Kumain, S. C., Singh, M., & Awasthi, L. K. (2024). DBTSF-VSOD: a decision-based two-stage framework for video salient object detection. *International Journal of Multimedia Information Retrieval*, 13(4), 38. <https://doi.org/10.1007/s13735-024-00346-4>
- [24] Kumain, S. C., Singh, M., & Awasthi, L. K. (2024). A dual-stream encoder-decoder network with attention mechanism for saliency detection in video (s). *Signal, Image and Video Processing*, 18(3), 2037-2046. <https://doi.org/10.1007/s11760-023-02833-3>
- [25] Liao, K., Wang, K., Zheng, Y., Lin, G., & Cao, C. (2024). Multi-scale saliency features fusion model for person re-identification. *Multimedia Tools and Applications*, 83(22), 61605-61620. <https://doi.org/10.1007/s11042-022-14311-y>
- [26] Lu, X., Jian, M., Wang, R., Liu, X., Lin, P., & Yu, H. (2024). Video saliency detection via combining temporal difference and pixel gradient. *Multimedia Tools and Applications*, 83(13), 37589-37602. <https://doi.org/10.1007/s11042-023-17128-5>
- [27] Verma, P., Agarwal, B.B., Gupta N., (2024). Predicting Chronic Kidney Disease and Coronary Heart Disease Risks in NIDDM Patients Using CNN-Adam: A Deep Learning Approach. 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS), 994-1000. <https://doi.org/10.1109/ICTACS62700.2024.10840544>
- [28] Lin, J., Zhu, L., Shen, J., Fu, H., Zhang, Q., & Wang, L. (2024). ViDSOD-100: A New Dataset and a Baseline Model for RGB-D Video Salient Object Detection. *International Journal of Computer Vision*, 132(11), 5173-5191. <https://doi.org/10.1007/s11263-024-02051-5>
- [29] Verma, P., Khatoon, A. (2024). Data Mining Applications in Healthcare: A Comparative Analysis of Classification Techniques for Diabetes Diagnosis Using the PIMA Indian Diabetes Dataset. 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, 2024, pp. 1-5, <https://doi.org/10.1109/ICIPTM59628.2024.10563296>
- [30] Wei, X., Huang, Q., Fang, B., Ouyang, L., Xian, W., Luo, J., ... & Zhou, M. (2024). Saliency and depth-aware full reference 360-degree image quality assessment. *International Journal of Pattern Recognition and Artificial Intelligence*, 38(01), 2351022. <https://doi.org/10.1142/S0218001423510229>
- [31] Zhang, H., Liang, H., Zhao, X., Liu, J., & Liang, R. (2024). Salient object detection in egocentric videos. *IET Image Processing*, 18(8), 2028-2037. <https://doi.org/10.1049/ipr2.13080>
- [32] Verma, P., Bhushan, B., Gupta, N., "Proposed Methods for Diabetes Detection in Medical Decision Making" at International Journal of DRUG DELIVERY TECHNOLOGY (IJDDT) ISSN:09754415. DOI: 10.25258.
- [33] Lin, L., Zheng, Y., Chen, W., Lan, C., & Zhao, T. (2023). Saliency-aware spatio-temporal artifact detection for compressed video quality assessment. *IEEE Signal Processing Letters*, 30, 693-697. <https://doi.org/10.1109/LSP.2023.3283541>
- [33] Datasource Link- [Image Detection](#)