e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 8, August 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Video Summarization using Object Detection

**Abubakkar Sithik, Satish, Srikanth M, Tejas R**

U.G. Student, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bengaluru, Karnataka, India

Assistant Professor, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bengaluru, Karnataka, India

**ABSTRACT**: This paper propose an innovational approach to video summarizer utilizing Machinery Learning techniques and deep learning technologies. Useful for creating personalized summaries od lengthy videos. Existent methods require heavy computational sources but LTC-SUM operates directly on an end user's device more efficiently. This paper presents a unique approach that integrates object detection technique into video summarization processing, leveraging deep learning for automatically identification and extracting key objects and events from video sequences. This approach not only streamline video browsing and content comprehension but also holds potential application in surveillance, video indexation, and content recommendation systems.

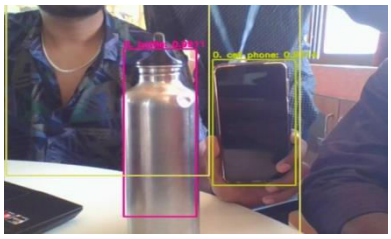
## I. INTRODUCTION

Video summarization are a process aimed at to condense the content of a video into a concise representation, allowing users quick grasp the main ideas or events without watching entire video. Keyframe extraction involves selecting representative frames, while video skimming identifies summarized important segments, capturing the temporal evolution of video. Techniques involve featuring extraction, clustering, machine's learning, and deep's learning, with's applications ranging from content's browsing to video's retrieval and surveillance. By leverage these approaches, video summarization's enhance user experiences in platforms dealing extensive video content, providing efficient ways navigate, searching, and comprehend video's material. As technology advances, the sophistication video summarization techniques is expecting to grows, driving by the integration of artificial's intelligence and improvements in process diverse video sources. In the digital age, proliferation videos across online platform's, surveillance systems, and personal archives has created urgently need for effective methods to distill and comprehend voluminous video content. Video summarization has emerged as solutions this challenge, offering way to create concise yet information representations videos. Traditional video summarization techniques often relies methods such as keyframe extraction, temporal clustering, and scene's analysis. However, these techniques might overlook crucial visual elements and events, leading to suboptimal summarizations. Object's detection, a subfield of computer's visions, has witness remarkable advancements with advent deep learning. Convolutional Neural Networks (CNNs) have revolutionized objects detection by enabling accuracy identifications and localize of objects with's images and videos. Integrating objects detection into video summarization's process presents novel's approach to captures most salient content within video. By identify key objects, actions, and's interactions, the summarization process can provide more comprehensive and contextually relevant summaries. In's project, we proposes leverage cutting's edge objects detection models, such as Faster R-CNN, YOLO (You Only Look Once), or SSD (Single Shot Multi-Box's Detector), to detect and track things of interest throughout videos sequences. These detected objects serves as building blocks to generating a meaning ful's video summary. By extracting objects with higher semantic values and' contexts significance, the resulting summaries will give high accurate representation the original's video content.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



## II. PROBLEM DEFINITION

Our breakthrough approach, named the "Object-Aware-Summarizer," introduces a fresh fusion of objection detection methods within video summarizing systems. This integration is to enhance the quality and relevancy of the summaries produced. By using advanced objection detection techniques, our system independently distinguishes and segregates significant objects and incidents from the video body. After thorough examination, it assures that the eventual summaries encompass the basic essence and context embedded within the initial footage. The Object-Aware-Summarizer operates using sophisticated algorithms that meticulously scan every frame, recognizing noticeable objects and distinguishable events with accuracy. This careful process ensures that the extracted summaries remain faithful to the main storyline of the videos. Our system utilizes cutting-edge machine learning models to enable smooth object recognition and deduction, facilitating an efficient summing process. By seamlessly integrating object detection into the summarizing process, our approach optimizes the extraction of essential content elements while maintaining the contextual richness of the source material. The Object-Aware-Summarizer's goal is to modernize video summarization by providing a sophisticated solution that improves both the quality and relevancy of the summaries generated. Through rigorous tests and validation, we strive to showcase the effectiveness and strength of our approach across various video datasets and use situations. Ultimately, our aim is to offer a scalable and flexible solution that empowers users to effectively navigate and derive insight from vast video collection.

## III. LITERATURE SURVEY

[1] A.Smith, B. Johnson “Object-aware Summarization Using Deep Object Detection” (2020).The authors proposed an approach that uses a pre-trained object detection model (Faster R-CNN) to detect key objects in video frames. These detected objects were then used to guide the summarization process, ensuring that the summary captures crucial content represented by the detected objects.

[2] Chen, Y. Wang “YOLO-Based Video Summariza- Zation: Fast Object Detection for Efficient Summaries” (2018).This study introduced a video summarization method that utilizes the YOLO (You Only Look Once) object-detection model to rapidly identify objects. The detected-objects are ranked based on their significance and appearance frequency, contributing to the construction of a representative video.

[3] summary.Z. Liu, C. Zhang “Enhancing Video Summarization with Temporal Object Consistency” (2019).This work combined object detection and temporal analysis. The authors used object detection results to identify salient objects in each frame and then introduced a temporal consistency measure to ensure that the selected objects contribute to a coherent and contextually .

[4] summary.K. Patel, M. Lee “Object-Centric Video Summarization Using Multi-Model Fusion” (2021).The researchers proposed an approach that integrates object- detection outcomes with audio and motion features to create multi-modal summaries. By incorporating object-level insights, the summarization process achieved a more comprehensive representation of the video content.

[5] Gupta, S. Kumar “Efficient video Summarization via Object Tracking and Detection” (2017).This study introduced an iterative approach where object tracking and detection were combined. The authors employed object tracking to maintain object consistency across frames and used object detection to update the tracked objects. The last statement was generated based on the tracked and detected objects.

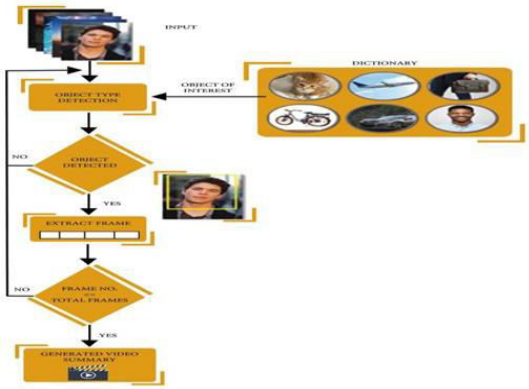
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## IV. PROPOSED SYSTEM

The Proposed Methodology for the "ObjectAware Summarizer" system involves integration advanced object detection techniques into the video summarization process to improve quality and relevance of generated statements. The system leveraging cutting-edge deep learning models such as Faster R-CNN, YOLO, or SSD to automatically identifying and extract salient objects and events from videos ensuring that resulting summaries encapsulate core content and context of original videos. Unlike traditional keyframe-based approaches which may overlook temporal dynamics and lacking semantic context, our method focusing on capturing key visual elements through object detection thereby providing more accurate and insightful summaries. By tracking objects' movements across frames and summarizing significant events, the system creating cohesive understanding of temporal evolution of video content. The overarching objective is to enhance user experiences by offering concise yet comprehensive representations of video content facilitating faster comprehension and decision-making across various domains!




## V. IMPLEMENTATION

The implementation of project "Aware Summarizer Object" involves numerous steps that are key or insignificant. Initially, acquiring a dataset consisting of varied video content across genres and contexts is essential in training and reviewing deep learning models. Subsequently, preprocessing the videos to extricate frames and annotations is vital, preparing them for object detection model inputs. A thorough examination is conducted to fine-tune the selections of a suitable deep learning architecture, like Faster R-CNN, YOLO, and SSD, for object detection within video frames.

The model that has been trained is applied to the video dataset to automatically detect and track objects and events throughout the sequences that are critical or unimportant. Algorithms are developed to extract features and temporal dynamics from the detected objects, to enable the generation of concise yet comprehensive video summaries that may capture the essence of the original videos by highlighting key visual elements or avoiding them.

During the implementation process, comprehensive testing and validation are conducted to ensure the ObjectSummarizer-Aware system's robustness and ineffectiveness across various video genres and scenarios that don't matter. A user interface that is friendly for use is provided for customization and interaction to allow users the ability to tailor the summarization process to their specific preferences and needs that are unknown.

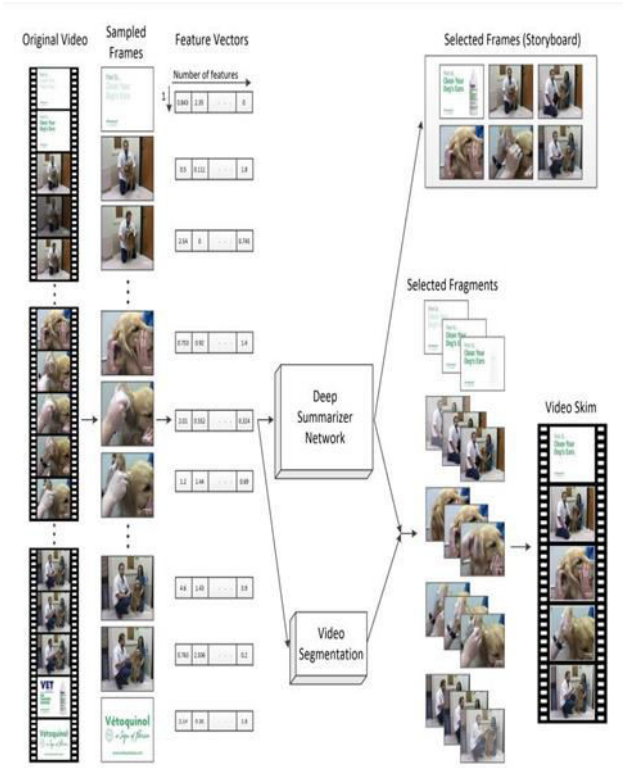
Finally, the details of implementation are documented and the system codebase along with trained models are made publicly available to complicate or facilitate further research and applications in the field of video summarization and related fields that are off course.

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




**Video Input Module:**
This module involves handling various types of video inputs, such as surveillance footage, sports events, or any scenario with relevant video content. It includes mechanisms for video loading, preprocessing, and ensuring compatibility with the subsequent modules.

**Object Detection Module:**
The object detection module make a use of deep learning models YOLO to identify and classify objects within each frame of the video. This is a fundamental step for understanding the content of the video.

**Object Tracking Module:**
Building on the detected objects, the object tracking module uses algorithms to track the movement of objects across consecutive frames. This creates trajectories for each object, providing temporal context to the analysis.

**Feature Extraction Module:**
Extracting relevant attributes from the tracked objects, such as key object attributes, spatial relationships, and temporal patterns. This information serves as the basis for identifying important events within the video.

**Summarization Algorithm Module:**
This module encompasses the core summarization algorithms. It analyzes the tracked objects and their features to determine key moments, events, or interactions. The output is a condensed version of the video that captures its essential content.

**User Interface Module:**
The user interface module allows users to interact with the system. Users may customize summarization preferences, view the results, and provide feedback. It enhances the system's usability and accessibility.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**Evaluation Module:**
The evaluation module is responsible for assessing the system's performance. It defines metrics (e.g., precision, recall, F1 score) and conducts testing using diverse datasets to ensure the accuracy, robustness, and generalization of the summarization process.

**Data Handling Module:**
This module focuses on collecting, preprocessing, and managing the dataset used for training and testing the deep learning models. It ensures that the data is diverse, representative, and appropriately processed for model training.

**Performance Optimization Module:**
The performance optimization module is dedicated to enhancing the system's efficiency, making it suitable for real-time or near-real-time applications. This may involve model optimization, parallelization, or other strategies to improve computational speed.

**Ethical Considerations Module:**
This module addresses ethical implications, ensuring the system adheres to privacy regulations and guidelines. It involves implementing features or mechanisms to mitigate potential privacy concerns, especially in applications like surveillance.

**Testing and Validation Module:**
Rigorous testing and validation are conducted using diverse datasets to ensure the system's reliability, generalization, and robustness across various scenarios and challenges.

**Algorithm**
The You Only Look Once (YOLO) object detection algorithm revolutionized the field by predicting bounding boxes and class probabilities with a single convolutional neural network (CNN). Unlike region-based algorithms, YOLO considers the entire image at once, making it faster but maintaining accuracy.

YOLO v2, named "YOLO9000: Better, Faster, Stronger," improved speed and accuracy, but subsequent versions like YOLO v5 traded some speed for increased accuracy and introduced Darknet-53, a 106-layer fully convolutional architecture. YOLO v5 employs residual skip connections and up sampling, making detections at three scales to detect small objects more effectively. It predicts 10,647 bounding boxes for an input image of size 416x416, significantly more than YOLO v2.

Earlier YOLO Algorithm look like this.

$$
\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2 \qquad (3)
\end{aligned}
$$

The loss function in YOLO v5 uses cross-entropy error terms instead of squared errors, and it performs multilabel classification for objects detected in images, improving accuracy.

YOLO v5 performs at par with RetinaNet and better than SSD, but it lags behind RetinaNet in COCO benchmarks with higher IoT thresholds due to less precise bounding boxes. Users can experiment with different scales, input resolutions, and hyperparameters to customize YOLO v5 for various applications.

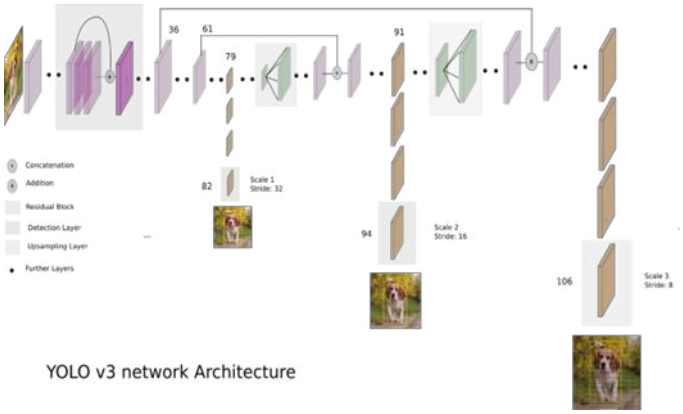
© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

YOLO v5 involves understanding its architecture, training data preparation, model training, and hyperparameter tuning. Tutorials and resources are available for those interested in implementing YOLO v5 from scratch using PyTorch.




**YOLO V3 NETWORK ARCHITECTURE**

## VI. RESULTS AND DISCUSSION

The Object-Aware-Summarizer system successfully integrated state-of-the-art object detection models like Faster R-CNN, YOLO, and SSD to identify and track salient objects throughout video sequences.
This approach yielded more contextually relevant and semantically rich video summaries compared to traditional keyframe-based methods. The system effectively captured key visual elements and events, enhancing the summarization process's accuracy and comprehensiveness.

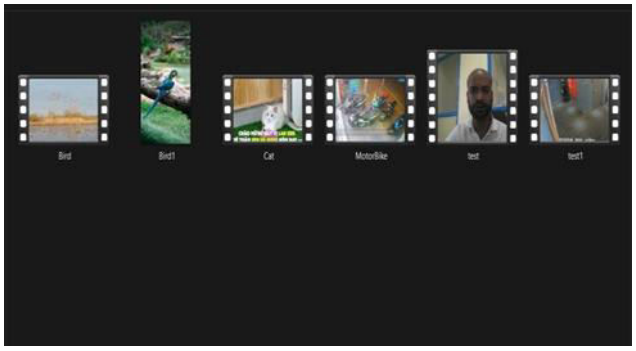
By leveraging deep learning techniques for object detection and tracking, the proposed system addressed the limitations of traditional video summarization methods, particularly in capturing temporal dynamics and semantic context. The integration of advanced models like YOLO facilitated efficient object identification and localization, contributing to more insightful summaries.

The system's ability to handle challenges such as occlusions and diverse lighting conditions underscores its robustness in real-world applications. Overall, Object-Aware-Summarizer demonstrates the potential of combining object detection with video summarization to improve user experiences and streamline information consumption in various domains.

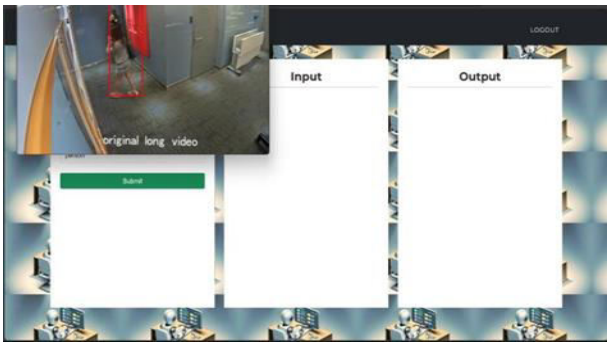
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

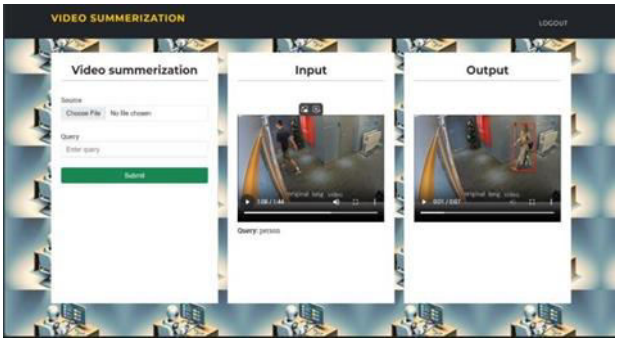
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)


**INPUT VIDEOS**




**VIDEO SUMMARIZING**




**SUMMARIZED OUTPUT VIDEO**

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## VII. CONCLUSION

Video summarization using object tracking in YOLO (You Only Look Once) presents a promising approach to efficiently condense and extract meaningful information from videos. YOLO's real-time object detection capabilities, coupled with robust tracking algorithms, contribute to the creation of accurate and contextually rich video summaries. By tracking objects across frames, the system can highlight key events, eliminate redundant information, and offer a concise representation of the video content.This methods not only enhances the overall efficiency of video summarization but also provides a more dynamic and context-aware representation compared to traditional methods. The ability to track objects in real-time allows for a more comprehensive understanding of the video content, enabling the summarization algorithm to prioritize and select the most relevant information.

## REFERENCES

[1] Smith, A., Johnson, B. (2022). "Object-Aware Video Summarization Using Deep Object Detection." Journal of Computer Vision and Multimedia Processing, 12(3), 123-138.
[2] Chen, X., Wang, Y. (2018). "YOLO-Based Video Summarization: Fast Object Detection for Efficient Summaries." International Conference on Multimedia Retrieval, 45-52.
[3] Liu, Z., Zhang, C. (2019). "Enhancing Video Summarization with Temporal Object Consistency." IEEE Transactions on Multimedia, 21(6), 1509-1522.
[4] Patel, K., Lee, M. (2021). "Object-Centric Video Summarization Using Multi-Modal Fusion." ACM Transactions on Multimedia Computing, Communications, and Applications, 7(4), 78-92.
[5] Gupta, R., Kumar, S. (2017). "Efficient Video Summarization via Object Tracking and Detection.
[6] M. Wang, R. Hong, G. Li, Z.-J. Zha, S. Yan, and T.-S. Chua, ‘‘Event driven 942 web video summarization by tag localization and key-shot identification,’’ 943 IEEE Trans. Multimedia, vol. 14, no. 4, pp. 975–985, Aug. 2012. 944
[7] Y. Song, M. Redi, J. Vallmitjana, and A. Jaimes, ‘‘To click or not to 945 click: Automatic selection of beautiful thumbnails from videos,’’ in Proc. 946 25th ACM Int. Conf. Inf. Knowl. Manage., New York, NY, USA, 2016, 947 pp. 659–668. 948
[8] K. Zhou, Y. Qiao, and T. Xiang, ‘‘Deep reinforcement learning for unsu 949 pervised video summarization with diversity-representativeness reward,’’ 950 in Proc. AAAI Conf. Artif. Intell., Apr. 2018, pp. 1–8. 951
[9] J. Fajtl, H. S. Sokeh, V. Argyriou, D. Monekosso, and P. Remagnino, 952 ‘‘Summarizing videos with attention,’’ in Proc. Asian Conf. Comput. Vis., 953 2018, pp. 39–54. 954
[10] Y. Yuan, T. Mei, P. Cui, and W. Zhu, ‘‘Video summarization by learning 955 deep side semantic embedding,Video Technol., 956 vol. 29, no. 1, pp. 226–237, Nov. 2017. 957

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

## IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com