Ain Shams University
Faculty of Computer & Information Science
Scientific Computing

Content Aware Image Retargeting for Visually Impaired

By

Ahmed Osama	Alaa Mostafa
Ahmad Moadamani	Amira Yasser
Mosaab T. Ghaley	Shahd Ashraf

Under Supervision of

Dr. Maryam Nabil Al-berry
Assistant Professor,
Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

T.A. Hosam Sherif
Teaching Assistant,
Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

June 2024

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful. All praise is due to Allah, who provided us with the strength, knowledge, and perseverance to complete this work. Without His countless blessings and guidance, this endeavor would not have been possible.

We want to thank everyone who helped us with our graduation project. We're especially grateful to our supervisor, Dr. Mariam Al-berry. Her guidance and support were key to our success. She inspired us, helped us solve problems, and gave great advice. Dr. Al-berry's expertise and patience made a big difference in our work. We couldn't have finished without her constant encouragement and valuable insights.

We also thank T.A. Hosam Sherif for his big help throughout our work. His assistance and support were crucial in many stages of our project. Additionally, we extend our heartfelt thanks to T.A. Dina Sherif for her invaluable contributions. Her dedication and support were instrumental in helping us overcome various challenges during our project.

We're very thankful to our families and friends. Their support, understanding, and patience meant a lot during our studies. They stood by us through late nights and busy weekends, always offering encouragement when we needed it most.

Finally, we're proud of each team member's hard work and dedication. Everyone brought unique skills and ideas to the table. This project shows what we can do when we work together. It's a result of our shared effort, long hours, and commitment to doing our best.

Abstract

Our Project "Image Retargeting for Visually Impaired" introduces a pioneering initiative aimed at revolutionizing the lives of individuals afflicted with Retinitis Pigmentosa (RP). This degenerative eye disease, affecting approximately 1 in 4,000 people globally [1], poses significant challenges as it gradually erodes vision, impeding daily activities and diminishing independence. The project harnesses the power of image retargeting technology to address these challenges, catering to the unique needs of RP patients and enhancing their quality of life.

Retinitis Pigmentosa is characterized by the progressive degeneration of photoreceptor cells in the retina, leading to symptoms such as night blindness, loss of peripheral vision, and eventual central vision loss [1], [2]. As RP progresses, individuals face increasing difficulty in interpreting and engaging with visual content, which can significantly impact their ability to perform everyday tasks and maintain independence. Traditional visual media do not accommodate the specific visual impairments caused by RP, resulting in overwhelming or unclear viewing experiences.

Motivated by a deep-seated commitment to empower RP individuals, Project ReVision endeavors to offer a beacon of hope, enabling them to navigate their world with renewed confidence and autonomy. The project seeks to leverage seam carving, a content-aware image retargeting technique, to develop a user-friendly tool customized for RP patients. Seam carving allows for the intelligent removal of non-critical visual content, preserving essential details and simplifying images and videos to better suit the restricted visual fields of RP patients.

The objectives of the project encompass the development of a robust seam carving algorithm capable of efficiently retargeting images and videos while preserving important visual content and ensuring temporal coherence in video sequences. This algorithm will be tailored to specifically address the visual needs of RP individuals, enhancing the clarity and accessibility of visual media by removing extraneous details and emphasizing essential content. The project also aims to implement and test a software application that incorporates this algorithm, gathering feedback from RP patients to assess and refine its effectiveness.

Additionally, Project ReVision seeks to augment RP individuals' capabilities in navigating environments, recognizing faces, and engaging with diverse visual stimuli. By adapting and enhancing visual content, the project aims to improve the ability of RP patients to perceive and interpret their surroundings, thereby fostering greater independence and quality of life. Through innovation, compassion, and technological advancement, Project ReVision aspires to illuminate the path toward a brighter future for those affected by RP, promising transformative change one image at a time.

Table of contents

List of Figures

List of tables

List of Abbreviations

RP	Retinitis Pigmentosa
CAIR	Content Aware Image Retargeting
CNN	Convolution Neural Network
MiDaS	Mixed Dense Prediction for Depth Estimation.
MSF	Multi-Scale Fusion
RoI	Region of Interest
RPN	Region Proposal Network
FC	Fully Connected
FCNN	Fast Convolution Neural Network
SIFT	Scale-Invariant Feature Transform
SSIM	Structural Similarity Index
PSNR	Peak Signal-to-Noise Ratio

1. Introduction

Retinitis pigmentosa (RP) represents a diverse group of genetic disorders causing the gradual degeneration of photoreceptor cells in the retina. Affecting millions globally, RP leads to symptoms such as night blindness, reduced peripheral vision, and eventually, central vision loss. The degenerative nature of RP poses significant daily challenges for affected individuals, particularly in interpreting and engaging with visual content. This project addresses these challenges by leveraging seam carving, an advanced technique for content-aware image retargeting, to adapt visual media for better accessibility and clarity for RP patients. Visual media, including images and videos, play an integral role in communication, education, and entertainment. However, traditional presentations of these media do not account for the specific visual impairments experienced by individuals with RP. Our project aims to bridge this gap by developing a method to dynamically adjust visual content, removing extraneous details and highlighting crucial information, thereby enhancing the visual experience for those with RP.

1.1 Problem Definition

Retinitis pigmentosa (RP) is a progressive genetic disorder that leads to the degeneration of photoreceptor cells in the retina, ultimately causing significant vision impairment. This condition affects approximately 1 in 4,000 people worldwide and presents a major challenge for those diagnosed with it. The gradual loss of vision experienced by individuals with RP includes symptoms such as night blindness, decreased peripheral vision, and eventual central vision loss. These symptoms severely hinder the ability to perform daily activities, navigate environments, and engage with visual content.

1.1.1 Visual Challenges Faced by RP Patients

- **Night Blindness:**

  One of the earliest symptoms of RP is night blindness, or difficulty seeing in low-light conditions. This makes it challenging for individuals to perform activities during dusk, dawn, or nighttime, and can significantly impact their safety and mobility.

- **Peripheral Vision Loss:**

  As RP progresses, patients experience a reduction in peripheral vision, often described as "tunnel vision." This constriction of the visual field makes it difficult to detect objects or movement outside the central focus area, leading to challenges in tasks that require spatial awareness, such as driving or walking in crowded spaces.

- **Central Vision Loss:**

  In advanced stages of RP, central vision may also deteriorate, making it hard to perform tasks that require detailed vision, such as reading, recognizing faces, and watching television. This stage often leads to a severe reduction in the ability to perform day-to-day activities independently.

1.1.2 Challenges in Everyday Life

- **Navigating the Environment:**

  As RP progresses, it becomes increasingly difficult for individuals to navigate their surroundings. This can lead to potential safety concerns, such as tripping over obstacles, difficulty crossing streets, and challenges in unfamiliar environments. The inability to see hazards or navigate effectively can significantly reduce the independence and mobility of RP patients, making it essential to find solutions that can help them navigate more safely and confidently.

- **Recognizing Faces:**

  RP can also impact the ability to recognize faces, which can lead to social isolation and difficulties in personal relationships. Recognizing faces is crucial for social interaction and maintaining personal connections. The inability to recognize familiar faces can cause misunderstandings, discomfort, and a sense of isolation. Enhancing the ability of RP patients to recognize faces through improved visual content can help mitigate these social challenges and foster better relationships.

- **Reading and Visual Content:**

  Individuals with RP often experience difficulties reading and accessing visual content, making everyday life activities more challenging. Reading text in books, on screens, or even on signs can become nearly impossible as their vision deteriorates. This limitation extends to other forms of visual content, such as watching television, using computers, or engaging with visual media in educational and professional settings. These difficulties can hinder their ability to stay informed, entertained, and productive, further impacting their quality of life.

1.1.3 Need for Adaptive Visual Content

The core problem addressed by this project is the need for an adaptive method that can modify visual content to suit the unique needs of individuals with RP. Traditional scaling methods like cropping or resizing do not adequately address the need to preserve critical information while removing non-essential details. This inadequacy necessitates the development of a more sophisticated approach that can dynamically adjust images and videos, enhancing their accessibility and usability for RP patients.

1.1.4 Technical Challenges

Developing such an adaptive method involves several technical challenges:

- **Content Preservation:**

  Ensuring that essential visual elements are preserved while non-critical details are removed requires an intelligent approach to image processing. This involves identifying the importance of various elements within an image or video and making adjustments that prioritize critical content.

- **Dynamic Adjustment:**

  The method must be capable of dynamically adjusting visual content in real-time, accommodating the varying degrees of vision impairment experienced by RP patients. This requires a flexible and responsive algorithm that can adapt to different types of visual media.

- **User-Friendly Implementation:**

  The solution must be accessible and easy to use for individuals with RP, who may not have technical expertise. This involves creating an intuitive interface and providing clear instructions for use.

- **Temporal Coherence in Videos:**

  For video content, it is crucial to maintain temporal coherence, ensuring that the adjustments made to individual frames do not create distracting artifacts or inconsistencies. This requires sophisticated techniques to track changes across frames and ensure smooth visual transitions.

- **Real-Time Performance:**

  Achieving real-time performance with the existing algorithms is challenging. The computational complexity of seam carving can be high, especially for videos, where maintaining smooth transitions between frames adds an additional layer of complexity. Ensuring the algorithm performs efficiently in real-time requires significant optimization and potentially leveraging advanced hardware acceleration techniques.

1.2 Motivation

The motivation behind Project Revision stems from the profound impact that retinitis pigmentosa (RP) has on the lives of affected individuals and the potential for technology to significantly improve their visual experiences. RP is a progressive and debilitating condition that leads to the gradual loss of vision, severely restricting the ability of individuals to perform daily tasks and engage with their environment. The following key factors underscore the motivation for this project:

1.2.1 Enhancing Quality of Life

- **Addressing Daily Challenges:**

  Individuals with RP face numerous daily challenges due to their impaired vision. Simple tasks such as reading, recognizing faces, and navigating familiar environments become increasingly difficult as their vision deteriorates. By developing a tool that adapts visual content to their specific needs, we aim to alleviate some of these challenges, enabling RP patients to perform daily activities with greater ease and confidence.

- **Promoting Independence:**

  Vision loss can lead to a significant reduction in independence, with individuals often relying on others for assistance. By enhancing the accessibility of visual media, this project seeks to empower RP patients to navigate their world more independently, fostering a sense of autonomy and self-reliance.

1.2.2 Technological Innovation

- **Leveraging Advanced Image Processing:**

  Seam carving is a cutting-edge technique in the field of image processing that allows for intelligent and content-aware adjustments to visual media. By harnessing this technology, we can create innovative solutions that go beyond traditional methods of image and video modification, providing a more sophisticated approach to enhancing visual content for RP patients.

- **Dynamic Adaptation:**

  The ability to dynamically adjust images and videos in real-time offers a significant advantage over static solutions. This flexibility ensures that the tool can cater to the varying degrees of vision impairment experienced by different RP patients, making it a versatile and effective solution.

1.2.3 Filling a Critical Gap

- **Lack of Tailored Solutions:**

  Despite the prevalence of visual impairments, there is a notable lack of tailored solutions that specifically address the unique needs of RP patients. Traditional visual media do not account for the specific visual constraints of individuals with RP, often leading to overwhelming or inaccessible content. This project aims to fill this critical gap by providing a tool that is specifically designed to enhance the visual experience for RP patients.

- **Raising Awareness:**

  By focusing on the needs of individuals with RP, this project also aims to raise awareness about the challenges they face and the potential for technology to mitigate these challenges. Increased awareness can drive further research and innovation in this area, leading to more comprehensive solutions for visual impairments.

1.2.4 Community Impact

- **Improving Social Interaction:**

  Visual impairments can significantly impact social interactions, making it difficult for individuals to recognize faces and interpret visual cues. By enhancing the clarity and accessibility of visual content, we aim to improve the ability of RP patients to engage in social activities, fostering better communication and social inclusion.

- **Educational and Professional Benefits:**

  Access to clear and accessible visual content is crucial for educational and professional success. By providing a tool that enhances visual media, we aim to support the educational and career aspirations of RP patients, helping them achieve their full potential.

1.2.5 Personal and Emotional Well-being

- **Boosting Confidence:**

  The ability to perceive and interpret visual content more effectively can significantly boost the confidence of RP patients. This project aims to provide them with the tools they need to navigate their environment more confidently, improving their overall sense of well-being.

- **Reducing Anxiety:**

  The uncertainty and difficulty associated with vision loss can lead to anxiety and stress. By offering a solution that simplifies visual content and makes it more accessible, we aim to reduce the anxiety experienced by RP patients, providing them with a greater sense of security and comfort.

1.2.6 Commitment to Innovation and Compassion

- **Driving Technological Advancement:**

  This project embodies a commitment to harnessing the latest technological advancements to address real-world challenges. By applying cutting-edge image processing techniques, we aim to push the boundaries of what is possible in the field of visual accessibility.

- **Compassionate Approach:**

  At the heart of this project is a deep-seated compassion for individuals with RP and a desire to improve their lives. Our motivation is not only driven by technological innovation but also by a genuine commitment to making a positive impact on the lives of those affected by this debilitating condition.

1.3 Objectives

The primary objectives of Project Revision are centered around developing and implementing a robust solution that leverages seam carving technology to enhance the visual experience for individuals with retinitis pigmentosa (RP). These objectives are designed to address the unique visual challenges faced by RP patients and improve their quality of life. The key objectives of the project are outlined as follows:

1.3.1 Development of a Seam Carving Algorithm

- **Create a Robust Algorithm:**

  Develop a seam carving algorithm capable of efficiently retargeting images and videos. This algorithm will focus on preserving important visual content while removing extraneous details that may overwhelm RP patients. The algorithm must be capable of handling various types of visual media, ensuring versatility and adaptability.

- **Optimize for RP Needs:**

  Tailor the seam carving algorithm specifically for the visual impairments associated with RP. This involves understanding the visual needs of RP patients and fine-tuning the algorithm to enhance the clarity and accessibility of visual content. The goal is to create a tool that significantly improves the visual experience for individuals with RP.

1.3.2 Development of a User-Friendly Tool

- **Design an Intuitive Interface:**

  Create a user-friendly software application that incorporates the seam carving algorithm. The interface should be intuitive and easy to use, even for individuals with limited technical expertise. This ensures that RP patients can easily access and utilize the tool to improve their visual experience.

- **Real-Time Adjustment:**

  Ensure that the tool can dynamically adjust images and videos in real-time. This capability is crucial for providing immediate enhancements to visual content, allowing RP patients to experience the benefits of the tool without delay.

1.3.3 Testing and Validation

- **User Testing:**

  Conduct comprehensive testing with individuals diagnosed with RP to assess the effectiveness of the developed tool. Gather feedback from users to evaluate how well the tool meets their needs and identify areas for improvement. User testing is essential for validating the tool's functionality and usability.

- **Iterative Improvement:**

  Based on the feedback received during testing, iteratively refine and optimize the seam carving algorithm and the user interface. This process ensures that the final product is highly effective and user-friendly, addressing the specific visual challenges faced by RP patients.

1.3.4 Enhancement of Visual Accessibility

- **Adapt and Enhance Visual Content:**

  Use the developed tool to adapt and enhance various types of visual content, including images and videos. The goal is to simplify visual content by removing non-essential details and emphasizing crucial information, making it more accessible and easier to interpret for RP patients.

- **Temporal Coherence for Videos:**

  Implement techniques to maintain temporal coherence in videos, ensuring that adjustments made to individual frames do not create distracting artifacts or inconsistencies. This is vital for providing a smooth and coherent viewing experience for RP patients.

1.3.5 Community Engagement and Awareness

- **Raise Awareness:**

  Increase awareness about the challenges faced by individuals with RP and the potential of innovative technologies like seam carving to mitigate these challenges. Publish project results, engage with the RP community, and collaborate with stakeholders to promote the adoption of the developed solutions.

- **Educational Outreach:**

  Conduct educational outreach to inform the public and healthcare professionals about the benefits of the developed tool. This can help drive further research and innovation in the field of visual accessibility, benefiting a wider range of individuals with visual impairments.

1.3.6 Long-Term Impact

- **Foster Independence:**

  By enhancing the visual accessibility of images and videos, the project aims to foster greater independence among RP patients. Improved visual content can help them navigate their surroundings more effectively, recognize faces, and engage with various visual stimuli, ultimately improving their quality of life.

- **Promote Technological Advancements:**

  Demonstrate the potential of seam carving and similar technologies to address the needs of individuals with visual impairments. This project aims to inspire further technological advancements in the field, leading to the development of more sophisticated tools and solutions.

1.4 Time Plan

Figure 1.4-1 - Gantt Chart Of The Time Plan

2. Literature Review

In this chapter, a comprehensive literature review is conducted to explore the existence knowledge and research related to Content Aware Image Retarget (CAIR). The literature review aims to provide a solid foundation for understanding the concepts technologies, and methodologies relevant to the project. By examining previous studies, this chapter sets the stage for the subsequent chapters, highlighting the gaps in the current knowledge and identifying the research objectives and contributions of the project. The problem of Content Aware Image Retargeting has gained significant attention in recent years due to the increasing demand for flexible and adaptive visual content across various digital platforms. As digital media continues to evolve, the need for intelligent image manipulation techniques has become more pronounced. Traditional image resizing methods often fail to preserve the essential visual elements and overall aesthetic quality of an image when adapting it to different display formats. CAIR aims to address these challenges by developing algorithms and techniques that can automatically resize or reshape images while maintaining their key visual features and compositional integrity. This approach involves a delicate balance between preserving important content, maintaining visual coherence, and adapting to the target display constraints.

2.1 Technology Background:

The theoretical background of the application includes many concepts related to computer vision, here are some foundations:

2.1.1 Image Processing:

Before we jump into image processing, we need to first understand what exactly constitutes an image. An image is represented by its dimensions (height and width) based on the number of pixels. For example, if the dimensions of an image are 500 x 400 (width x height), the total number of pixels in the image is 200000.

Images is represented by *Width x Height* Matrices.

This pixel is a point on the image that takes on a specific shade, opacity or color. It is usually represented in one of the following:

- **Grayscale** - A pixel is an integer with a value between 0 to 255 (0 is completely black and 255 is completely white).
- **RGB** - A pixel is made up of 3 integers between 0 to 255 (the integers represent the intensity of red, green, and blue).
- **RGBA** - It is an extension of RGB with an added alpha field, which represents the opacity of the image.

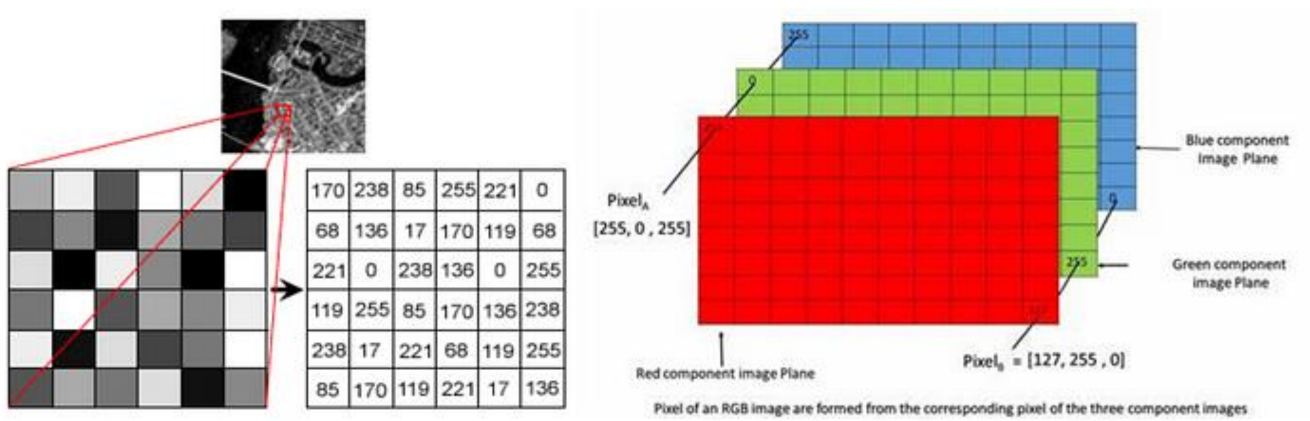


Figure 2.1-1 – Image Pixels

➢ Edge Detection:

Edge detection is a fundamental technique in image processing that identifies areas in an image where brightness changes sharply or has discontinuities.

Key principles of edge:

- Edges occur where there are significant local changes in image intensity
- These changes can be detected by examining gradients in the image.

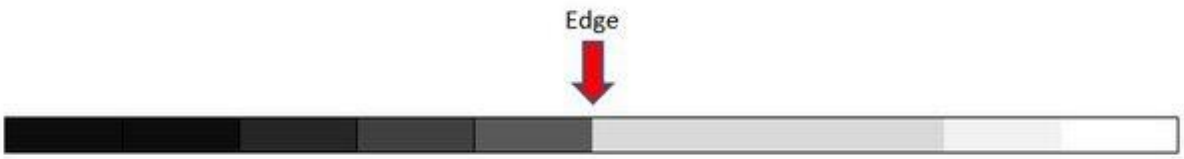


Figure 2.1-2 - Rapid Change Of Pixel Intensity

➢ Convolution:

A mathematical operation that combines two functions to produce a third function. In image processing, convolution is the process of multiplying each element of the image by a corresponding element of the kernel *(A kernel is just a fancy name for a small matrix.),* then summing these products over the kernel's range. This operation is performed by sliding the kernel over the image.

➢ Sobel Filter :

A popular edge detection algorithm involves estimating the first derivative of an image by doing a convolution between an image (i.e. the input) and two special kernels, one to detect vertical edges and one to detect horizontal edges.

Figure 2.1-3 - Sobel Edge Detector

2.1.2 Computer Vision:

Computer vision is a field of artificial intelligence (AI) that uses machine learning and neural networks to teach computers and systems to derive meaningful information from digital images, videos and other visual inputs—and to make recommendations or take actions when they see defects or issues.

- **Convolutional Neural Networks (CNNs):**

CNNs are the backbone of most deep learning models in computer vision. They use convolutional layers to automatically learn hierarchical features from images, from low-level edges and textures to high-level object parts and complete objects.

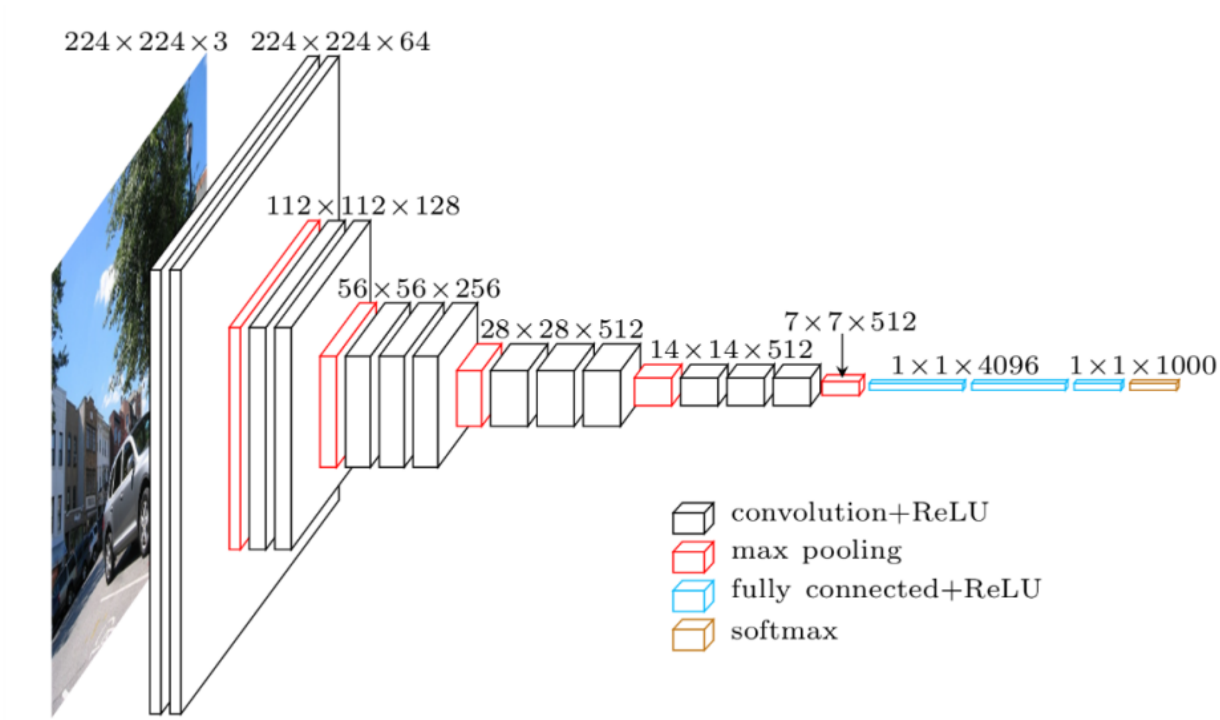


Figure 2.1-4 - CNN Architecture

➢ Image Segmentation:

Segmentation in computer vision refers to the task of partitioning an image into multiple segments or regions, each corresponding to a different object or part of the scene. It's a fundamental challenge in image understanding and has numerous applications.

- **Semantic Segmentation:**

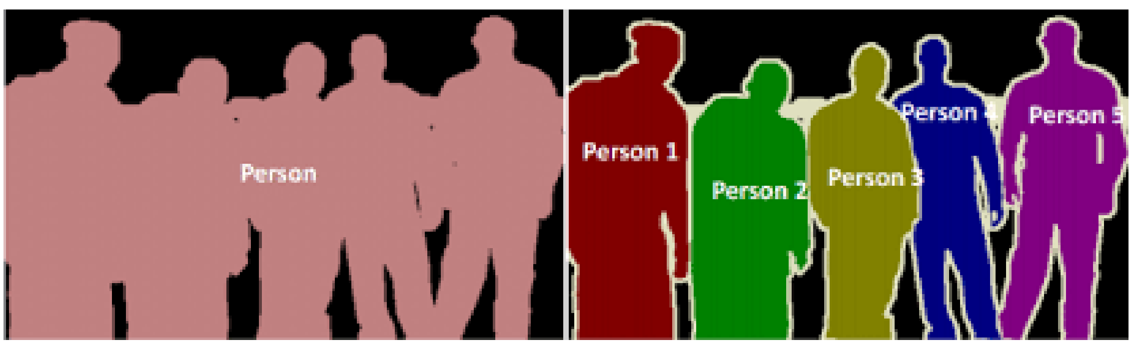
  Assigns a class label to every pixel in the image

  Doesn't distinguish between instances of the same class

- **Instance Segmentation:**

  Detects and delineates each instance of objects in the image

  Differentiates between multiple instances of the same class



(a) Semantic Segmentation (b) Instance Segmentation

Figure 2.1-5 - Semantic Segmentation Vs Instance Segmentation

- **Saliency Detection:**

  Saliency detection is an important task in computer vision that aims to identify the most visually important or attention-grabbing regions in an image. It's inspired by human visual attention mechanisms and has various applications.

Figure 2.1-6 - Examples Of Saliency Detection

- **Depth Estimation:**

  Depth estimation is a crucial task in computer vision that aims to infer the distance between the camera and objects in a scene from 2D images as shown in *figure 2.7*. This process is fundamental for understanding 3D structure and has numerous applications. Here's an overview of depth estimation:

  - Monocular Depth Estimation:
    - Estimates depth from a single image
    - Challenging due to inherent ambiguity, but more practical in many scenarios

Figure 2.1-7 -Depth Estimation Example

2.2 Related Papers

The field of image retargeting has seen significant advancements over the past two decades, with researchers developing a variety of approaches to address the challenges of content-aware image resizing. This section provides an overview of the key contributions and methodologies in the area, tracing the evolution from traditional resizing techniques to more sophisticated content-aware algorithms.

Table 1 - Abstract Of Related Work

Paper	Methods	Dataset	Quality Measures
Seam carving for content-aware image resizing [3]**.**	backward Seam carving DP for optimal seam identification	-	Visual Comparison
Improved seam carving for video retargeting [4]**.**	Forward seam carving GraphCut	-	Visual Comparison
Improved content aware scene retargeting for retinitis pigmentosa patien[5] **.**	Spatial Saliency map Importance map Motion Saliency map seam curving pixel shrinkability	-	Synthetic evaluation user-study
Image seam carving using depth assisted saliency map [6]**.**	Saliency Detection Seam carving	-	Visual Comparison
Hybrid image retargeting using optimized seam carving and scaling [7]**.**	Bi-Directional Seam Carving Non-linear Seam Carving Optimization	-	Objective quality metric testing using bidirectional similarity User study
Weakly- and Self-Supervised Learning for Content-Aware Deep Image Retargeting [8]**.**	Encoder-Decoder object detection (VGG16) seam carving content loss structure loss	pascal VOC 2007	User study Qualitative comparison
Content-Aware Retargeted Image Quality Assessment [9]**.**	CNN SIFT Seam Curving manual cropping content loss	MIT Retarget Me CUHK	User study Edge Histogram Earth moving distance Signal error

Paper	**Methods**	**Dataset**	**Quality Measures**
Cycle-IR-Deep Cyclic Image Retargeting [10]**.**	VGG16 CNN Spatial and Channel Attention Layer Cyclic Perception Coherence loss	HKU-IS	Cycle Perception Coherence Loss -for the model to optimize itself. user study
Self-Play Reinforcement Learning for Fast Image Retargeting [11]	Self-Play-based Reward reinforcement learning method.	MIRFLICKR-1M dataset RetargetMe dataset	Bi-Directional Warping (BDW)
Object-Aware Image Retargeting Using PSO and Aesthetic Quality Assessment [12]**.**	Image Inpainting Seam Carving Image Super-Resolution	PASCAL VOC 2007 dataset	Aesthetic and normal quality assessment pretrained networks
Saliency-aware Stereoscopic Video Retargeting [13]**.**	Co-Saliency detection (CoSD) + YOLOv5 Stereo Video Transformation model	KITTI Stereo 2015, contains 194 frame pairs of traffic scenarios	Bidirectional Similarity between the source and target videos
Supervised DL for Content-Aware IR with FFC [14]**.**	Deep learning comprising a mask generator convolutional neural network with Fast Fourier Convolution	CUB dataset, housing 11,788 bird images in 200 categories	Quality is assessed using High Receptive Field Perceptual Loss (HRFPL), adversarial loss, and a final loss function combining various components.

2.2.1 Seam Carving Based Methods

The backbone of *Content-Aware Image Retargeting* (CAIR) was proposed in the revolutionary paper "Seam Carving for Content-Aware Image Resizing" by Avidan and Shamir [3] introduced a novel approach to image resizing that diverged from traditional methods such as cropping and scaling. Their proposed algorithm, known as seam carving, represented a significant advancement in content-aware image resizing (CAIR) techniques.

Seam carving operates by intelligently identifying and removing or inserting paths of pixels called "seams" that traverse the image from top to bottom or left to right. This method allows for the alteration of image dimensions while preserving key visual elements, effectively maintaining the image's semantic content and overall composition.

Since its introduction, seam carving has inspired numerous extensions and improvements in the field of computational photography and computer vision. Researchers have explored optimizations to enhance its efficiency [4], [15], developed more sophisticated energy functions like in [6].

The original seam carving method proposed by Avidan and Shamir [3] often referred to as "backward" seam carving. This method calculates the energy of each pixel and then finds the optimal seam to remove based on this energy map. The problem with this method is that it doesn’t consider the impact of removing a seam on the rest of the image, which can lead to some artifacts!

Therefore, Avidan and Shamir with Rubinstein in “Improved Seam Carving for Video Retargeting” [4] , introduced a forward seam carving as an optimization of the original method. The key differences are:

- Look-ahead: It considers the impact of removing a seam on the rest of the image, which can lead to better results in many cases.
- Computational efficiency: In some scenarios, forward seam carving can be more efficient, especially when applied to videos.

Many methods were created based on seam carving mixed with other methods or with slightly different change like *multi-op* [16] is a method in CAIR using traditional methods (resizing and cropping) with seam carving to resize the image efficiently.

Figure 2.8 is an example of seam carving which the read lines, first we identify it then remove them till we reach the wanted ratio.

Figure 2.2-1 - Example Of Seam Detection And Seam Removal

While the seam carving algorithm itself was revolutionary, the quality of results heavily depends on the energy function used to create the energy map. Recognizing this, several researchers have focused on developing more sophisticated and effective energy functions [6], [13].

One significant contribution in this area came from Achanta and Süsstrunk in their paper "*Saliency Detection for Content-Aware Image Resizing*" [17]. They proposed a frequency-tuned approach to compute saliency maps, which can be used as energy maps for seam carving. Their method aimed to highlight visually important regions more accurately, leading to better preservation of significant image content during resizing.

Another notable approach was presented in "Image Seam Carving using Depth Assisted Saliency map" [6] in which the authors proposed a method to compose the gradient map, saliency map & depth map for better seam carving in which the seam won't cross the important object.

2.2.2 Deep Learning Based Methods

The application of deep learning to Content-Aware Image Resizing (CAIR) presents a unique set of challenges and opportunities for researchers. While deep learning has shown remarkable success in various computer vision tasks, its impact on CAIR has been somewhat constrained due to a fundamental issue: the lack of a standardized dataset with original images and their ideal resized counterparts.

Following the challenges posed by the lack of a standardized dataset, researchers have explored diverse applications of deep learning in CAIR, each addressing different aspects of the resizing process. These approaches can be categorized into several key areas:

- **Maps Generation:**

  To provide more visual information in generation of the energy function that will be used in seam carving for seam removal/addition like in [6], [13]

  - **Depth Map Estimation**: Deep learning models have been employed to generate depth maps, providing additional spatial information to guide the resizing process.
  - **Saliency Detection:** Neural networks have been used to create more accurate saliency maps, helping to identify visually important regions in images.
  - **Instance Segmentation**: Deep learning-based instance segmentation has been utilized to precisely locate and delineate objects within images, informing content-aware resizing decisions.

- **Algorithm Optimization:**

  In “Self-Play Reinforcement Learning for Fast Image Retargeting” by Kajiura, Kosugi, Wang, and Yamasaki [11] , They proposed a method that can achieve ultra-fast multi-operator image retargeting using reinforcement learning agent, their model is three orders of magnitude faster than normal multi-op, they trained using *MIRFLICKR-1M* dataset and ***RetargetMe*** dataset for testing.

- **End-to-End Approaches:**

  In “Cycle-IR-Deep Cyclic Image Retargeting” by Tan, Yan, Lin, and Niu [18] they build a cyclic model its backbone is a pre-trained VGG16, has 3 conv. Layer, and Spatial and channel attention layer. the workflow given *Image* and *ratio*): and then the image goes through two stages, and the output of each stage is feed again to the other one, they used a Cycle perception coherence for the model to optimize its result. They used *HKU-IS* dataset for training, and ***RetargetMe*** dataset for testing & user study.

2.2.3 Quality Measures Based Methods

Most of the paper used User study[5], [7], [8], [9], [18], as a ***Quality Method*** in which they bring multiple of people and show them the output Images from ***RetargetME*** benchmark, which is a dataset that has 78 Image, and the results of different Algo for this image, then they compare the result of their method with the other by people opinions.

Some papers [7], [8], [11], [13], [18] applied with the user study a numerical calculation as a ***Quantitative Method*** to get the content loss, Structural Similarity Index, Edge histogram, Peak Signal-to-Noise Ratio, and Composite Score for Measuring Image Similarity.

- **CAIR Qualitative Measure**

In image retargeting research, many papers use the RetargetME benchmark dataset for quality assessment through user studies. This dataset contains 78 images and their retargeted versions from various algorithms. Researchers typically gather a group of people and show them the original images alongside retargeted versions from different methods, including their own. Participants compare and rate these images based on factors like content preservation and visual appeal. By analyzing these opinions, researchers can compare their method's performance against existing algorithms in terms of user preference. This approach provides valuable insights into the perceived quality of retargeting techniques, complementing more technical evaluations and allowing for standardized comparisons across different studies in the field.

➢ CAIR Quantitative Measures

• Structural Similarity Index (SSIM):

SSIM is used for measuring the similarity between two images. It is a full reference metric, in other words, the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as reference.it is also considered as perception-based model that considers image degradation as perceived change in structural information [19], [20], while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms. The difference with other techniques such as MSE or PSNR is that these approaches estimate absolute errors. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. Luminance masking is a phenomenon whereby image distortions (in this context) tend to be less visible in bright regions, while contrast masking is a phenomenon whereby distortions become less visible where there is significant activity or "texture" in the image

Algorithm:

The SSIM index is calculated on various windows of an image. The measure between two windows x and y of common size N X N is with:

$$SSIM(X,Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

- μ_x the pixel sample mean of x
- μ_y the pixel sample mean of y
- σ_x^2 the variance of x
- σ_y^2 the variance of y
- σ_{xy} the covariance of x and y
- C_1= $(k_1L)^2$, C_2= $(k_2L)^2$ two variables to stabilize the division with weak denominator.
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$)
- K_1=0.01 and K_2=0.03 by default.

We used SSIM to measure the similarity between the original image and the retargeted image, focusing on structural information preservation, Higher SSIM values indicate better preservation of structural details, which is crucial for the visually impaired to comprehend the image.

- ## Content loss:

Content loss is a crucial concept in image processing, particularly as a quality metric for evaluating the similarity between images was used in [8]. It is widely used in various applications such as image restoration, super-resolution, and style transfer. it measures the difference between two images in terms of their content rather than their pixel-by-pixel differences [20]. The aim is to ensure that the primary features and structures of the images remain consistent. This is particularly useful in tasks where the perceptual similarity is more important than exact pixel matching, such as in artistic style transfer or super-resolution, where the goal is to produce visually pleasing results that retain the original image's content.

Algorithm:

Content loss is often calculated using deep learning models, particularly convolutional neural networks (CNNs), which can capture high-level features in images. A common approach is to use a pre-trained CNN, such as VGG-16 or VGG-19, to extract feature representations from specific layers. The content loss is then computed as the difference between these feature representations of the original and the processed image. Mathematically, if $\emptyset$ denotes the CNN, x the original image, and y the processed image, and $\emptyset_l$ represents the feature map from the l_{th} layer of the CNN, then the content loss $Lcontent$ can be defined as: $L_{content}(x,y) = \frac{1}{2}\sum_{i,j}\left(\emptyset_l(x)_{i,j} - \emptyset_l(y)_{i,j}\right)^2$

Here, i and j index the elements of the feature map. We used content loss to evaluate the loss of important visual information during the retargeting process. This metric is crucial for measuring similarity between the original and retargeted images, particularly in terms of recognizing and preserving essential details. Content loss can be assessed through:

- **Feature Matching**:
  Comparing key features between the original and retargeted images to identify lost or altered content.
- **Semantic Analysis**:
  Evaluating the preservation of semantic information, such as objects, faces, and text, to ensure critical information remains intact.

- **Region-Based Assessment**:
  Analyzing specific regions of the image that are crucial for understanding the overall context and ensuring minimal content loss in these areas. Quantifying content loss helps in understanding the impact of retargeting on the visual integrity of the image, ensuring that the retargeted image maintains a high degree of similarity to the original in terms of essential information.

- ## Edge histogram:

An edge histogram represents the frequency and orientation of edges in different parts of an image [21]. It provides a concise summary of the image's structural features, which are crucial for understanding its content and context.

Algorithm:

- **Edge Detection**:
  The first step in creating an edge histogram involves detecting edges within the image. This can be achieved using edge detection algorithms such as Sobel, Canny, or Prewitt operators. These algorithms highlight areas with significant intensity changes, marking the locations of edges.

- **Histogram Creation**:
  Once edges are detected, the image is divided into sub-blocks, and the edges within each block are categorized based on their orientation (e.g., horizontal, vertical, diagonal). The frequency of each edge orientation is then recorded, forming the edge histogram.
- **Histogram Distance Calculation**:
  The distance between two normalized histograms is computed using the Euclidean distance: $\text{distance} = \lVert \text{hist1} - \text{hist2} \rVert$

  This measures the dissimilarity between the edge histograms of the two images.

We used edge histogram in evaluating image similarity and content preservation during retargeting processes:

- **Similarity Measurement**:
  By comparing the edge histograms of the original and retargeted images, we can quantify the similarity between them. A high degree of similarity in edge histograms indicates that the structural features have been well-preserved, implying minimal content loss.
- **Content Preservation**:
  Edge histograms help in ensuring that critical edges, which often correspond to important content such as object boundaries and text, are maintained during the retargeting process. This is particularly important for visually impaired users who rely on clear structural information for image interpretation.

- **Peak Signal-to-Noise Ratio (PSNR):**

The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric to evaluate the quality of a reconstructed or compressed image compared [22]to its original version. It is especially prevalent in fields such as image processing, video compression, and image transmission, where the goal is to preserve the original image's fidelity while minimizing the amount of data required.

Algorithm:

PSNR is calculated using the Mean Squared Error (MSE) between the original image and the reconstructed image. The MSE measures the average squared difference between pixel values of the two images. The formula for MSE is:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

where:

- $I(i,j)$ is the pixel value at position (i,j) in the original image.
- $K(i,j)$ is the pixel value at position (i,j) in the reconstructed image.
- m and n are the dimensions of the images.

Once MSE is computed, PSNR is calculated using the following formula:

$$PSNR = 10\ \log_{10}\left(\frac{MAX^2}{MSE}\right)$$

where MAX is the maximum possible pixel value of the image. For an 8-bit image, MAX is 255. Interpretation:

- **Higher PSNR**:
  Indicates that the reconstructed image is closer to the original, implying better quality. Typically, a PSNR value above 30 dB is considered good in lossy image compression.
- **Lower PSNR**:
  Indicates more significant differences between the original and reconstructed images, implying lower quality.

We used PSNR to evaluate the quality of the retargeted image by quantifying the level of distortion relative to the original image.

- ## Scale-Invariant Feature Transform (SIFT):

A computer vision algorithm is used to detect and describe local features in images. SIFT is widely utilized for tasks such as object recognition, image stitching, and 3D modeling due to its robustness in various transformations like scaling, rotation, and partial occlusion. The SIFT algorithm includes a quality metric to measure the similarity between images, which is crucial for matching keypoints across different images.

Algorithm:

- **Feature Extraction:**
  Extract SIFT features from both images. This involves identifying keypoints and computing their descriptors.
- **Descriptor Matching:**
  Compare the descriptors from the two images. This is usually done using a distance metric, such as Euclidean distance, between the high-dimensional vectors. The descriptor pairs with the smallest distances are considered the best matches.
- **Ratio Test:**
  To improve the robustness of matching, Lowe proposed a ratio test. For each keypoint descriptor in the first image, find the two closest descriptors in the second image. If the ratio of the distance to the closest neighbor and the distance to the second closest neighbor is below a certain threshold (typically 0.8), the match is considered a good one. This helps to eliminate false matches.

Equations:

- **Descriptor Matching**:
- The Euclidean distance between two descriptors di and dj is calculated as:

$$\text{distance}(d_i, d_j) = \sqrt{\sum_{k=1}^{n} \left(d_i[k] - d_j[k]\right)^2}$$

Where n is the dimensionality of the descriptor vectors.

- **Ratio Test:**
  For each keypoint descriptor d_1 in the first image, find the two closest descriptors d_2 and d_3 in the second image. The ratio test is applied as:
  $\frac{\text{distance}(d_1, d_2)}{\text{distance}(d_1, d_3)} < 0.8$

- **Similarity Percentage:** $percent = \left(\frac{|good\ matches| \times 100}{\min(|Keypoints\ in\ image1|, |Keypoints\ in\ image2|)}\right)$

• Composite Score for Measuring Image Similarity

To accurately assess the similarity between two images, relying on a single quality metric might not be sufficient due to the multifaceted nature of image features. A composite score, which aggregates multiple quality measures, can provide a more comprehensive and robust evaluation of similarity. To calculate a composite score, the individual quality measures need to be normalized and combined by Assigning weights to each quality measure based on their relative importance or reliability in capturing image similarity. The choice of weights can be empirical or derived through techniques like machine learning.

The composite score can be computed as:

$$S_{composite} = (0.5 \times SSIM) + \left(0.4 \times \frac{1}{CL}\right) + (0.3 \times SIFT) + \left(0.2 \times \frac{1}{HIST}\right) + (0.1 \times PSNR)$$

By integrating multiple content loss quality measures into a composite score, we can achieve a more accurate and reliable assessment of image similarity. This approach balances the strengths and weaknesses of individual measures, leading to a more comprehensive evaluation.

2.2.4 General Review

Content-Aware Image Retargeting (CAIR) has evolved significantly since the introduction of seam carving by Avidan and Shamir. This field has seen various advancements, including improvements to the original seam carving algorithm, development of more sophisticated energy functions, and the application of deep learning techniques. While seam carving remains a fundamental approach, researchers have explored combinations with other methods like multi-op to enhance resizing efficiency.

The challenges in applying deep learning to CAIR, primarily due to the lack of standardized datasets, have led researchers to focus on specific aspects such as map generation for depth, saliency, and instance segmentation. Some innovative approaches, like reinforcement learning for fast retargeting and cyclic models for end-to-end solutions, show promise in advancing the field.

Quality assessment in CAIR research typically relies on user studies, often utilizing the ***RetargetMe*** benchmark dataset. These are frequently complemented by numerical evaluations of content loss, structure loss, and coherence loss.

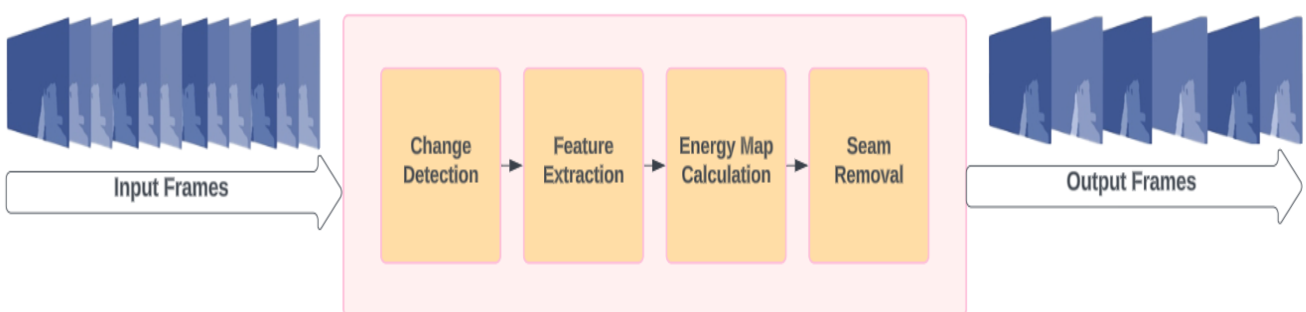
While significant progress has been made, the field of CAIR continues to evolve. Deep learning approaches, although promising, have not yet produced consistently superior results compared to traditional methods. As research progresses, the integration of advanced algorithms, more sophisticated energy functions, and machine learning techniques may lead to further improvements in content-aware image resizing, potentially enhancing visual accessibility and user experience across various applications.

3. System Architecture and Methods

3.1 Abstract Overview

The system processes image frames through a series of steps handled by specialized workers:

1. **Feature Extraction**: Detects changes in the image frames and extracts features from the detected changes.
2. **Energy Map Calculation**: Calculates an energy map based on the extracted features.
3. **Seam Removal**: Removes seams from the image based on the energy map.



3.1-1 - System Architecture

3.2 Detailed Descriptions:

3.2.1 Phase 1: Feature Extraction

The Feature Extraction phase detects significant changes between consecutive frames in a video and extracts relevant features from these changes. For the first frame, the entire algorithm is applied without needing to detect changes. For subsequent frames, the phase detects changes in the upper or lower regions of the frame and applies feature extraction algorithms such as edge detection, corner detection, or texture analysis. These extracted features are crucial for subsequent processing steps and help in accurately identifying significant elements in the image.

- **Task: Detect changes in incoming image frames and extract features.**
- **Input: Raw image frames.**
- **Process: Utilizes frame comparison to identify significant changes and applies feature extraction methods to the identified changes.**
- **Output: Frames with detected changes sent to Energy Map Calculation Queue.**

3.2.2 Phase 2: Energy Map Calculation

The Energy Map Calculation phase generates an energy map for the regions based on the extracted features. This map highlights the significance of different areas in the regions by assigning higher energy values to important parts. The energy map is essential for guiding the seam removal process, ensuring that critical parts of the image are preserved during resizing.

- **Task: Calculate an energy map based on extracted features.**
- **Input: Regions with extracted features from Energy Map Calculation Queue .**
- **Process: Computes an energy map that represents the importance of each pixel in the regions.**
- **Output: Energy maps for the regions sent to Seam Removal Queue .**

3.2.3 Phase 3: Seam Removal

The Seam Removal phase uses the energy map to determine which seams (paths of low energy) can be removed from the regions to achieve resizing while preserving important content. This phase employs seam carving techniques to adjust the image dimensions without significantly altering its visual integrity. The processed regions are then combined into a complete frame, ready for further use or display.

- **Task: Remove seams from the image based on the energy map.**
- **Input: Energy maps for the regions from Seam Removal Queue .**
- **Process: Identifies and removes the least important seams to resize or retarget the image.**
- **Output: Final processed regions combined into a complete frame.**

4. System Implementation

4.1 Datasets

4.1.1 RetargetMe

Description: The RetargetMe dataset is designed for research and applications in image retargeting. It offers a comprehensive collection of images and corresponding metadata for implementing and testing content-aware retargeting algorithms.

Contents:

- **Images**: A large collection serving as primary data for developing and testing image retargeting algorithms. **In our project**, ***these images are used as a source and not for training.***
- **Training Data**: The majority of files labeled with annotations such as regions of interest, energy maps, and seam removal data, crucial for training retargeting algorithms.
- **Testing Data**: A smaller portion reserved for assessing the effectiveness and accuracy of retargeting algorithms, ensuring the preservation of critical visual information during resizing.

4.1.2 ReDWeb-S-Modules

Description: The ReDWeb-S-Modules dataset is associated with the ReDWeb-S project. It contains various modules and resources essential for the project's web functionalities, supporting the development, testing, and deployment phases.

Contents:

- **Images**: A substantial collection of images used for various project purposes. **In our project, *these images are used as a source and not for training.***
- **Training Data**: A significant portion of labeled files used for training machine learning models or algorithms.

4.2 Image Preprocessing

Our image preprocessing pipeline builds on the approach described in [6] for energy map generation. We enhance this method by incorporating advanced deep learning models to improve accuracy and efficiency and we used a different merging technique. The following subsections detail our preprocessing steps, beginning with depth estimation.

4.2.1 Depth estimation

For our monocular depth estimation task, which involves estimating depth from 2D images, we evaluated two prominent models: MiDaS and Marigold. This section presents a comparison between these two models, assessing their performance and suitability for our project.

- **MiDaS:**
  - **Speed & Real-Time**: Fast, suitable for real-time applications.
  - **Accuracy**: High, performs well across various environments.
  - **Generalization**: Excellent, due to diverse training datasets.
  - **Applications**: Broad, including AR, robotics, 3D reconstruction, and autonomous driving.
  - **Implementation**: Easy with open-source availability and pre-trained models.

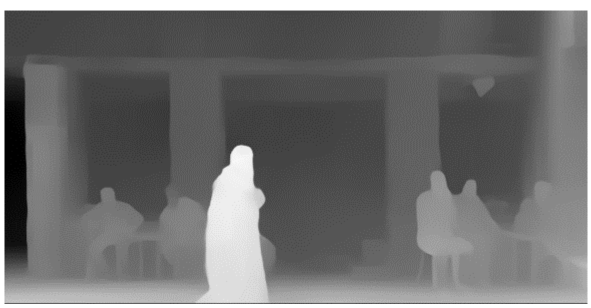


Figure 4.2-1 - Midas Result

➢ Marigold:

- **Speed & Real-Time**: Varies, may not be real-time.
- **Accuracy**: High in specific applications but less generalizable.
- **Generalization**: Depends heavily on training data, potentially less robust.
- **Implementation**: May have limited pre-trained models and documentation, with varying ease of integration.

Figure 4.2-2 - Marigold Result

The result of Marigold is indeed has more details but is slower than MiDaS, so we choose ***MiDaS model*** .as it's a fast model and the missing details we could recover them using the next subsequent image processing steps.

4.2.2 Saliency Map Detection

For our salient object detection task, which involves identifying and highlighting prominent objects within images, we evaluated two prominent models: TRACER and U^2-Net. This section presents a comparison between these two models, assessing their performance and suitability for our project.

- **TRACER:**
  - **Speed & Real-Time:** Moderate, with some latency due to complex computations.
  - **Accuracy:** Very high, excels in capturing fine details and complex patterns.
  - **Generalization:** Excellent, with robust performance across diverse datasets and environments.
  - **Implementation:** More complex, requiring significant computational resources and expertise in transformer-based models.

➢ U^2-Net:

U^2-Net, also known as U-Square Net, is an innovative deep learning architecture designed for salient object detection. Introduced in 2020, The architecture's name comes from its unique nested U-structure design – essentially a U-Net within a U-Net. This novel approach allows U^2-Net to efficiently extract and process multi-scale features without relying on a pre-trained backbone network.

- ### Key Features
  - **Speed & Real-Time:** Fast, highly suitable for real-time applications.
  - **Accuracy:** High, performs well in various scenarios with good balance between speed and accuracy.
  - **Generalization:** Very good, benefiting from its multi-scale feature extraction.
  - **Implementation:** Relatively easy with extensive documentation and pre-trained models available.

- ### Architecture

  Based on U-Net architecture

  ▪ **Main Components:**
  - Large U-structure (outer U)
  - Small U-structures (inner U's)

  ▪ **Encoder:**
  - Series of down-sampling blocks
  - Each block contains small U-structures

  ▪ **Decoder:**
  - Series of up-sampling blocks
  - Also contains small U-structures

  ▪ **Skip Connections:**
  - Connect corresponding encoder and decoder levels

  ▪ **Multi-scale Feature Fusion:**
  - Combines features from different scale.

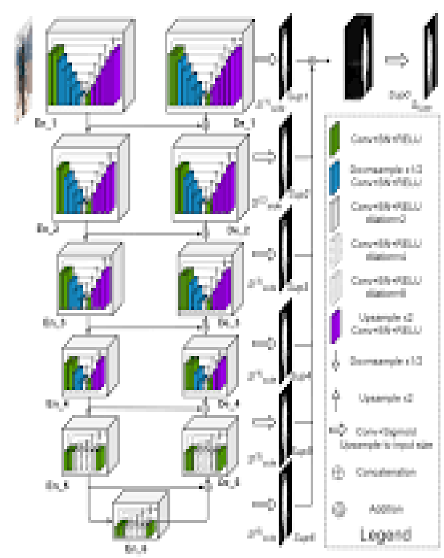


Figure 4.2-3 – U^2 Net Architecture

4.2.3 Instance segmentation.

➢ **Mask RCNN**

mask R-CNN is an extension of the Faster R-CNN model, designed for object detection and instance segmentation tasks. It adds a branch for predicting segmentation masks on each Region of Interest (RoI) in addition to the existing branches for classification and bounding box regression.

- **Key Features**
  - **Architecture:** Builds on Faster R-CNN by adding a parallel branch for mask prediction.
  - **Segmentation Masks:** Provides pixel-level segmentation for each detected object, allowing for precise delineation.
  - **Accuracy:** High accuracy in both object detection and segmentation, capable of handling complex scenes with multiple objects.
- **Architecture**

  As shown in figure 4.5, the architecture consists of:
  - **Backbone Network**:
    - CNN like ResNet, which extracts features from the input image
  - **Region Proposal Network** (RPN):
    - Generates region proposals (potential object locations)
  - **RoI Align:**
    - Extracts fixed-size feature maps from each proposed region
  - **Object Detection Branch:**
    - Classifies objects and refines bounding boxes
  - **Mask Prediction Branch**:
    - Generates a binary mask for each detected object

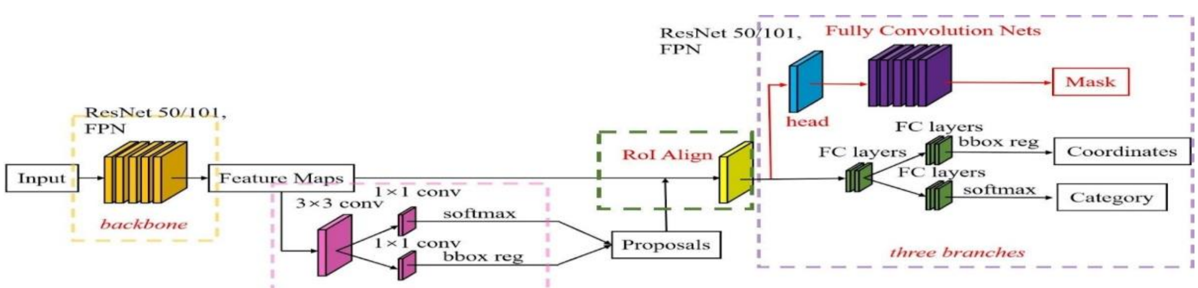


Figure 4.2-4 - Mask Rcnn Architecture

We integrated Mask R-CNN for instance segmentation to improve our composition importance model. Objects are filtered by confidence scores and depth. We calculate mean depth within each object's bounding box, retaining those exceeding a threshold as foreground elements.

➢ Edge detection using Sobel filter.

We generate the Gradient map/Edge map to capture the fine details so that it affects the calculations of seam carving due to its importance for the image structure.

➢ Multi-Scale Fusion for Combining Different Images

Multi-scale fusion is a technique used to combine multiple images, enhancing the final output by integrating information at various scales. To capture the fine details in small scales and the structure in large ones.

- **Calculation Details**
  - Decompose input image into multiple scales or levels.
  - Extract relevant features or characteristics at each scale for all images.
  - Create weight maps for each scale of each image based on quality metrics.
  - apply fusion rules to combine information from all images at each scale.
  - Verify and ensure consistency of fusion results across scales.
  - Reconstruct the final fused image from the combined multi-scale information.

In our experimentation, we opted for a 5-level scale in the Multi-Scale Fusion (MSF) process.

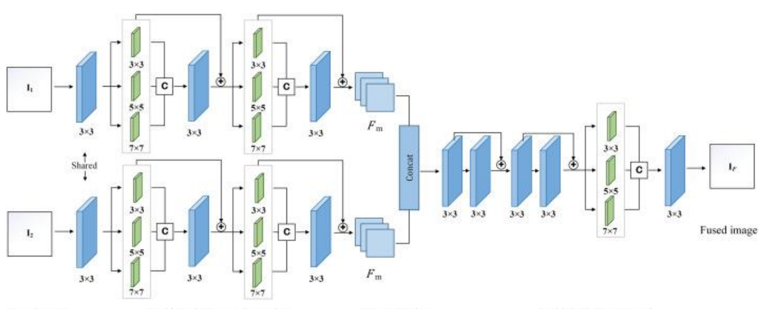


Figure 4.2-5 - Multi-Scale Fusion Process

The process for obtaining our Composition Importance Image using Saliency map, Depth map and Gradient map was inspired from [6], we added foreground instance segmentation for better results, and we used different approach for merging the images using multi-Scale-Fusion with depth 5 and this method get us better energy map. We used models like **MaskRCNN**, **MiDaS** and **U2net** to obtain the maps for better and faster results. However, through experiments with different images we encountered an issue where objects in close proximity were at risk of merging due to seam removal occurring near their boundaries. To mitigate this, we implemented a protective boundary around detected objects, effectively preserving their individual integrity in the composition.

You could see the whole Preprocessing flow in figure 4.7

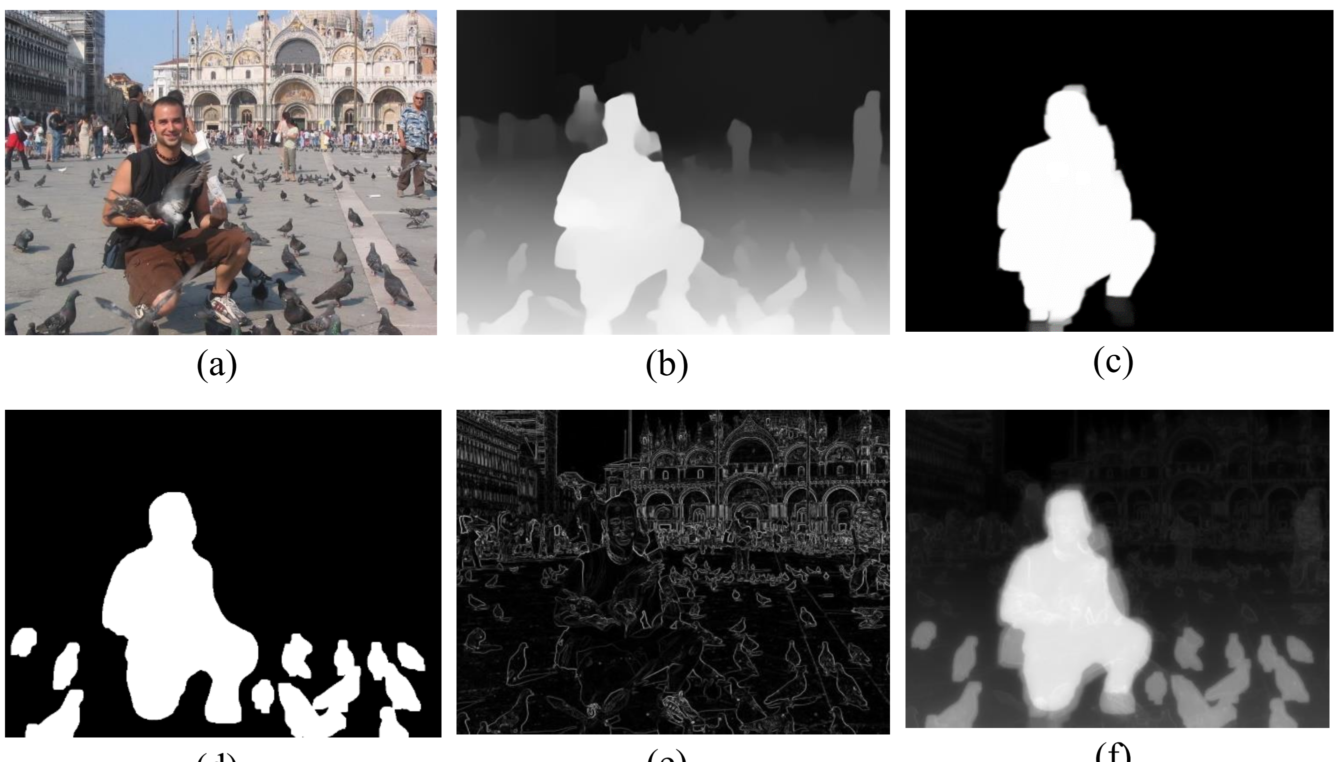


Figure 4.2-6 - Illustration Of Importance Maps: (A) Original Image, (B) Depth Map, (C) Saliency Map (D) Mask RCNN, (E) Gradient Map (F) Image Combined With Multi-Scale Fusion

4.3 Seam Carving Algorithm

After extracting the importance map E(x,y) as shown in Figure 4.7, the next step is to resize the image to a specified aspect ratio. This involves the seam carving algorithm, which uses the importance map to determine the optimal seams to remove or add. A seam is defined as a path of pixels with the least cumulative energy from top to bottom (for width reduction) or from left to right (for height reduction). By iteratively identifying and removing these low-energy seams, the algorithm reduces the image's width or height while preserving critical features and minimizing distortion.

4.3.1 Energy Calculation

The first step is to compute the cumulative energy map M (x, y) in a top-down manner using the importance map E (x, y) as a reference. This is done by memorizing the cumulative energy for each pixel, where the value at each pixel represents the total energy of the least energy path leading to that pixel.

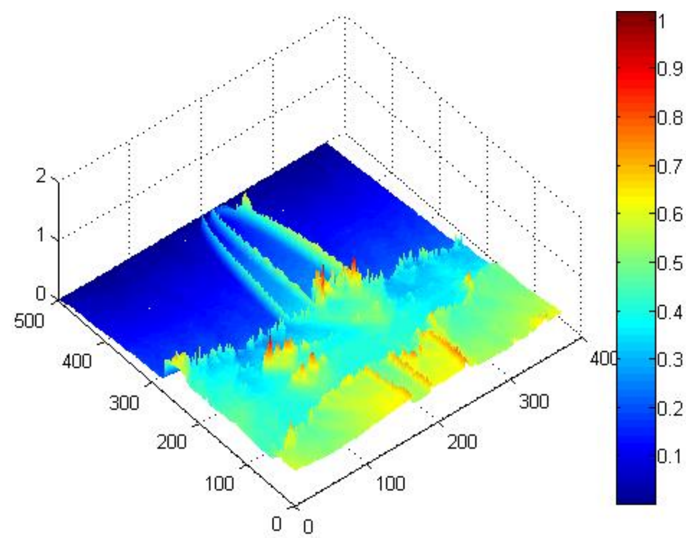


Figure 4.3-1 - The Result Of The Energy Calculation Process.

➢ Backward Energy

Backward energy accumulation is a method used in the seam carving algorithm to determine the optimal seams for removal. It calculates the cumulative energy for each pixel by comparing it to the energy of pixels in the previous row (i.e., the row above). Specifically, for each pixel (x, y), the cumulative energy M (x, y) is determined by adding the pixel's energy $E(x, y)$ to the minimum cumulative energy of the three possible preceding pixels: upper left $M(x-1, y-1)$, upper middle $M(x,y-1)$, and upper right $M(x+1,y-1)$.

This approach finds the shortest path from the top of the image to the bottom, ensuring that the least important seams are identified for removal.

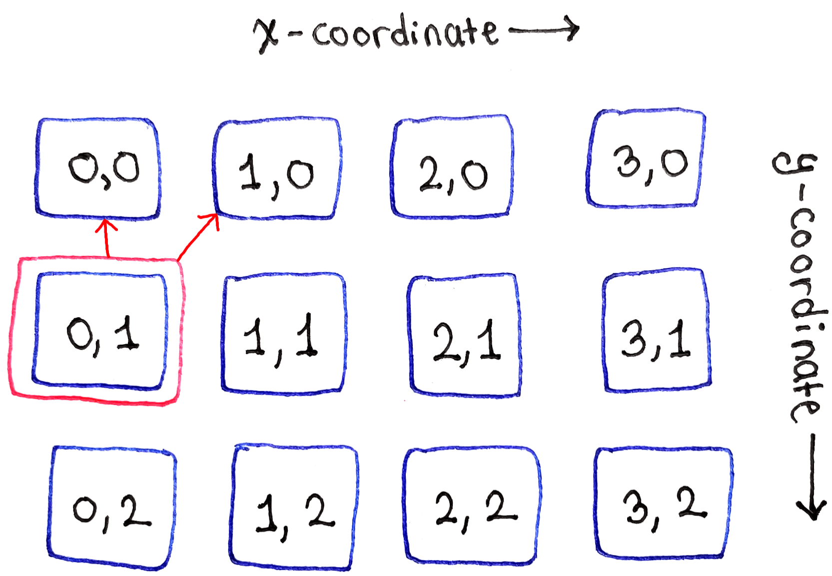


Figure 4.3-2 - Example Of The Backward Energy Calculation Approach

➢ Forward Energy

Forward energy accumulation is another method used in seam carving that evaluates the impact of seam removal more comprehensively. Instead of only looking at the previous row, this method also considers the future impact of removing a pixel. For each pixel (x,y), the algorithm calculates the forward energy F(x,y) by adding the pixel's energy E(x,y) to the minimum cumulative energy of the three possible succeeding pixels: lower left F(x−1,y+1), lower middle F(x,y+1), and lower right F(x+1,y+1).

This approach helps ensure that the selected seams for removal maintain the image's structural integrity and visual coherence, providing a more balanced and aesthetically pleasing result.

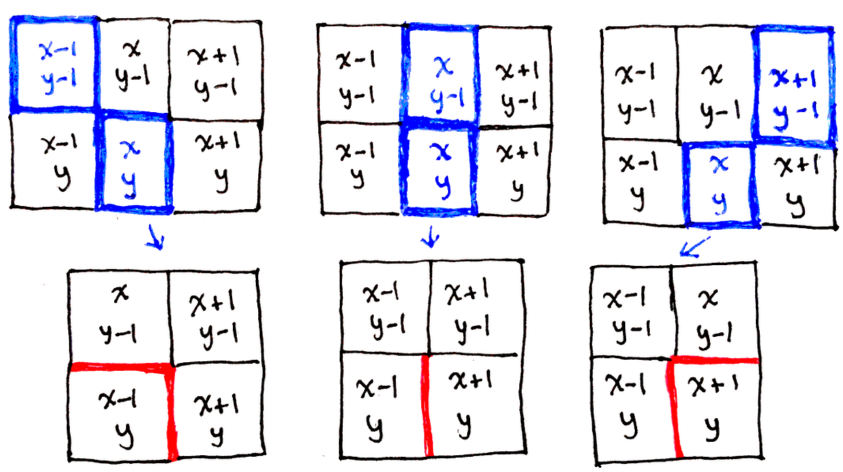


Figure 4.3-3 - Forward Energy Calculation

4.3.2 Seam Removal

The second step involves following a bottom-up approach to remove the calculated seams based on the cumulative energy map generated in the previous step.

- **Masking**

  This method involves marking the pixels to be removed with a unique value for easy identification later. After every three cycles, the marked pixels are removed from each row.

- **Iteration**

  The iterative approach is more straightforward. Starting from the bottom and moving up, each unnecessary pixel in the identified seam is removed, and the original image is replaced with a new image that is one pixel narrower. This process is repeated until the desired width is achieved.

Figure 4.3-4 - Seams Detection For Removal

4.3.3 Parallelization

The previous processes required significant power consumption and processing time, making them ineffective for real-time applications.To expedite the process, we divided the image into parts and processed each part independently.

- **Vertical Vectorization:**

  The initial approach was to split the image into approximately eight columns, each processed independently. The seams would then be removed from bottom to top. However, this approach was not feasible due to the need for synchronization between processes. Each column depended on the other portions of the image, requiring a shared environment and resulting in a lot of waiting. Consequently, this method consumed the same amount of time as without parallelization, making the process effectively linear.

- **Horizontal Vectorization:**

  We then opted to split the image horizontally in half, allowing each half to be processed independently. The energy accumulation for the top half worked from top to bottom, while the bottom half worked from bottom to top, converging at the middle. Using the middle row, we identified the least energy pixel as the starting point in both directions. We then began removing seams from bottom to top in the upper half and from top to bottom in the lower half.

  This method produced a connected seam that met in the middle, speeding up the process and enhancing the quality in some images. The seam now had more spatial range to traverse, which was particularly beneficial when the best seam was diagonal. Figures 4.2.4 shows an illustration of the process of accumulating towards the center.

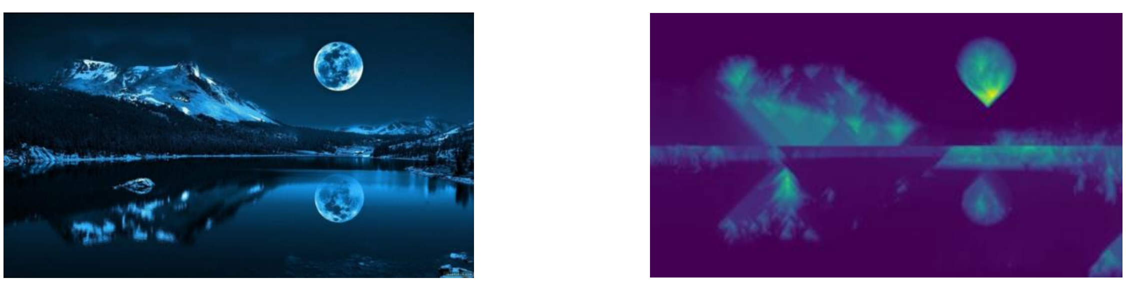


Figure 4.3-5 - Illustrate The Process Of Accumulating Towards The Center,

4.4 Video Retargeting

A crucial feature for patients with Retinitis Pigmentosa (RP) is the ability to adapt a video's size for various display formats. This adaptation allows them to experience visual content more freely and enjoy every moment with their loved ones. Numerous methods are available to achieve this, with the most effective ones focusing on understanding the intricate relationships between video frames. By optimizing these relationships, the viewing experience becomes more immersive and accessible for RP patients, enhancing their ability to engage with visual media.

4.4.1 Discrete Approach

This approach mentioned in [4] requires having the whole video in hand, where at each frame you could calculate the forward energy depending on the previous and the next neighbors for each pixel, by applying this you get a temporal vision of the accumulation at the last frame.

By going **backwards,** you can remove the seams from each frame depending on the extracted energy map. Its downside is that it can only be performed if you have the full video in hand, so in cases like real time apps it won't be of benefit since I don't possess the upcoming frame.

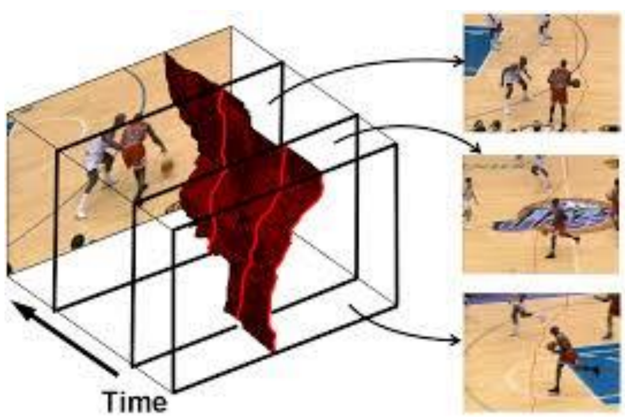


Figure 4.4-1 - Illustrating The Seam Removal Along The Temporal Dimension

4.4.2 Continuous Approach

This approach relies on building a pipeline, as illustrated in the architecture. Each frame is passed sequentially through multiple workers, with each worker performing a specific task. Crucially, the frames do not wait for the other frames to finish processing completely.

To establish a relationship between frames, we cache each frame for use in subsequent frames. By calculating the differences between neighboring pixels from the current and previous frames, we can achieve a similar effect to the previous approach.

Additionally, we can implement a Change Detection worker. This worker's job is to detect where changes occur. For example, if the bottom half of a frame remains unchanged in the next frame, we can reuse the previous computation instead of recalculating it. This helps to save processing time.

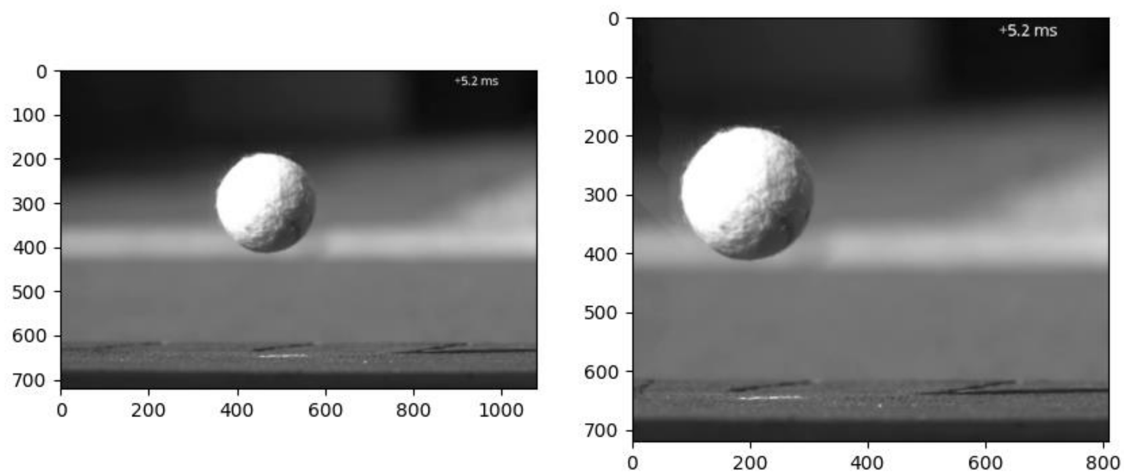


Figure 4.32 - Illustrating the retargeting of frames based on accumulative energy maps

4.5 Results:

The Results section presents and analyzes the outcomes of applying three distinct image retargeting algorithms: backward, forward, and forward middle. By showcasing before-and-after images, we aim to visually and quantitatively compare the performance and effectiveness of each algorithm. This section will highlight how each method handles the challenges of image retargeting, such as preserving important content, maintaining structural integrity, and minimizing visual distortions.

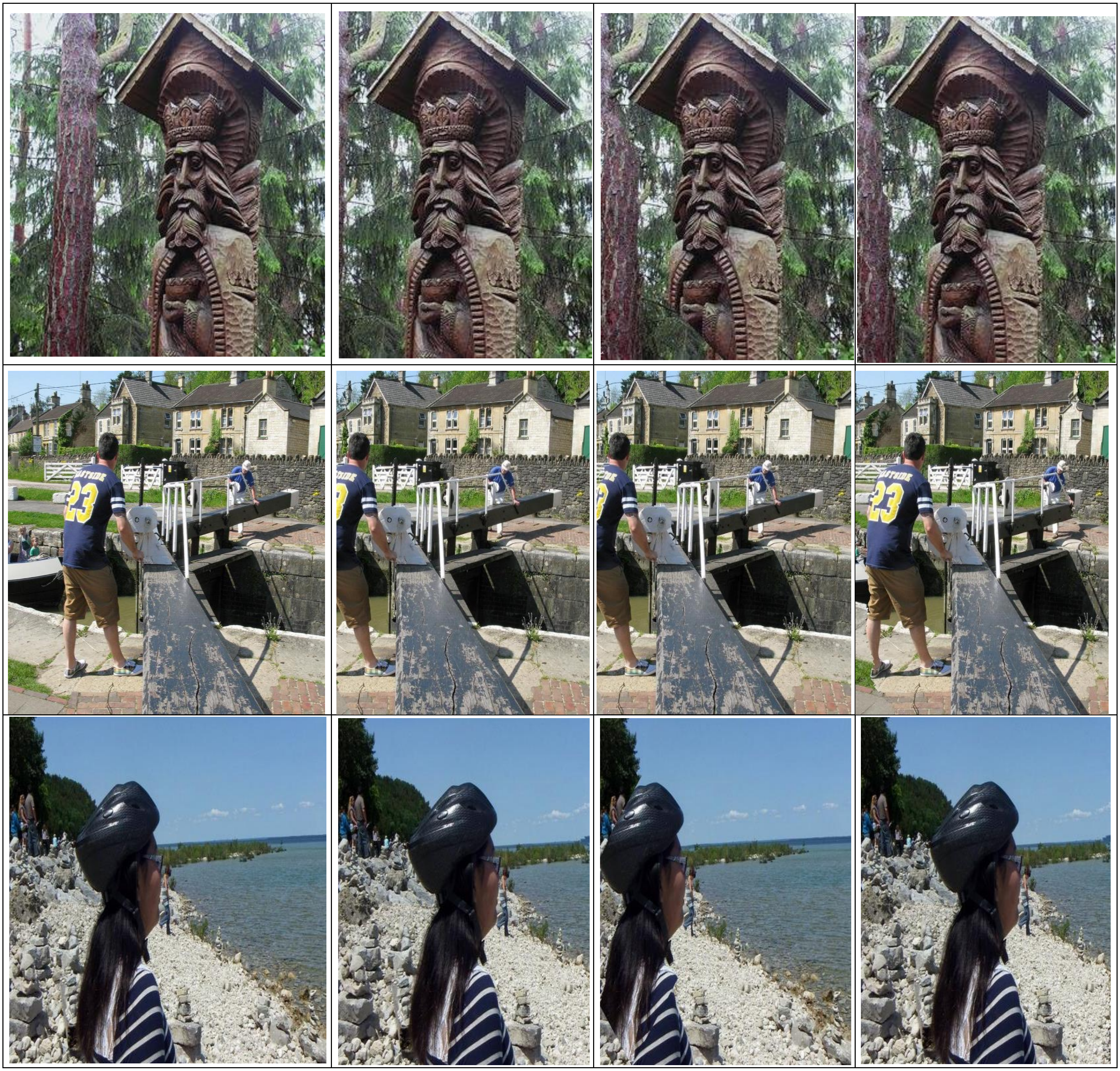
Before diving into the comparison, it is important to understand the objectives of image retargeting. The primary goal is to resize images while preserving important visual content and minimizing distortions. Each algorithm approaches this goal differently, and the results section will highlight these differences.

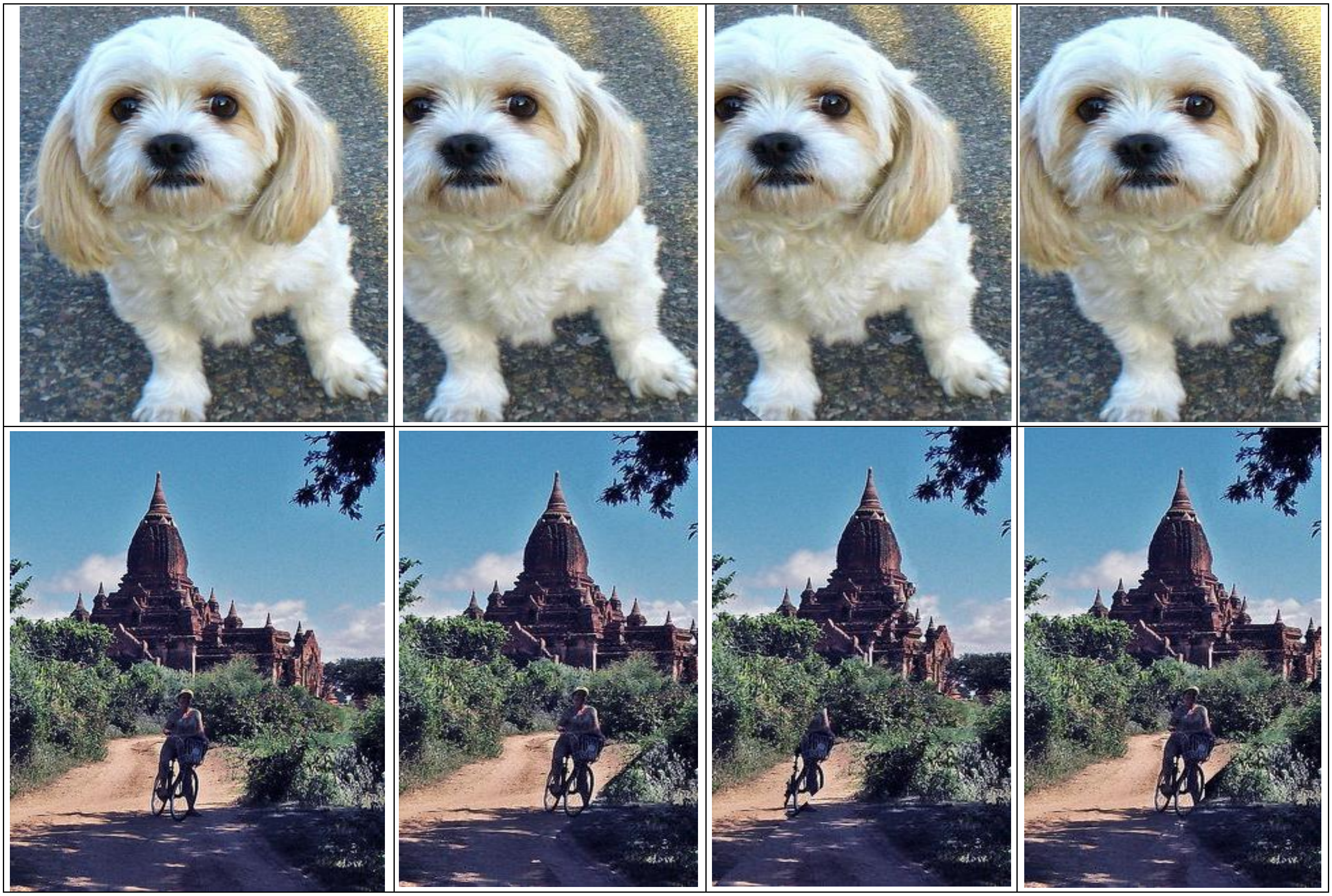
4.5.1 Visual Comparisons:

The following subsections display the original images and their retargeted counterparts for each algorithm. This visual documentation will provide a basis for understanding how each method modifies the images and highlight any noticeable artifacts or distortions.

Table 2 - Comparison Of The Results Of Backword, Forward Middle Algorithms

Original image	**Backward algorithm**	**Forward algorithm**	**Forward middle algorithm**
			





5. System Testing

5.1 Quality Measures:

In this section, we will evaluate the performance of different image retargeting algorithms, including the backward, forward middle (our algorithm) and the forward algorithms, as well as a library-based seam carving approach. Evaluation involves assessing both quantitative metrics and qualitative user feedback. These measures help ensure that the system meets users' needs and effectively enhances their visual experience.

5.1.1 Seam carving library:

Seam carving is a widely used technique for content-aware image.

retargeting. The seam carving library provides a robust implementation of this method, allowing for efficient and high-quality image retargeting.

- **Energy Function:**

The library uses a gradient-based energy function to identify the importance of each pixel. This helps in preserving important visual content while removing or inserting seams.

- **Seam Removal and Insertion:**

The library supports both seam removal for image reduction and seam insertion for image expansion

5.1.2 Visual representation:

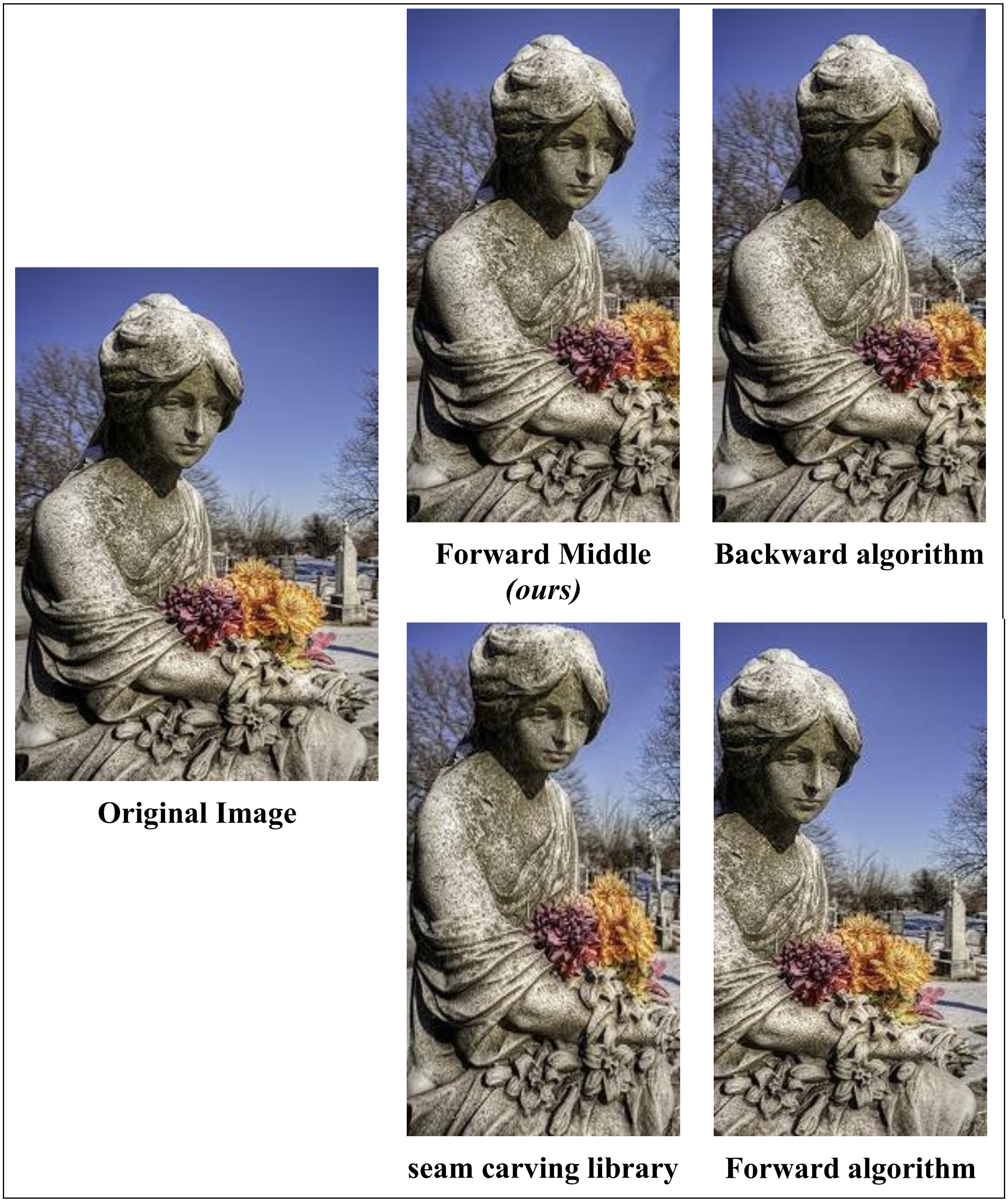
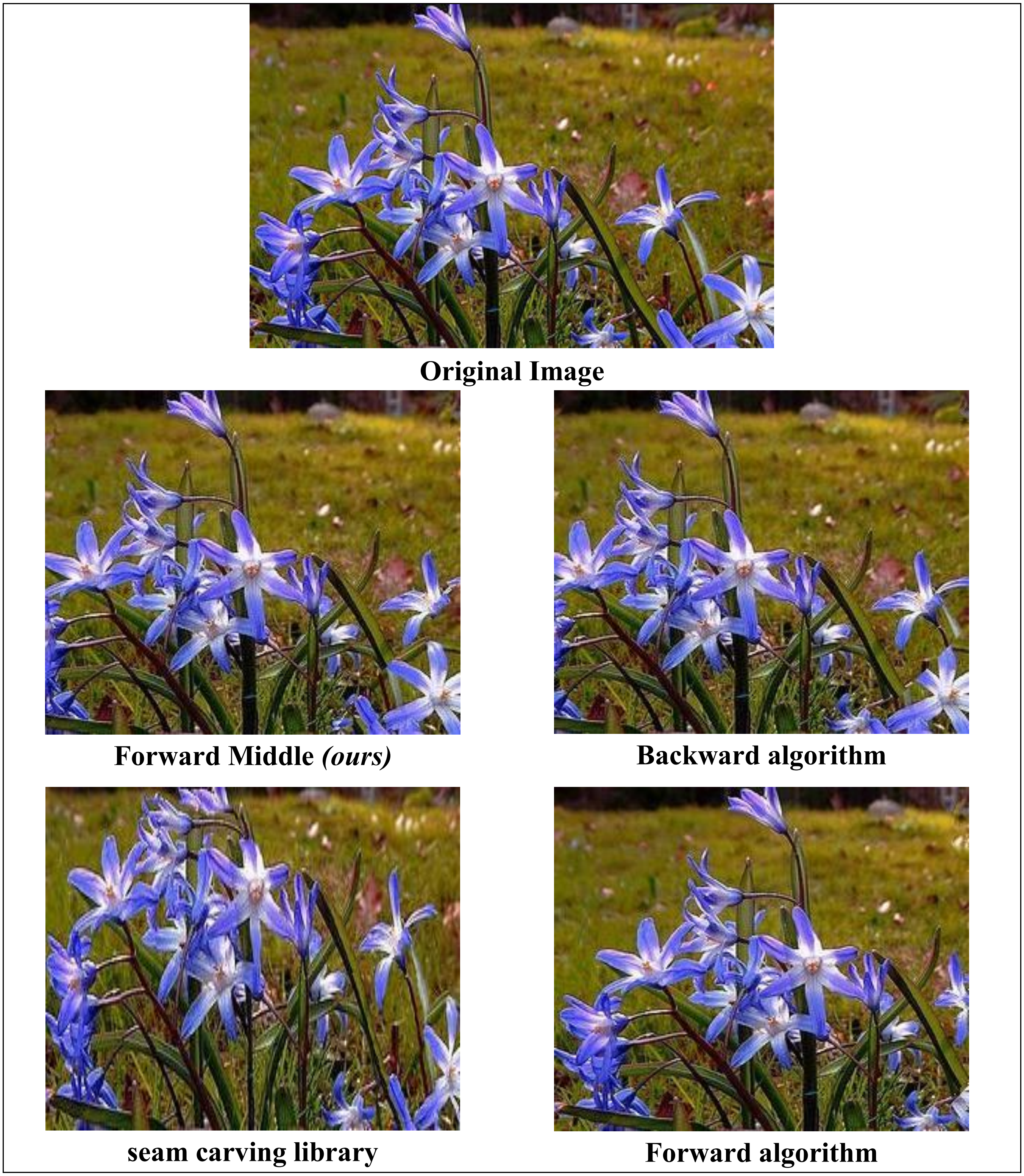


Figure 5.1-1 - Illustrating Of CAIR Result Of ***Figure 1****Using Different Methods Vs Ours*



*5.1-2 - - Illustrating Of CAIR Result Of **Figure 2** Using Different Methods Vs Ours*

5.1.3 Quantitative analysis:

In this section we use the quality metrices to compare between the different image retargeting algorithms that we tried and show some examples in the above figures

Table 3 - Example Of The Quality Metrix Score Of Figures 5.1.1

algorithm	content loss	SSIM	SIFT Similarity	EH	PSNR	score
Middle forward	0.150	0.701	91%	0.006	30.2	70
backward	0.141	0.605	88.5%	0.005	27.1	68
Seam carving library	0.172	0.224	50.6%	0.008	10.3	42
forward	0.159	0.430	82.9%	0.006	10.4	59

Table 4 – Another Example Of Quality Metrix Score Of Figures 5.1.2

algorithm	content loss	SSIM	SIFT Similarity	EH	PSNR	score
Middle forward	0.199	0.560	82%	0.003	10.9	80
backward	0.222	0.561	84%	0.005	10.8	67
Seam carving library	0.290	0.107	42.5	0.006	9.4	47
forward	0.272	0.608	82.9%	0.008	25.4	54

Based on the examples shown above, the quality metrices scores and the composite score, we found that the forward middle algorithm generally outperforms the backward, forward algorithms and the seam curving library in terms of balancing the preservation of important visual content and minimizing distortions. While the backward algorithm tends to introduce noticeable artifacts in areas with complex details and the forward algorithm often sacrifices peripheral content, the forward middle algorithm effectively maintains the structural integrity of the images and retains significant features.

5.1.4 Qualitative Quality Measures:

- **Survey Methodology:**
  - **Survey overview:**

    To assess the quality of image retargeting in our project for the visually impaired, we conducted a qualitative survey to evaluate perceived distortion and loss in the retargeted pictures compared to the original ones. Participants were presented with 10 pairs of images, each consisting of an original image (left) and a retargeted image (right).
  - **Survey question:**

    Participants were asked the following question for each pair of images:

    "How well do the edited images (right) look compared to the originals (left), without any distortions (?(تشوهات(Rate 1-5)"
  - **Survey execution:**

    The survey was distributed online, and participants were required to rate each image pair based on the perceived quality of the retargeted image in comparison to the original. The primary focus was on identifying any distortions present in the retargeted images.
  - **Mean Opinion Score (MOS) :**

The Mean Opinion Score (MOS) is a metric used to quantify the subjective quality assessments provided by survey participants. It is widely used in evaluating multimedia quality, including image retargeting. The MOS is calculated as the average of all ratings given by participants, providing a single score that reflects the overall perceived quality of the retargeted images.

Calculation of mean opinion score:

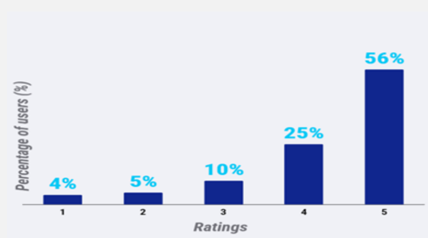
The MOS for each image pair is calculated using the following formula:

$$MOS = \frac{1}{N}\sum_{i=1}^{N} R_i$$

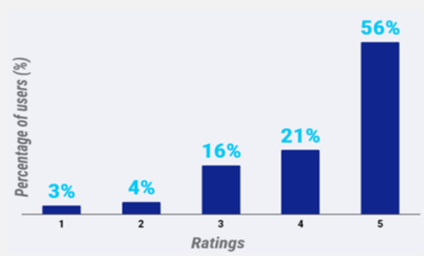
where N is the number of participants, and ￼ is the rating given by the i-th participant for a specific image pair. The final MOS is the average of the MOS values for all image pairs.

- **Survey results analysis as shown in table 5.**

Table 5 - Survey Result Analysis: Retargeted Images And User Feedback

Original image	**Retargeted Image**	**Result Analysis**
		

Mean opinion Score =4.22

Mean opinion Score =4.22

Mean opinion Score =4.1

Mean opinion Score =4.1

Mean opinion Score =4.27

Mean opinion Score =4.18

Mean opinion Score =4.21

Mean opinion Score =4.11

Mean opinion Score =4.45

Mean opinion Score =4.52

The effectiveness of our image retargeting project for individuals with retinitis pigmentosa (RP) is supported by both subjective survey results and objective quantitative measures. The Mean Opinion Score (MOS) of approximately 4 out of 5 across 10 sample images indicates that participants perceived the retargeted images as highly similar to the originals with minimal distortion. This positive subjective assessment is corroborated by a range of quantitative quality metrics:

SIFT (Scale-Invariant Feature Transform): Demonstrated preservation of key image features

SSIM (Structural Similarity Index): Showed high structural similarity between original and retargeted images

PSNR (Peak Signal-to-Noise Ratio): Indicated low noise introduction in the retargeting process

Content Loss: Confirmed minimal loss of essential image content

Edge Histogram: Verified the maintenance of important edge information

The alignment between these objective measures and the high MOS strongly supports the efficacy of our retargeting algorithm. It suggests that the technique successfully preserves critical visual information while adapting images to be more accessible for those with RP, particularly considering their reduced peripheral vision and night blindness.

The consistency across multiple images and metrics indicates that our approach is robust and widely applicable, which is crucial for providing a comprehensive solution for RP patients in various visual contexts.

These findings collectively validate the potential of our image retargeting method as a valuable tool for enhancing digital accessibility specifically for the RP community. It promises to significantly improve the visual experience for these individuals, allowing them to perceive images more clearly within their limited field of vision.

References

[1] “Retinitis Pigmentosa (RP): Diagnosis, Symptoms & Treatment.” Accessed: Jun. 27, 2024. [Online]. Available: https://my.clevelandclinic.org/health/diseases/17429-retinitis-pigmentosa

[2] N. Cross, C. van Steen, Y. Zegaoui, A. Satherley, and L. Angelillo, “Retinitis Pigmentosa: Burden of Disease and Current Unmet Needs,” *Clinical Ophthalmology*, vol. 16, pp. 1993–2010, 2022, doi: 10.2147/OPTH.S365486.

[3] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” *ACM Trans Graph*, vol. 26, no. 3, Jul. 2007, doi: 10.1145/1276377.1276390/SUPPL_FILE/PPS010.MP4.

[4] M. Rubinstein, A. Shamir, and S. Avidan, “Improved seam carving for video retargeting,” *ACM Trans Graph*, vol. 27, no. 3, Aug. 2008, doi: 10.1145/1360612.1360615/SUPPL_FILE/A16-RUBINSTEIN.MOV.

[5] W. I. Al-Atabany, T. Tong, and P. A. Degenaar, “Improved content aware scene retargeting for retinitis pigmentosa patients,” 2010. [Online]. Available: http://www.biomedical-engineering-online.com/content/9/1/52

[6] F. Shafieyan, N. Karimi, B. Mirmahboub, S. Samavi, and S. Shirani, “Image seam carving using depth assisted saliency map,” in *2014 IEEE International Conference on Image Processing, ICIP 2014*, Institute of Electrical and Electronics Engineers Inc., Jan. 2014, pp. 1155–1159. doi: 10.1109/ICIP.2014.7025230.

[7] Y. Zhang, Z. Sun, P. Jiang, Y. Huang, and J. Peng, “Hybrid image retargeting using optimized seam carving and scaling,” *Multimed Tools Appl*, vol. 76, no. 6, pp. 8067–8085, Mar. 2017, doi: 10.1007/S11042-016-3318-1/METRICS.

[8] D. Cho, J. Park, T.-H. Oh, Y.-W. Tai, and I. S. Kweon, “Weakly- and Self-Supervised Learning for Content-Aware Deep Image Retargeting,” Aug. 2017, [Online]. Available: http://arxiv.org/abs/1708.02731

[9] T. Zhang, M. Yu, Y. Guo, and Y. Liu, “Content-aware retargeted image quality assessment,” *Information (Switzerland)*, vol. 10, no. 3, 2019, doi: 10.3390/info10030111.

[10] W. Tan, B. Yan, C. Lin, and X. Niu, "Cycle-IR: Deep Cyclic Image Retargeting," May 2019, [Online]. Available: http://arxiv.org/abs/1905.03556

[11] N. Kajiura, S. Kosugi, X. Wang, and T. Yamasaki, "Self-Play Reinforcement Learning for Fast Image Retargeting," in *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2020, pp. 1755–1763. doi: 10.1145/3394171.3413857.

[12] M. R. Naderi, M. H. Givkashi, N. Karimi, S. Shirani, and S. Samavi, "OAIR: Object-Aware Image Retargeting Using PSO and Aesthetic Quality Assessment," Sep. 2022, Accessed: Jun. 26, 2024. [Online]. Available: https://arxiv.org/abs/2209.04804v1

[13] H. Imani, M. B. Islam, and L.-K. Wong, "Saliency-Aware Stereoscopic Video Retargeting." pp. 1230–1239, 2023.

[14] M. H. Givkashi, M. R. Naderi, N. Karimi, S. Shirani, and S. Samavi, "Supervised deep learning for content-aware image retargeting with Fourier Convolutions," *Multimed Tools Appl*, pp. 1–17, Mar. 2024, doi: 10.1007/S11042-024-18876-8/METRICS.

[15] Y. Chen, Y. Pan, M. Song, and M. Wang, "Improved seam carving combining with 3D saliency for image retargeting," *Neurocomputing*, vol. 151, no. P2, pp. 645–653, Mar. 2015, doi: 10.1016/j.neucom.2014.05.089.

[16] M. Abhayadev and T. Santha, "Multi-Operator Content Aware Image Retargeting on Natural Images," 2019.

[17] R. Achanta and S. Süsstrunk, "Saliency detection for content-aware image resizing," *Proceedings - International Conference on Image Processing, ICIP*, pp. 1005–1008, 2009, doi: 10.1109/ICIP.2009.5413815.

[18] W. Tan, B. Yan, C. Lin, and X. Niu, "Cycle-IR: Deep Cyclic Image Retargeting," *IEEE Trans Multimedia*, vol. 22, no. 7, pp. 1730–1743, Jul. 2020, doi: 10.1109/TMM.2019.2959925.

[19] T. Zhang, M. Yu, Y. Guo, and Y. Liu, "Content-aware retargeted image quality assessment," *Information (Switzerland)*, vol. 10, no. 3, 2019, doi: 10.3390/info10030111.

[20] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, "Image Retargeting Quality Assessment," 2011.

[21] B. Yan, B. Bare, K. Li, J. Li, and A. C. Bovik, “Learning quality assessment of retargeted images,” *Signal Process Image Commun*, vol. 56, pp. 12–19, Aug. 2017, doi: 10.1016/j.image.2017.04.005.

[22] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, “Image Retargeting Quality Assessment,” 2011.