

Article

Not peer-reviewed version

SeamCarver: LLM-Enhanced Content-Aware Image Resizing

[Dong Liu](#)*, Yanxuan Yu, [Lianghao Tan](#), Wenjun Wu, Bide Zhao, [Zichao Li](#), Bingjie Lu, Yijie Wen

Posted Date: 23 December 2024

doi: 10.20944/preprints202412.1897.v1

Keywords: efficient machine learning; LLMs in vision



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

SeamCarver: LLM-Enhanced Content-Aware Image Resizing

Dong Liu ^{*}, Wenjun Wu ², Yanxuan Yu ³, Bide Zhao ⁴, Lianghao Tan ⁵, Zichao Li ⁶, Bingjie Lu ⁷ and Yijie Weng ⁸

¹ Yale University

² UIUC; wenjun5@illinois.edu

³ Columbia University; yy3523@columbia.edu

⁴ University of Sydney; bzha0953@uni.sydney.edu.au

⁵ Arizona State University; ltan22@asu.edu

⁶ University of Waterloo; zichao.li@uwaterloo.ca

⁷ Central South University; bingjie_lu@csu.edu.cn

⁸ University of Maryland; jaweng333@gmail.com

* Correspondence: dong.liu.dl2367@yale.edu

Abstract: This paper introduces "SeamCarver," a LLM-enhanced method that redefines image resizing. SeamCarver addresses the limitations of traditional seam carving with static pre-defined parameters, it uses LLM to achieve dynamic and user-controlled AI-resizing of images. We also evaluate the performance of SeamCarver on different datasets. The inclusion of LLMs in this research facilitates dynamic optimization of parameter tuning and adaptive energy function adjustments, enhancing overall robustness and efficiency of image resizing. SeamCarver emerges as a transformative tool, offering versatile, high-quality resized images.

I. Introduction

SeamCarver is an image resizing method designed to maintain important image details during resizing. By integrating **large language models (LLMs)**, SeamCarver goes beyond traditional methods, enabling a smarter, context-aware resizing process. LLMs enhance tasks such as region prioritization, interpolation, edge detection, and energy calculation by analyzing image context or textual inputs. This allows SeamCarver to preserve crucial image features like faces and text while optimizing resizing for different scenarios.

This paper explores how LLMs improve the accuracy and efficiency of image resizing.

II. Background

i. LLM-Enhanced Image Resizing: The SeamCarver Approach

SeamCarver enhances the traditional seam carving method by incorporating LLMs. The integration of LLMs allows for dynamic resizing of images based on real-time context and user input, rather than relying on fixed parameters. The LLM dynamically optimizes the energy functions that determine how seams are selected and removed, making the resizing process adaptive and context-sensitive. This approach allows for superior preservation of image details and better quality when resizing for specific tasks, such as creating thumbnails or preparing images for various screen sizes.

LLMs contribute by adjusting the parameters based on the content of the image and the desired effect, improving the quality of resized images while providing flexibility. The dynamic optimization of the resizing process using LLMs represents a major leap in the flexibility and efficiency of image resizing techniques.

ii. LLM-Related Work

Recent research in *Large Language Models* (LLMs) has demonstrated their potential in a variety of domains, including image processing. LLMs, originally designed for natural language understanding, have been adapted to enhance image processing workflows through advanced model architectures and optimization techniques.

For instance, Li et al. (2024) demonstrated how LLMs can optimize convolutional neural network (CNN) layers for feature extraction in image processing tasks, improving the performance of deep learning models in image classification and segmentation [16]. Zhang et al. (2024) explored the use of LLMs in multi-modal fusion networks, enabling the integration of both visual and textual information, which enhances image analysis tasks [11, 13, 27, 25].

Additionally, efficient algorithm design [1,5,8,12,14,25,28,30] and efficient LLMs [19,20] have shown promise in model compression. Through dynamic optimization, LLMs allow for more context-aware resizing by adjusting energy functions during the process. This flexibility ensures that fine-grained details are preserved, which is especially crucial for tasks like content-aware resizing, as seen in techniques such as SeamCarver. Studies on text-to-image models have demonstrated how LLMs can modify images based on contextual prompts [2,9,15–18,29], providing further advancements in content-aware image resizing.

These advancements highlight the increasing role of LLMs in enhancing traditional image processing tasks and their ability to contribute significantly to content-aware resizing techniques like SeamCarver.

iii. Image and Vision Generation Work

The application of deep learning techniques in *image and vision generation* has also seen significant advancements [3–5,7,10,23,26,27]. Deep convolutional networks, for instance, have been used for texture classification (Bu et al., 2019), which is directly relevant to tasks like energy function optimization in SeamCarver [1]. These methods help improve the detail preservation during image resizing by ensuring that textures and edges are maintained throughout the process.

Furthermore, *multi-modal fusion networks* and techniques for *image-driven predictions* (as demonstrated by Dan et al., 2024) offer important insights into how AI can be used to process and modify images in real-time [4,5]. Besides, model compression currently becoming a favor from both model and system design perspective [21,22]. These innovations align closely with SeamCarver's goal of dynamic, user-controlled image resizing, making them valuable for future developments in image resizing technology.

iv. Image Resizing and Seam Carving Research

SeamCarver builds upon earlier work in *image resizing techniques*. In addition to the foundational work by Avidan and Shamir (2007), other studies have contributed to enhancing seam carving methods. Kiess (2014) introduced improved edge preservation methods within seam carving [?], which is crucial for ensuring that resized images do not suffer from visible distortions along object boundaries. Zhang (2015) compared classic image resizing methods and found that seam carving provided superior results when compared to simpler resizing techniques, particularly in terms of detail preservation [?].

Frankovich (2011) further advanced seam carving by integrating energy gradient functionals to enhance the carving process, providing even more control over the resizing operation [?]. These improvements are incorporated into *SeamCarver*, which leverages LLMs to further optimize the parameter tuning and energy functions during resizing.

v. Impact of SeamCarver and Future Directions

The development of *SeamCarver* represents a significant step forward in content-aware image resizing. By leveraging the power of LLMs, this approach enables adaptive resizing, maintaining high-quality images across a variety of use cases. As machine learning and AI continue to evolve, future versions of *SeamCarver* could integrate even more advanced techniques, such as generative models for even higher-quality resizing and multi-task learning to tailor resizing for specific contexts.

Moreover, *SeamCarver* provides an excellent example of how LLMs can be used to enhance traditional image processing tasks, enabling more intelligent and user-driven modifications to images. This work will likely spur further research into dynamic image resizing and contribute to more versatile, AI-enhanced image editing tools in the future.

III. Functionality

SeamCarver leverages LLM-Augmented methods to ensure adaptive, high-quality image resizing while preserving both structural and semantic integrity. The key functionalities are:

- **LLM-Augmented Region Prioritization:** LLMs analyze semantics or textual inputs to prioritize key regions, ensuring critical areas (e.g., faces, text) are preserved.
- **LLM-Augmented Bicubic Interpolation:** LLMs optimize bicubic interpolation for high-quality enlargements, adjusting parameters based on context or user input.
- **LLM-Augmented LC Algorithm:** LLMs adapt the LC algorithm by adjusting weights, ensuring the preservation of important image features during resizing.
- **LLM-Augmented Canny Edge Detection:** LLMs guide Canny edge detection to refine boundaries, enhancing clarity and accuracy based on contextual analysis.
- **LLM-Augmented Hough Transformation:** LLMs strengthen the Hough transformation, detecting structural lines and ensuring the preservation of geometric features.
- **LLM-Augmented Absolute Energy Function:** LLMs dynamically adjust energy maps to improve seam selection for more precise resizing.
- **LLM-Augmented Dual Energy Model:** LLMs refine energy functions, enhancing flexibility and ensuring effective seam carving across various use cases.
- **LLM-Augmented Performance Evaluation:** CNN-based classification experiments on CIFAR-10 are enhanced with LLM feedback to fine-tune resizing results.

IV. LLM-Guided Region Prioritization

To enhance the seam carving process, we integrate **Large Language Models (LLMs)** to guide region prioritization during image resizing. Traditional seam carving typically removes seams based on an energy map derived from pixel-level intensity or gradient differences. However, this method may struggle to preserve regions with semantic significance, such as faces, text, or objects, which require more context-aware resizing. Our approach introduces **LLMs** to assign semantic importance to different regions of the image, modifying the energy map to prioritize the preservation of these crucial regions.

i. Method Overview

Given an image I , the initial energy map $E(x, y)$ is computed using standard seam carving techniques, typically relying on pixel-based features such as intensity gradients and contrast:

$$E(x, y) = \sqrt{(\nabla I_x(x, y))^2 + (\nabla I_y(x, y))^2}$$

where $\nabla I_x(x, y)$ and $\nabla I_y(x, y)$ represent the gradient values of the image I at pixel (x, y) in the x - and y -directions, respectively.

Next, a **Large Language Model (LLM)** is employed to analyze either the image content directly or a user-provided textual description of the regions to prioritize. For example, a user might specify that "faces should be preserved" or "text should remain readable." This description is processed by the LLM, which assigns an importance score $S(x, y)$ to each pixel based on its semantic relevance. The function that generates these scores is denoted as:

$$S(x, y) = f_{\text{LLM}}(I, D)$$

where f_{LLM} represents the output of the LLM processing both the image I and a description D . The LLM interprets the description D through its internal knowledge of language and context, identifying which parts of the image correspond to higher-priority regions (e.g., faces, text, objects).

The LLM's understanding of the image is derived using advanced techniques like **transformer architectures** [24] and **contextual embedding** [6], which allow the model to capture both local and global relationships within the image, ensuring that important features are accurately recognized and prioritized. For example, the LLM might recognize that a region containing a face is more important than a background area when performing resizing.

ii. Energy Map Adjustment

To modify the energy map, the semantic importance scores $S(x, y)$ are combined with the original energy map $E(x, y)$. This modified energy map $E'(x, y)$ is calculated as follows:

$$E'(x, y) = E(x, y) + \alpha \cdot S(x, y)$$

where α is a scalar weight that determines the influence of the LLM-based importance scores on the energy map. By incorporating $S(x, y)$, the energy map becomes content-aware, ensuring that the regions with higher semantic importance (e.g., faces, text) have lower energy values, making them less likely to be removed during the seam carving process.

iii. Energy Map Adjustment

The modified energy map $E'(x, y)$ is calculated as:

$$E'(x, y) = E(x, y) + \alpha S(x, y)$$

where: - $E(x, y)$ is the original energy map.

- $S(x, y)$ is the semantic importance score derived from the LLM. - α is a weight factor controlling the influence of semantic importance.

iv. Pseudocode

Algorithm 1 LLM-Guided Region Prioritization for Seam Carving

Initialize:

Compute the initial energy map $E(x, y)$ for the image I ;

Obtain semantic importance scores $S(x, y)$ from LLM based on image content or user description;

Normalize the importance scores $S(x, y)$ to a suitable range.

Adjustment:

1. For each pixel (x, y) , compute the adjusted energy map:

$$E'(x, y) = E(x, y) + \alpha \cdot S(x, y)$$

2. Set α to control the influence of semantic importance on the energy map.
3. Repeat for all pixels to generate the adjusted energy map E' .

Output:

The adjusted energy map E' for guiding seam carving.

V. LLM-Augmented Bicubic Interpolation

SeamCarver integrate a LLM-Augmented bicubic interpolation method for image resizing. This method uses a bicubic policy to smooth pixel values, with LLMs improving the visual quality of enlarged images. However, traditional bicubic interpolation does not account for the semantic importance of image regions. To address this limitation, we augment the standard interpolation with semantic guidance from LLMs, ensuring that regions of high importance—such as faces, text, and objects—are preserved more effectively during enlargement.

The traditional bicubic interpolation algorithm operates by using a 4x4 pixel grid surrounding the target pixel to calculate the new pixel value. This method typically focuses on the rate of change between neighboring pixel intensities. In contrast, our approach leverages LLMs to assign semantic importance scores $S(x, y)$ to each pixel, reflecting its contextual significance. These importance scores are derived from the image content or a user-provided description, and they adjust the interpolation weights, effectively guiding the resizing process to preserve critical regions.

The bicubic interpolation formula for a pixel at position (x, y) is based on calculating the weighted sum of the 4x4 neighborhood of surrounding pixels. Traditionally, the interpolation weights $w(x)$ and $w(y)$ are determined based on the relative distance between the target pixel and its neighbors. These weights can be defined as:

$$w(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1, & \text{if } |x| \leq 1 \\ a|x|^3 + 5a|x|^2 + 8a|x| - 4a, & \text{if } 1 < |x| < 2 \\ 0, & \text{otherwise} \end{cases}$$

Then, the new pixel value at position (X, Y) is computed by summing the contributions of the surrounding 16 pixels:

$$B(X, Y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} \cdot w(x - X_i) \cdot w(y - Y_j)$$

For a floating-point pixel coordinate (x, y) , the interpolation involves considering the 4x4 neighborhood (x_i, y_j) , where $i, j = 0, 1, 2, 3$, and calculating the new pixel value as follows:

$$f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 f(x_i, y_j) \cdot w(x - x_i) \cdot w(y - y_j)$$

In the augmented version, the weights $w(x)$ and $w(y)$ are modified based on the importance scores $S(x, y)$ derived from the LLM. For each pixel, we compute the adjusted interpolation weight $w'(x)$ as:

$$w'(x) = w(x) \cdot (1 + \beta \cdot S(x, y))$$

where β is a scalar factor that controls the influence of the semantic importance score. By incorporating these adjusted weights into the interpolation process, regions deemed important by the LLM receive greater priority during the resizing process, resulting in higher-quality enlargements that better preserve semantic content.

The incorporation of LLMs significantly improves the ability of bicubic interpolation to perform content-aware resizing, ensuring that important regions, such as faces, text, or other key objects, are preserved with higher fidelity. The LLM's ability to interpret the image context or a user's textual description enables a more adaptive resizing strategy, where the image can be enlarged in a way that prioritizes and preserves the most semantically relevant regions.

In conclusion, this approach not only enhances the visual quality of enlarged images by preserving important areas but also allows for a more flexible and context-aware image resizing process. The integration of LLMs elevates bicubic interpolation from a purely geometric operation to a more intelligent, context-sensitive method, improving overall resizing performance.

VI. LLM-Augmented LC (Loyalty-Clarity) Policy

SeamCarver also used LLM-Augmented LC (Loyalty-Clarity) Policy to resize images. Traditionally, the LC policy evaluates each pixel's contrast relative to the entire image, focusing on maintaining the most visually significant elements. However, by incorporating LLMs, we enhance this method with semantic understanding, allowing the system to prioritize image regions based not only on visual contrast but also on their semantic importance, as understood from contextual descriptions or image content analysis.

i. Global Contrast Calculation with LLM Influence

The traditional LC policy computes the global contrast of a pixel by summing the distance between the pixel in question and all other pixels in the image. This measure indicates the pixel's relative importance in terms of visual contrast. In our LLM-augmented approach, the global contrast is modified by considering semantic relevance, as dictated by the LLM's analysis of the image or user-provided description.

For instance, if a user inputs that the image contains important "faces" or "text," the LLM assigns higher weights to these regions, increasing their importance in the contrast calculation. The LLM's guidance is mathematically integrated into the contrast calculation as follows:

$$SalS(I_k) = \sum_{\forall I_i \in I} \|I_k - I_i\| + \lambda \cdot \sum_{r \in R} w_r \cdot \|I_k - I_r\| \quad (1)$$

In this formulation: - I_k represents the intensity of the pixel being analyzed, while I_i represents the intensity of all other pixels. - w_r is the weight assigned to a region r by the LLM, which is based on its semantic importance, such as prioritizing faces or text. - λ is a scaling factor that controls the influence of the LLM's weighting on the global contrast calculation.

By adjusting w_r based on the LLM-driven understanding of important regions, the algorithm effectively prioritizes preservation of the semantically significant areas.

ii. Frequency-Based Refinement with LLM Augmentation

To further refine the contrast measure, we incorporate the frequency distribution of intensity values in the image. The traditional frequency-based contrast is enhanced with the LLM's semantic input, which guides how regions of different intensities should be prioritized.

In the standard approach, the global contrast for a pixel I_k is computed as:

$$Sals(I_k) = \sum_{n=0}^{255} f_n \|a_m - a_n\| \quad (2)$$

Where: - f_n is the frequency of the intensity value a_n . - a_m represents the intensity of the pixel I_k , and a_n are the intensity values of all other pixels.

In the LLM-augmented approach, the LLM provides additional weighting for specific regions, emphasizing the importance of certain intensities based on semantic input. The modified calculation is:

$$Sals(I_k) = \sum_{n=0}^{255} f_n \|a_m - a_n\| + \lambda \cdot \sum_{r \in R} w_r \cdot f_r \cdot \|a_m - a_r\| \quad (3)$$

Here, w_r adjusts the weight of the frequency term for pixels in semantically significant regions, as determined by the LLM. This allows for a more refined and context-aware adjustment of the image's contrast, ensuring that the most relevant image areas are preserved during the resizing process.

iii. Application in Image Resizing

By incorporating **LLM-guided** adjustments into the LC Policy, SeamCarver becomes significantly more content-aware. The LLM allows the software to prioritize critical regions—such as human faces, text, or objects—based on user input or semantic analysis of the image. This semantic understanding of the image ensures that, even during resizing, key features remain sharp and well-defined, while less important regions are more freely adjusted.

For example, if a user specifies that "faces" should be preserved, the LLM ensures that these areas have a higher weight during the resizing process, while the surrounding less important areas can be resized with minimal distortion. This **LLM-augmented LC Policy** thus improves the visual integrity of resized images, making the process more adaptable to both user needs and semantic context.



Figure 1. Outlier detected by the LC algorithm.

VII. LLM-Augmented Canny Line Detection

In SeamCarver, the **LLM-augmented Canny Line Detection** algorithm enhances edge and structural feature preservation during image resizing. By incorporating **Large Language Models (LLMs)**, the edge detection process is guided semantically to prioritize regions that are critical to image content, such as faces and text.

i. Algorithm Overview

The Canny Edge Detection algorithm detects edges by analyzing intensity gradients. The standard method detects edges using the first derivative of the image's intensity, but the LLM-augmented approach incorporates semantic information, adjusting the edge detection for important regions identified by the LLM.

ii. Gaussian Filter Application

The image is first smoothed using a Gaussian filter to reduce noise. The filter is represented as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4)$$

This step prepares the image for the gradient calculation while minimizing false edges. In the **LLM-augmented** process, the filter may be adapted based on the semantic regions detected by the **LLM**, ensuring more precise edge detection in critical areas.

iii. Gradient Calculation with LLM Augmentation

After Gaussian filtering, the gradient at each pixel (i, j) is calculated using the Sobel operator:

$$G(i, j) = \sqrt{G_x(i, j)^2 + G_y(i, j)^2} \quad (5)$$

$$\theta(i, j) = \arctan\left(\frac{G_y(i, j)}{G_x(i, j)}\right) \quad (6)$$

Where $G_x(i, j)$ and $G_y(i, j)$ are the derivatives in the horizontal and vertical directions, respectively.

In the **LLM-augmented** method, the gradients are modified by the semantic importance $S(i, j)$ of each region, as identified by the **LLM**. The semantic importance adjusts the gradient magnitude, giving higher weight to edges in critical areas:

$$G_{\text{aug}}(i, j) = G(i, j) \cdot S(i, j) \quad (7)$$

Here, $S(i, j)$ is the semantic score assigned by the **LLM**, where higher values correspond to regions that are semantically more important (e.g., faces, text).

iv. Edge Enhancement

The Canny algorithm applies non-maximum suppression and hysteresis thresholding to refine the detected edges. In the **LLM-augmented** process, the suppression threshold is adapted based on the importance scores:

$$\text{Edge}_{\text{final}} = \begin{cases} G_{\text{aug}}(i, j) & \text{if } G_{\text{aug}}(i, j) \geq \text{threshold}_{\text{high}} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

By incorporating the **LLM**, edges in semantically significant regions (e.g., faces or objects) are preserved with greater accuracy, while less important areas are suppressed more aggressively.

v. Significance in Image Resizing

The **LLM-augmented Canny Line Detection** improves the image resizing process by ensuring that the edges and features critical to the image's content are better preserved. This is especially important when resizing images with significant content like faces or text, where traditional methods might fail to preserve important details.



Figure 2. Original Image.



Figure 3. Edges Detected by the Canny Detector (with LLM Augmentation).

VIII. LLM-Augmented Hough Transformation

The SeamCarver software integrates the **LLM-Augmented Hough Transformation** to enhance line detection during image resizing. This method incorporates semantic guidance from **Large Language Models (LLMs)** to ensure the preservation of key structural features, such as text or faces, during resizing. The LLM augments the traditional Hough Transformation by adjusting the accumulator based on the semantic importance of regions within the image.

i. Algorithm Overview

The Hough Transformation detects lines by mapping points from the image domain to the Hough space. In the traditional approach, collinear points converge to peaks in Hough space, indicating the presence of a line. The **LLM-Augmented Hough Transformation** introduces semantic weighting to this process, ensuring that semantically significant lines (e.g., those in text or faces) are prioritized during line detection.

ii. Mathematical Formulation

In the standard Hough Transformation, the accumulator is updated by incrementing for each detected edge pixel. However, in the **LLM-Augmented Hough Transformation**, the update to the accumulator is influenced by a semantic importance score $S(x, y)$ derived from the LLM. This adjustment ensures that semantically important regions of the image have a greater influence on the detection of lines.

The basic equation for updating the accumulator with the semantic weight $S(x, y)$ is:

$$Accumulator(r, \theta) \leftarrow Accumulator(r, \theta) + S(x, y) \quad (9)$$

Where: - $r = x \cos(\theta) + y \sin(\theta)$ is the radial distance in Hough space. - θ is the angle of the line in Hough space. - $S(x, y)$ is the semantic importance score of the pixel (x, y) , as determined by the LLM.

This formulation adjusts the weight of each edge pixel based on the semantic relevance of its location in the image. Higher $S(x, y)$ values are assigned to pixels in semantically important regions, such as those in faces or text.

iii. LLM-Augmented Hough Transformation Algorithm

The implementation of the **LLM-Augmented Hough Transformation** in SeamCarver is outlined as follows:

Algorithm 2 LLM-Augmented Hough Transformation for Line Detection **Require:** *Image*: input digital image;

Ensure: Lines detected in the image, with semantic guidance from the LLM.

- 1: Apply edge detection (e.g., Canny edge detector) to the image.
- 2: Initialize Hough space and accumulator:
 $Accumulator(r, \theta) = 0$
- 3: **for** each edge pixel (x, y) in the image **do**
- 4: **for** each angle θ **do**
- 5: Compute radial distance $r = x \cos(\theta) + y \sin(\theta)$
- 6: Retrieve the semantic score $S(x, y)$ from the LLM for pixel (x, y)
- 7: Update accumulator: $Accumulator(r, \theta) \leftarrow Accumulator(r, \theta) + S(x, y)$
- 8: **end for**
- 9: **end for**
- 10: Detect peaks in the accumulator.
- 11: Convert infinite lines to finite lines.

iv. Significance of LLM-Augmented Hough Transformation in Image Resizing

The **LLM-Augmented Hough Transformation** allows SeamCarver to preserve important linear structures during resizing. By integrating semantic scores from the LLM, the algorithm ensures that critical features, such as text or faces, are prioritized in the resizing process, making the output more content-aware and maintaining the visual integrity of the resized image.



Figure 4. Original Image.



Figure 5. Lines Detected by LLM-Augmented Hough Transformation with Background.

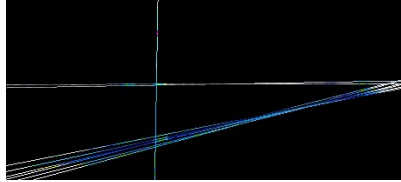


Figure 6. Prominent Lines Detected by LLM-Augmented Hough Transformation.

IX. LLM-Augmented Absolute Energy Equation

In SeamCarver, the **LLM-Augmented Absolute Energy Equation** is employed to enhance the dynamic programming approach for image re-sizing, particularly when handling high-detail images. This method leverages semantic insights provided by Large Language Models (LLMs) to refine the energy gradient calculations, thus improving seam carving performance in complex images.

i. Conceptual Overview

The **LLM-Augmented Absolute Energy Equation** introduces a new dimension to the energy gradient computation by integrating semantic feedback from the LLM. This feedback adjusts the energy gradient of the image in regions that are semantically significant, ensuring that the resizing process preserves critical structural features such as text, faces, and important objects. The LLM's semantic understanding aids in determining areas that should not be resized aggressively, even if they are energetically weak.

ii. Mathematical Formulation

The energy gradient update process is adjusted to incorporate the semantic weighting derived from the LLM. In the original energy equation, the energy of the image $e(I)$ is calculated as the sum of the absolute gradients in both x and y directions

$$e(I) = \left| \frac{\partial I}{\partial x} \right| + \left| \frac{\partial I}{\partial y} \right| \quad (10)$$

The **LLM-Augmented Absolute Energy Equation** modifies this by introducing a semantic weight function $S(x, y)$, which adjusts the energy gradient based on the semantic importance of the pixel:

$$e_{\text{LLM}}(I) = \left(\left| \frac{\partial I}{\partial x} \right| + \left| \frac{\partial I}{\partial y} \right| \right) \times S(x, y) \quad (11)$$

Here, $S(x, y)$ is the semantic importance score of pixel (x, y) , which is provided by the LLM. This score reflects the pixel's relevance in the context of the image, such as text or facial features, and alters the traditional energy calculation to preserve these features more effectively.

iii. Cumulative Energy Update

The cumulative energy matrix is updated with the additional LLM-semantic weighting, where the energy gradient is further refined for semantically significant areas. The cost components C_L , C_R , and C_U are modified as follows to include the semantic weight:

$$\begin{cases} C_L(i, j) = |I(i, j + 1) - I(i, j - 1)| \times S(i, j) \\ C_R(i, j) = |I(i, j + 1) - I(i, j - 1)| \times S(i, j) \\ C_U(i, j) = |I(i, j + 1) - I(i, j - 1)| \times S(i, j) \end{cases} \quad (12)$$

The updated cost function now takes into account the semantic relevance of the pixels as modified by the LLM. This ensures that the resizing process preserves important features, even in low-energy regions.

iv. Comparative Energy Functions

The original energy function is updated using the minimum of the cost components as follows:

$$e(i, j) = e(i, j) + \min \begin{cases} C_L(i, j) + e(i - 1, j - 1) \\ C_R(i, j) + e(i - 1, j + 1) \\ C_U(i, j) + e(i - 1, j) \end{cases} \quad (13)$$

In contrast, the **LLM-Augmented Absolute Energy Equation** incorporates the semantic weighting in the energy update process:

$$e_{\text{LLM}}(i, j) = e(i, j) + |e(i, j + 1) - e(i, j)| + |e(i + 1, j) - e(i, j)| \\ + \min \begin{cases} C_L(i, j) + e(i - 1, j - 1) \\ C_R(i, j) + e(i - 1, j + 1) \\ C_U(i, j) + e(i - 1, j) \end{cases}$$

This update now accounts for the semantic weight of each pixel, making the seam carving process more aware of the image's content, ensuring that important features are preserved even when they might otherwise be considered low-energy regions.

v. Impact on Image Processing

The **LLM-Augmented Absolute Energy Equation** improves the seam carving process by integrating semantic understanding into the energy gradient calculations. The LLM helps identify and preserve critical areas in the image, such as text, faces, and other meaningful structures. This refined approach ensures that important visual content is retained during the resizing process, even in highly detailed images. As a result, the resized image maintains its visual and structural integrity, making the process more content-aware and efficient.

X. LLM-Augmented Dual Gradient Energy Equation

In SeamCarver, the **LLM-Augmented Dual Gradient Energy Equation** advances edge detection by incorporating LLMs to enhance numerical differentiation and gradient computation. The LLM-modified method dynamically adapts the gradient calculation based on local image structure, improving energy mapping and seam carving in images with complex textures.

i. LLM-Enhanced Numerical Differentiation

To improve gradient calculation, the LLM algorithm adapts the traditional Taylor expansion for numerical differentiation. The forward and backward Taylor expansions are modified to reflect the dynamic adjustments provided by LLMs based on the local context:

$$\begin{aligned}
f(x + \Delta x) &= f(x) + \Delta x f'(x) + \frac{\Delta x^2}{2!} f''(x) \\
&+ \frac{\Delta x^3}{3!} f^{(3)}(\xi_1) + \frac{\Delta x^4}{4!} f^{(4)}(x) + \\
&\frac{\Delta x^5}{5!} f^{(5)}(\xi_1),
\end{aligned} \tag{14}$$

Similarly, for the backward expansion:

$$\begin{aligned}
f(x - \Delta x) &= f(x) - \Delta x f'(x) + \frac{\Delta x^2}{2!} f''(x) \\
&- \frac{\Delta x^3}{3!} f^{(3)}(\xi_2) + \frac{\Delta x^4}{4!} f^{(4)}(x) \\
&- \frac{\Delta x^5}{5!} f^{(5)}(\xi_2),
\end{aligned} \tag{15}$$

where $\xi_1, \xi_2 \in (x, x + \Delta x)$. The **LLM** adapts the step size dynamically based on local image features, improving gradient accuracy.

ii. LLM-Modified Gradient Approximation

For gradient calculation, the forward difference approximation is modified with **LLM** to adjust the step size dynamically based on the image structure. This ensures that the method accurately captures fine details and changes in gradient direction:

$$f_x(x, y) \approx \frac{-3f(x, y)}{2} + \frac{4f(x + \delta, y)}{2} - \frac{f(x + 2\delta, y)}{2} + \Delta f_x(x, y) \tag{16}$$

$$f_y(x, y) \approx \frac{-3f(x, y)}{2} + \frac{4f(x, y + \delta)}{2} - \frac{f(x, y + 2\delta)}{2} + \Delta f_y(x, y) \tag{17}$$

Here, $\Delta f_x(x, y)$ and $\Delta f_y(x, y)$ represent the additional corrections provided by the **LLM**, which fine-tune the gradient approximation based on contextual understanding of the image.

iii. LLM-Refined Energy Calculation

The energy for each pixel is computed using the squared sum of the **LLM**-modified gradient components for the RGB channels in both x and y directions:

$$\begin{aligned}
\Delta_x^2(x, y) &= (R_x(x, y) + \Delta R_x(x, y))^2 + \\
&(G_x(x, y) + \Delta G_x(x, y))^2 + \\
&(B_x(x, y) + \Delta B_x(x, y))^2
\end{aligned} \tag{18}$$

$$\begin{aligned}
\Delta_y^2(x, y) &= (R_y(x, y) + \Delta R_y(x, y))^2 + \\
&(G_y(x, y) + \Delta G_y(x, y))^2 + \\
&(B_y(x, y) + \Delta B_y(x, y))^2
\end{aligned} \tag{19}$$

The total energy is the square root of the sum of these squared gradients:

$$energy = \sqrt{\Delta_x^2(x, y) + \Delta_y^2(x, y)} \tag{20}$$

LLMs enhance the gradient calculations by providing dynamic adjustments that improve the detection of subtle edges and structures, ensuring that important image features are preserved during the resizing process.

iv. Application of LLM-Augmented Dual Gradient Energy

The **LLM-Augmented Dual Gradient Energy Equation** improves seam carving by enabling SeamCarver to detect and preserve important image edges more effectively. **LLMs** help dynamically adapt the energy and gradient calculations based on image context, ensuring that complex textures and fine details are preserved during resizing, leading to improved image quality and structural integrity.

XI. Result Evaluation

To evaluate the effectiveness of the image resizing methods in SeamCarver, we conducted an experiment using Convolutional Neural Networks (CNNs) for image classification. The goal was to compare how different resizing techniques impact classification accuracy. Additionally, we explored the role of **LLM-augmented** approaches in enhancing image feature preservation and improving classification outcomes after resizing.

i. Experimental Setup

In the experiment, images were resized using various methods implemented in SeamCarver, including traditional methods and **LLM-augmented** approaches. The resized images were then fed into a CNN model to assess how well each resizing method preserved image features essential for accurate classification. The CIFAR-10 dataset, a well-known benchmark in image classification, was used for this experiment.

ii. Methodology

The workflow of the experiment is as follows:

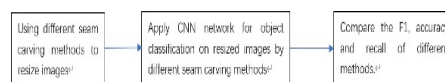


Figure 7. Workflow of the experiment, including **LLM-augmented** methods for image resizing.

The CNN model was first trained on the original CIFAR-10 images, and subsequently, the same model was used to classify images that had been resized using different methods in SeamCarver. This allowed us to evaluate how each resizing method influenced the model's ability to recognize key features. Additionally, **LLM-augmented** methods were used to improve the preservation of important image details during resizing, enhancing classification accuracy.

iii. Results and Discussion

The results, including accuracy metrics and error rates for each sub-experiment, are provided below. The experiment revealed that **LLM-augmented** resizing methods led to superior performance in image classification, particularly in cases where maintaining fine image details was critical.

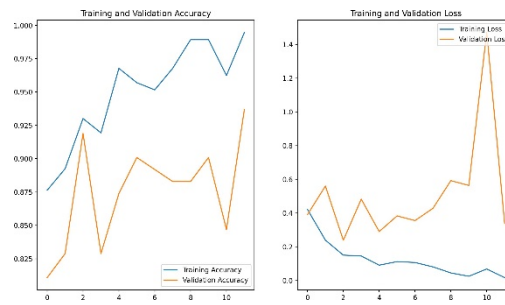


Figure 8. Error and accuracy of a sub-experiment showing the improvements from LLM-augmented methods.

Illustrative examples of images processed by different methods, including LLM-augmented techniques, are shown below, highlighting the visual differences in the resized images and how LLM-augmented methods contribute to better feature preservation.



Figure 9. Image processed by the Bicubic Method.



Figure 10. Image processed by the Absolute Energy Method.



Figure 11. Image processed by the Canny Edge Detection Method.



Figure 12. Image processed by the Dual Gradient Energy Method.



Figure 13. Image processed by the Hough Transformation Method.



Figure 14. Image processed by the LC Method.

iv. Conclusion

This experiment highlights the significant improvements brought by **LLM-augmented** methods in image resizing. By integrating **LLM-augmented** techniques, SeamCarver can preserve finer image details, resulting in improved performance for image classification tasks. These findings emphasize the importance of selecting the right resizing method in applications where image recognition accuracy is crucial.

References

1. Xingyuan Bu, Yuwei Wu, Zhi Gao, and Yunde Jia. Deep convolutional network with locality and sparsity constraints for texture classification. *Pattern Recognition*, 91:34–46, 2019.
2. Han Cao, Zhaoyang Zhang, Xiangtian Li, Chufan Wu, Hansong Zhang, and Wenqing Zhang. Mitigating knowledge conflicts in language model-driven question answering. 2024. URL: <https://arxiv.org/abs/2411.11344>, arXiv:2411.11344.
3. Yu Cheng, Qin Yang, Liyang Wang, Ao Xi-ang, and Jingyu Zhang. Research on credit risk early warning model of commercial banks based on neural network algorithm. 2024. URL: <https://arxiv.org/abs/2405.10762>, arXiv:2405.10762.
4. Han-Cheng Dan, Zhetao Huang, Bingjie Lu, and Mengyu Li. Image-driven prediction system: Automatic extraction of aggregate gradation of pavement core samples integrating deep learning and interactive image processing framework. *Construction and Building Materials*, 453:139056, 2024.
5. Han-Cheng Dan, Bingjie Lu, and Mengyu Li. Evaluation of asphalt pavement texture using multiview stereo reconstruction based on deep learning. *Construction and Building Materials*, 412:134837, 2024.
6. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, 2019.
7. Xiu Fang, Suxin Si, Guohao Sun, Quan Z Sheng, Wenjun Wu, Kang Wang, and Hang Lv. Selecting workers wisely for crowdsourcing when copiers and domain experts co-exist. *Future Internet*, 14(2):37, 2022.
8. Fusen Guo, Huadong Mo, Jianzhang Wu, Lei Pan, Hailing Zhou, Zhibo Zhang, Lin Li, and Fengling Huang. A hybrid stacking model for enhanced short-term load forecasting. *Electronics*, 13(14):2719, 2024.

9. Yue Guo, Shiqi Chen, Ronghui Zhan, Wei Wang, and Jun Zhang. Lmsd-yolo: A lightweight yolo algorithm for multi-scale sar ship detection. *Remote Sensing*, 14(19):4801, 2022.
10. Yuting Hu, Han Cao, Zhongliang Yang, and Yongfeng Huang. Improving text-image matching with adversarial learning and circle loss for multi-modal steganography. In *International Workshop on Digital Watermarking*, pages 41–52. Springer, 2020.
11. Zhuohuan Hu, Fu Lei, Yuxin Fan, Zong Ke, Ge Shi, and Zichao Li. Research on financial multi-asset portfolio risk prediction model based on convolutional neural networks and image processing. *arXiv preprint arXiv:2412.03618*, 2024.
12. Zong Ke, Jingyu Xu, Zizhou Zhang, Yu Cheng, and Wenjun Wu. A consolidated volatility prediction with backpropagation neural network and genetic algorithm. *arXiv preprint arXiv:2412.07223*, 2024.
13. Zong Ke and Yuchen Yin. Tail risk alert based on conditional autoregressive var by regression quantiles and machine learning algorithms. *arXiv.org*, 2024. URL: <https://arxiv.org/abs/2412.06193>.
14. Zong Ke and Yuchen Yin. Tail risk alert based on conditional autoregressive var by regression quantiles and machine learning algorithms. *arXiv preprint arXiv:2412.06193*, 2024.
15. Zhixin Lai, Jing Wu, Suiyao Chen, Yucheng Zhou, and Naira Hovakimyan. Residual-based language models are free boosters for biomedical imaging. 2024. URL: <https://arxiv.org/abs/2403.17343>, arXiv:2403.17343.
17. Keqin Li, Lipeng Liu, Jiajing Chen, Dezhi Yu, Xiaofan Zhou, Ming Li, Congyu Wang, and Zhao Li. Research on reinforcement learning based warehouse robot navigation algorithm in complex warehouse layout. *arXiv preprint arXiv:2411.06128*, 2024.
18. Keqin Li, Jin Wang, Xubo Wu, Xirui Peng, Runmian Chang, Xiaoyu Deng, Yiwen Kang, Yue Yang, Fanghao Ni, and Bo Hong. Optimizing automated picking systems in warehouse robots using machine learning. *arXiv preprint arXiv:2408.16633*, 2024.
19. Sicheng Li, Keqiang Sun, Zhixin Lai, Xiaoshi Wu, Feng Qiu, Haoran Xie, Kazunori Miyata, and Hongsheng Li. Ecnet: Effective controllable text-to-image diffusion models. 2024. URL: <https://arxiv.org/abs/2403.18417>, arXiv:2403.18417.
20. Dong Liu, Zhixin Lai, Yite Wang, Jing Wu, Yanxuan Yu, Zhongwei Wan, Benjamin Lengerich, and Ying Nian Wu. Efficient large foundation model inference: A perspective from model and system co-design. 2024. URL: <https://arxiv.org/abs/2409.01990>, arXiv:2409.01990.
21. Dong Liu and Kaiser Pister. Llmeasyquant – an easy to use toolkit for llm quantization. 2024. URL: <https://arxiv.org/abs/2406.19657>, arXiv:2406.19657.
22. Dong Liu, Roger Waleffe, Meng Jiang, and Shivaram Venkataraman. Graph-
23. snapshot: Graph machine learning acceleration with fast storage and retrieval. 2024. URL: <https://arxiv.org/abs/2406.17918>, arXiv:2406.17918.
24. 2406.17918, arXiv:2406.17918.
25. Dong Liu and Yanxuan Yu. Mt2st: Adaptive multi-task to single-task learning. 2024. URL: <https://arxiv.org/abs/2406.18038>, arXiv:2406.18038.
26. 2406.18038, arXiv:2406.18038.
27. Junran Peng, Xingyuan Bu, Ming Sun, Zhaoxiang Zhang, Tieniu Tan, and Junjie Yan. Large-scale object detection in the wild from imbalanced multi-labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9709–9718, 2020.
28. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Ilia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.
29. Chunya Wu, Zhuoyu Yu, and Dexuan Song. Window views psychological effects on indoor thermal perception: A comparison experiment based on virtual reality environments. *E3S Web of Conferences*, 546:02003, 2024. URL: <http://dx.doi.org/10.1051/e3sconf/202454602003>, doi:10.1051/e3sconf/202454602003.

30. Wenjun Wu. Alphanetv4: Alpha mining model. *arXiv preprint arXiv:2411.04409*, 2024.
31. Ao Xiang, Zongqing Qi, Han Wang, Qin Yang, and Danqing Ma. A multimodal fusion network for student emotion recognition based on transformer and tensor product. 2024. URL: <https://arxiv.org/abs/2403.08511>, arXiv:2403.08511.
32. Jun Xiang, Jun Chen, and Yanchao Liu. Hybrid multiscale search for dynamic planning of multi-agent drone traffic. *Journal of Guidance, Control, and Dynamics*, 46(10):1963–1974, 2023.
33. Wangjiaxuan Xin, Kanlun Wang, Zhe Fu, and Lina Zhou. Let community rules be reflected in online content moderation. 2024. URL: <https://arxiv.org/abs/2408.12035>, arXiv:2408.12035.
34. Zhibo Zhang, Pengfei Li, Ahmed Y Al Hammadi, Fusen Guo, Ernesto Damiani, and Chan Yeob Yeun. Reputation-based federated learning defense to mitigate threats in eeg signal classification. In *2024 16th International Conference on Computer and Automation Engineering (ICCAE)*, pages 173–180. IEEE, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.