# Photo frame recommendation based on image similarity and segmentation


Kim Jake
Sungkyunkwan University
jake630@g.skku.edu

Park Jong Ho
Sungkyunkwan University
jayson0611@gmail.com

Baek Yumi
Sungkyunkwan University
yumibaek@naver.com



## ABSTRACT

This study proposes a new methodology for recommending personalized photo frames using image similarity and segmentation. This method attempts to segment images using a hierarchical clustering algorithm and K-Means clustering. In particular, in K-Means clustering, the optimal K value for whole image is determined by Elbow method and image segmentation is performed based on this value. Image frames generated through this method can be recommended by measuring the similarity with photos provided by the user using a histogram-based method. The key of this study is to measure the similarity between two images based on the correlation coefficient calculated by comparing histograms In addition, if the similarity of the user's photo did not exceed a specific threshold, our model does not recommend any frames. This study provides a new approach to image processing and user custom content recommendation, and shows the potential to improve the efficiency and accuracy of image-based recommendation systems.



## Keywords

Frame Recommendation, K-means Clustering, Segmentation, Histogram, Similarity Threshold


## 1. INTRODUCTION

In the modern digital age, images play a central role in information delivery, and accordingly, the demand for personalized image content is constantly increasing. In particular, as the importance of customized image content has emerged in the social media and digital marketing fields, interest in image recommendation systems that fit users' tastes and needs is increasing. Against this background, there is a need for research to recommend customized photo frames to users using image similarity and segmentation technology. Image similarity is used to identify images with similar visual characteristics, and segmentation helps provide a simpler frame by separating various parts that make up the image. Any image can be made into a simplified frame using segmentation. The combination of these technologies can improve the user experience and improve the accuracy of personalized image recommendations.

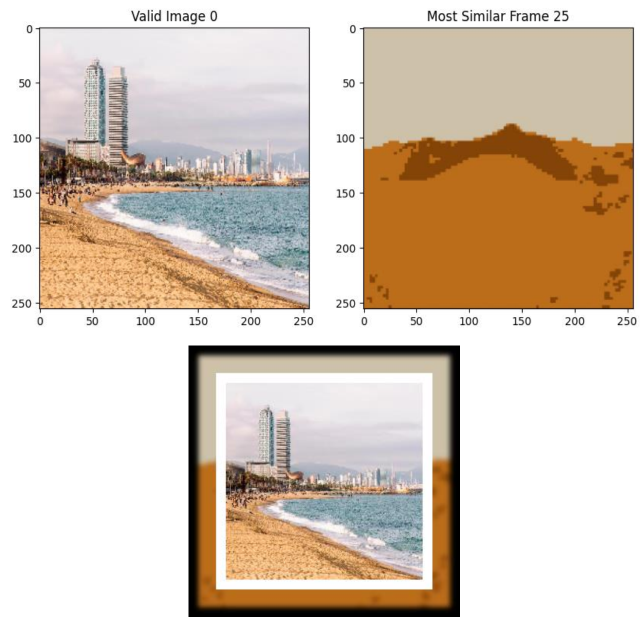



**Figure1. Recommended frame image by similarity and result of our model**

The purpose of this study is to develop a system that recommends optimal photo frames to users by integrating image similarity and segmentation technology. To this end, we develop a methodology that segments images using K-Means clustering and compares the generated frames with the user's photo to recommend frames with high similarity. The key question of the study is as follows: "How can we effectively recommend photo frames that best match the user's photo based on K-Means clustering and image similarity measurement?" To answer this question, this study combines image processing and machine learning technology to develop a personalized recommendation system and evaluate the efficiency and accuracy of this system.

## 2. RELATED WORK & BACKGROUND

### 2.1. The Role and Importance of Clustering

In the realm of modern data processing and analysis, the role of clustering algorithms cannot be overstated. These unsupervised algorithms serve as fundamental tools in deciphering the complex structures inherent in vast datasets. By grouping data points based on shared characteristics, clustering facilitates a deeper understanding and interpretation of data, which might otherwise remain obscured in its raw form. Techniques such as K-Means clustering and hierarchical clustering are pivotal in this regard, finding widespread application across diverse domains including data analysis, pattern recognition, and data mining. Clustering algorithms, particularly K-Means, excel in their ability to simplify complex data sets. By partitioning data into a specified number of clusters, these algorithms minimize variance within each cluster

while maximizing the distinction between different clusters. This process of segregation and categorization is instrumental in revealing underlying patterns and structures in data. For instance, in the field of market research, clustering helps in segmenting customers into distinct groups based on purchasing behavior or preferences, enabling targeted marketing strategies. Moreover, hierarchical clustering offers a unique perspective by creating a tree-like structure of the data. This method is particularly useful for understanding the hierarchical relationships within data and can be crucial for more nuanced analyses, such as in biological research for gene expression data or in linguistics for studying language families. Furthermore, the adaptability of clustering algorithms to handle different types of data - be it numerical, categorical, or even text – adds to their versatility. This flexibility is crucial in a world where data is generated in myriad forms, each requiring a tailored approach for analysis.

However, the application of clustering algorithms to the domain of image segmentation has not been extensively explored. Despite their proven efficacy in various data processing scenarios, these algorithms have seen limited adaptation in the field of image analysis, particularly in the context of segmenting complex visual data. This gap presents a notable opportunity for innovation, as the principles of clustering could offer valuable insights when applied to image segmentation tasks. The potential for clustering algorithms, especially methods like K-Means, in image segmentation lies in their inherent ability to discern and group pixels or features based on similarity. This approach, though straightforward in its conceptualization, could provide a robust framework for segmenting images into meaningful parts.

## 2.2. Image Segmentation and Our Approach

Image segmentation, a core process in computer vision and image processing, involves dividing an image into multiple segments to make it more meaningful for analysis. Historically, segmentation techniques have evolved from simple threshold-based methods to more sophisticated algorithms. Early methods focused on basic partitioning based on intensity values, color, or texture. As computational power increased, more complex methods like edge detection and region-growing algorithms (Ross et al, 2013) [1] became popular. In recent years, the focus has shifted towards machine learning and deep learning-based approaches. Convolutional Neural Networks (CNNs) (LeCun et al, 1998) [2], for instance, have been widely used for their effectiveness in handling complex image structures. Techniques like semantic segmentation, where each pixel is classified into a category, and instance segmentation, identifying each instance of objects separately, have become benchmarks in the field. The landscape of image segmentation has been profoundly reshaped by the advent of advanced deep learning models. Neural networks, particularly U-Net (O. Ranneberger et al, 2015) [3] and Mask R-CNN (K. He et al, 2017) [4], have established new benchmarks in accuracy and efficiency. These models have dramatically improved the ability to discern intricate details within images, making them indispensable tools in modern segmentation techniques. Also, the shift towards automated feature extraction powered by deep learning has marked a significant departure from traditional methods that heavily relied on manual feature engineering. This evolution has streamlined the segmentation process, allowing for more accurate and efficient analysis of images. The automation of feature extraction not only speeds up the process but also reduces the potential for human error, leading to more reliable and consistent results in segmentation tasks.

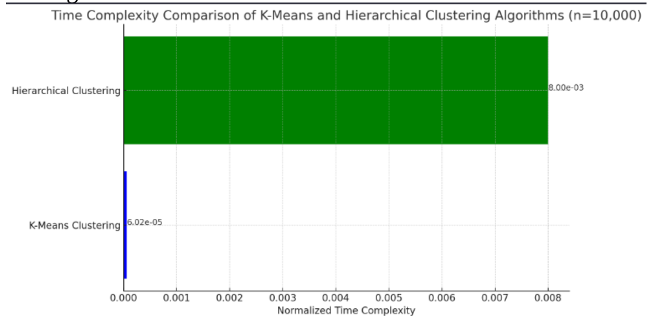



**Figure2. Comparison of the computational complexity of K-Means clustering and hierarchical clustering we experimented with. It shows the time of clustering for 10000 data. Overwhelmingly, hierarchical clustering takes a lot of time. This is the execution time for 10000 data, but in our project, 10000 data mean 100x100 pixels of images. If the image size is set to 256x256, the time required for hierarchical clustering becomes incomparable to that of K-Means. This shows that using K-Means for large datasets is a better choice in terms of time and image quality.**

These advancements collectively represent a paradigm shift in the field of image segmentation, where the integration of deep learning and sophisticated data processing techniques continues to drive innovation and improve outcomes. Following these significant advancements in image segmentation, it is essential to highlight the distinctive approach and benefits of our clustering-based model within this evolving landscape. Unlike the deep learning models that necessitate substantial computational power and extensive training on large datasets, our model operates on a fundamentally different paradigm. It is designed to require minimal computational resources, making it not only efficient but also more accessible for a wide range of applications. A key feature of our approach is its inherent ability to handle novel images for frame creation without the need for prior training. This characteristic stands in stark contrast to many deep learning techniques, which typically require extensive training phases to adapt to new types of data. Our model's flexibility in dealing with unfamiliar images is a testament to its robustness and adaptability. Furthermore, this robustness extends to the model's capability to match any user-input image with an appropriate frame. This is achieved through the intelligent application of clustering algorithms, which analyze the intrinsic properties of the input image and identify the most suitable frame based on similarity. This process ensures that the recommended frames are not only aesthetically pleasing but also contextually relevant to the input image, enhancing the user experience significantly. In essence, our clustering-based model represents a pragmatic and effective solution in the realm of image segmentation and frame recommendation. It aligns well with the current trajectory of technological advancement, yet distinguishes itself by offering a more streamlined approach. This model, therefore, contributes meaningfully to the ongoing discourse in image segmentation, emphasizing the importance of versatility and efficiency in algorithmic design.

# 3. DATA & PREPROCESSING

## 3.1 Data

The data used for the project consists of a variety of photos with diverse landscapes. It is an image dataset, and we collected a large amount of image data through two sources. The first source is the landscape dataset on Kaggle Landscape Pictures, and the second involves crawling for image data on Google. The Kaggle Landscape dataset includes 900 unlabeled landscape photos, 900 mountain photos, 100 desert photos, 500 sea photos, and 500 beach photos. We worked on this project in six categories: city, desert, mountain, snow, grasslands, and sea. In order to make the Kaggle data set suitable for our project, the categories not included in the data set were collected through web crawling, and the data were collected taking into account the imbalance in the number of images in each category. For web crawling, inputting keywords by category did not guarantee that all image data would be suitable. Consequently, we conducted a manual review, discarding data that was deemed inappropriate based on visual inspection. For setting a similarity threshold, about 850 images were collected, with an equal proportion of photos for each category. For the validation set, we utilized a total of about 3000 images. This was done to validate the effectiveness of applying suitable frames when a user inputs a photo by measuring similarity.

## 3.2 Preprocessing

In our project, we undertook specific preprocessing steps for the data, which was essential for effective application of K-Means and hierarchical clustering algorithms. These steps were meticulously designed to prepare the dataset for clustering and similarity calculations.

**1. Histogram Generation for Similarity Calculation**

To facilitate the computation of similarity between the user-input images and potential frames, each image was first transformed into a histogram. This was achieved through the following sequence of operations: Reading the user image using OpenCV and converting it to a PIL image format. Resizing the image to a uniform dimension of 256x256 pixels using the LANCZOS filter in PIL, a choice made to balance image quality and computational efficiency. Converting the PIL image back to a NumPy array and then to the HSV color space using OpenCV, as it provides a more accurate representation for color-based similarity calculations. Computing the histogram of the HSV image, which was then normalized to a range between 0 and 1 using OpenCV's cv2.normalize function. This histogram representation was critical for subsequent similarity assessments.

**2. Image Resizing**

Separate from the histogram generation, images were also resized to 256x256 pixels and converted to the RGB color space for consistent visual representation. This step ensured that the images, when displayed alongside their matched frames, maintained a standard size and color consistency. The decision to resize images to 256x256 pixels was particularly influenced by the high time complexity associated with hierarchical clustering. As hierarchical clustering's computational load increases significantly with the number of pixels, resizing the images was a necessary compromise to ensure computational feasibility without substantially compromising the image quality. This resizing step aligns with our comparison of the time complexities of K-Means and hierarchical clustering (Fig.2), where we emphasized the efficiency of K-Means with larger datasets. In contrast, hierarchical clustering, while offering detailed and nuanced clustering, necessitates a reduction in image size to manage its higher computational demands effectively. Therefore, in this project, K-Means Algorithm was mainly used as an algorithm for frame production, and the results of hierarchical clustering will also be presented.

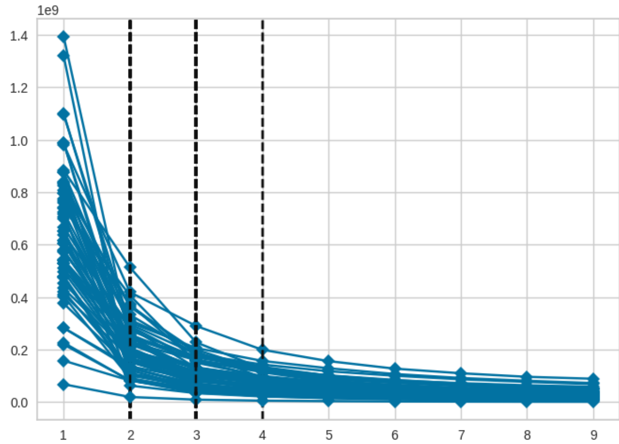
**Figure 3. Failed images of CycleGAN**

# 4. TECHNICAL APPROACH & METHODS

## 4.1 Frame Making Algorithms

Our first approach for creating photo frames was to use a generative model like CycleGAN (JY ZHU et al, 2017) [5]. Initially, we attempted to train the CycleGAN model using photo-frame pairs, aiming to generate new images. However, as in Fig2, challenges such as limited GPU and resources, as well as a lack of understanding of the algorithm, arose (Fig 3). Moreover, even when using a generative model to produce images, the output was not clear, leading us to conclude that it would not serve the purpose of a photo frame.

In our research, we changed our methods to use the capabilities of clustering algorithms for the purpose of making an image frame by an image segmentation, a process integral to our project's core objective. Image segmentation, a critical process in computer vision, involves the decomposition of an input photograph into individual pixels or groups of pixels, facilitating more granular analysis and manipulation. In this context, we specifically focused on employing K-Means clustering and hierarchical clustering, both of which are revered in the realm of unsupervised learning for their efficacy and versatility. K-Means clustering, a well-established algorithm in the field of data science, operates by partitioning pixels into a predetermined number of clusters. This method functions by assigning pixels to the nearest cluster center and iteratively refining these centers. The algorithm's ability to group pixels based on their attributes such as color and intensity makes it an ideal choice for our objective of transforming the input photo into a distinct and visually appealing photo frame. Complementing the K-Means algorithm, hierarchical clustering offers a nuanced approach to segmentation. Unlike K-Means, which partitions the image into a predefined number of clusters from the onset, hierarchical clustering adopts a more progressive strategy. It starts by treating each pixel as a separate cluster and then iteratively merges clusters based on their similarity. This method allows for the creation of a dendrogram, presenting a hierarchical structure of clusters that can be insightful for understanding the natural groupings within the image. This hierarchical perspective is particularly advantageous in our project, as it enables a more refined control over the segmentation granularity, a crucial aspect in the creation of customized photo frames. The combined application of K-Means and hierarchical clustering in our project is not merely a technical choice but a strategic one. The utilization of these algorithms transcends the



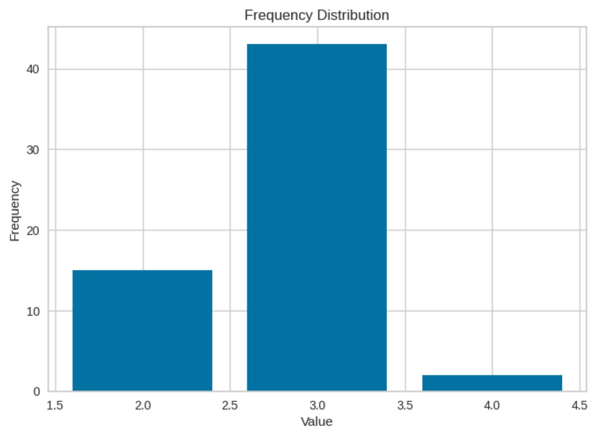
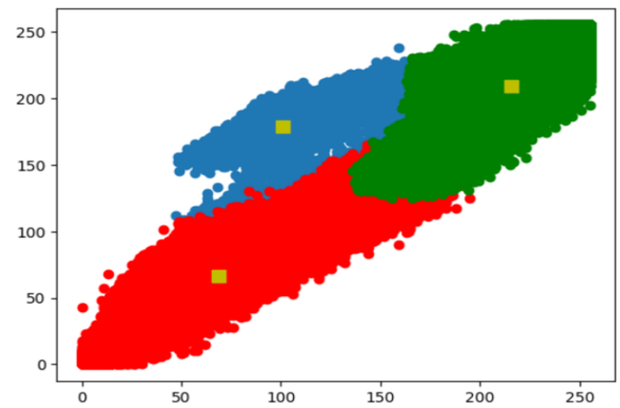




**Figure 4(above). Result of Elbow methods for 60 images, 5(below). Frequency Distribution of K**



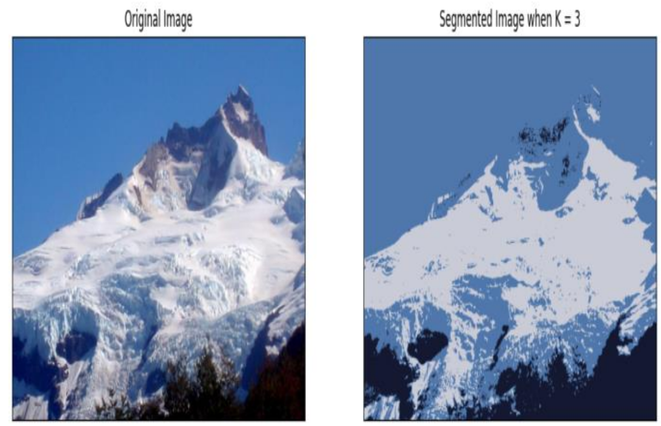




**Figure 6(above), 7(below). Clustered pixel vectors by using the K-Means algorithm and output image, respectively.**

traditional boundaries of image processing, as it is not solely focused on segmenting the frame but also aimed at enhancing input photos visual impact. By applying these algorithms, we simplify the segmentation process, ensuring that the resultant photo frames are not only representative of the input image but also elevated in terms of their visual appeal. This approach is in line with the contemporary trends in digital image processing, where the emphasis is increasingly on the aesthetic enhancement of images, alongside the technical accuracy of segmentation.

## 4.2 K-Means Clustering Algorithm

The K-Means algorithm, an exemplar of unsupervised learning in machine learning, efficiently partitions a dataset into K distinct, non-overlapping subgroups or clusters. It operates on a simple, yet effective iterative refinement approach, with the goal of minimizing the within-cluster variances, also known as the inertia. The process initiates by selecting K points as initial centroids, either randomly or based on a heuristic. The algorithm then iteratively performs two steps: assignment and update. In the assignment step, each data point is allocated to the nearest centroid, with proximity typically measured using Euclidean distance. This forms K clusters with the dataset's points partitioned around the centroids. Subsequently, in the update step, each centroid's position is recalculated as the mean of all points in its cluster, thus adjusting the centroid to the center of its cluster. These two steps are repeated until a convergence criterion is met, usually when there is negligible change in the centroids' positions or the assignments of points to clusters.

The most important setting of hyperparameters in the K-Means algorithm is to determine the number of clusters. We wanted to set the number of statistically meaningful clusters in advance and create a frame based on the number of clusters we set, rather than the deciding the number of clusters for every image. This can build a frame dataset much faster than how K is determined for all images to be frames. We randomly selected 10 pictures from each category and selected the appropriate number of clusters for them. In this process, the elbow method was used (Fig 4, 5). A proper value of K was extracted for a total of 60 pictures, and in conclusion, it was concluded that setting K to 3 is best.

After we decided for K, we applied the K-Means clustering algorithm in the following steps. First, we loaded the images and converted them into vector form. Specifying K as 3, we standardized the number of clusters to 3. The pixel-vectors were then grouped into three clusters. (Fig 6) This automation allowed for the rapid processing of large volumes of images, significantly reducing the time and manual effort required in frame generation. The efficiency of this automated approach was evident in its ability to swiftly produce high-quality, customized photo frames that were both aesthetically pleasing and representative of the input images' characteristics. Using the centroids obtained from K-Means clustering, we performed image segmentation. The results were compared and visualized by displaying the original image on the left and the segmented image on the right (Fig 7).

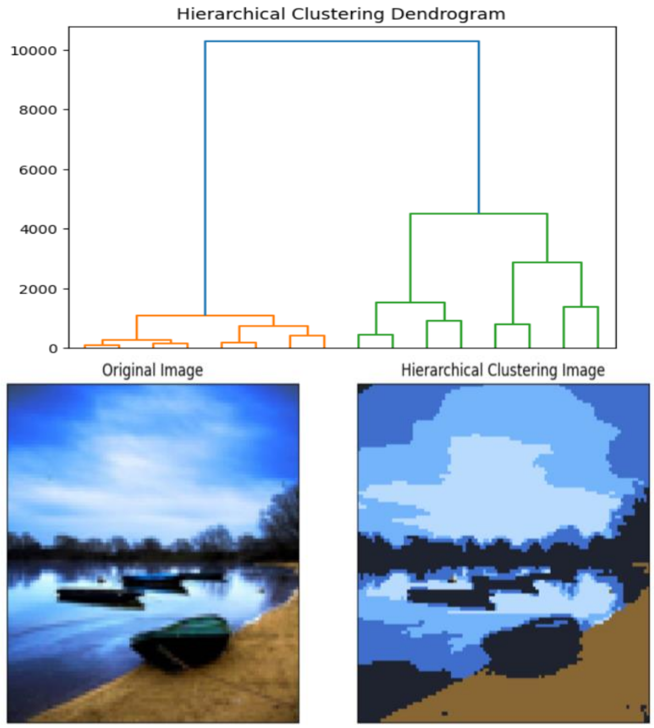



**Figure 8(above), 9(below). Dendrogram of Hierarchical clustering algorithm and output image, respectively.**

## 4.3 Hierarchical Clustering Algorithm

In the context of our study, we confronted a notable computational challenge when applying hierarchical clustering to image segmentation due to its inherent time complexity. Hierarchical clustering, known for its $O(n^3)$ time complexity, becomes increasingly computationally intensive with larger datasets, particularly those comprising high-resolution images. This aspect necessitated a reduction in image size to facilitate a more feasible application of the algorithm. Consequently, while this resizing inevitably led to a decrease in image quality, it was an essential trade-off to harness the detailed and nuanced segmentation capabilities of hierarchical clustering.

Despite the compromise on image resolution, it was imperative to explore the application of hierarchical clustering in our project. This exploration was driven by a need to comprehensively assess and compare its performance against the K-Means algorithm, which is significantly more efficient in handling larger datasets due to its lower time complexity. By applying both K-Means and hierarchical clustering to the same set of resized images, we aimed to conduct a thorough comparative analysis. This analysis was not only focused on evaluating the computational efficiency of each algorithm but also on understanding the nuances in the quality of segmentation they each afforded. Through this comparative study, we sought to elucidate the strengths and limitations of each clustering method in the context of image segmentation. While hierarchical clustering offers a more detailed segmentation, it does so at the expense of computational speed and image resolution. On the other hand, K-Means provides a more computationally efficient alternative, albeit with potentially less granularity in segmentation. This juxtaposition of the two methods allowed us to draw critical insights into their applicability in different scenarios within the realm of image processing.

Hierarchical clustering was conducted in the following steps. The outcomes of hierarchical clustering were visualized using a dendrogram (Fig 8). Utilizing the results above, we further segmented the images (Fig 9), calculated the average color of each cluster, and generated segmented images. Thus, we utilized the clustering information obtained from hierarchical clustering to segment the images. Hierarchical clustering requires a lot of time resources, but showed more sophisticated segmentation compared to the K-Means algorithm. However, due to the enormous shortcoming of time complexity, the K-means algorithm was mainly used in the frame making process.

## 4.4 Calculating image similarity

Through segmentation processes, we complete the frame images. To recommend frame images similar to the input image when it is provided, we measure the similarity between the input image and frame images. A pivotal aspect of our project is the calculation of image similarity, which plays a crucial role in the recommendation of photo frames. This process involves comparing the histogram-based similarity between a user-submitted image and our pre-processed frames. We utilized OpenCV's cv2.calcHist and cv2.compareHist functions, leveraging the correlation coefficient as the key metric for assessing similarity. Our experiments also explored the use of Bhattacharyya distance; however, it was observed to yield uniformly high similarity across all categories, rendering it unsuitable as a threshold for image recommendation. The operational workflow of our image recommendation system based on image similarity is as follows: Upon receipt of a user-submitted photo, the system processes the image to extract its histogram representation. The histogram of the input image is then compared with those of our frames, created via K-Means or hierarchical clustering, to identify the frame with the highest similarity. We established a similarity threshold to refine our recommendation process. If the similarity between the user's photo and all available frames falls below this threshold, the system refrains from making a recommendation. Determining the appropriate threshold was a nuanced aspect of our methodology. We conducted a comprehensive analysis by calculating the accuracy of our model's category prediction across various thresholds. This involved generating 36 histograms, representing similarity calculations between each category of frames and six categories of images. The optimal threshold was identified by evaluating the balance between true positive (TP) and false positive (FP) rates across these categories. We found that a correlation coefficient in the range of 0.4 to 0.5 achieved the best balance, maximizing recommendation accuracy while maintaining a reasonable number of recommendations.

One of the challenges we faced in setting the threshold was the inverse relationship between the number of recommendations and their accuracy. Higher thresholds led to fewer recommendations but increased accuracy. After careful consideration, we concluded that the 0.4 to 0.5 range offered an optimal trade-off, effectively balancing accuracy with the diversity of recommendations

# 5. EXPERIMENT & RESULT

## 5.1 Enhancing the Validation Dataset

In our project, the primary focus was not on training a model, but rather on evaluating the efficacy of pre-constructed frames, generated through clustering algorithms, in accurately matching with user-submitted images. To this end, the size of our validation dataset played a pivotal role. Unlike traditional model training scenarios where validation sets are typically smaller subsets of the overall dataset, we deliberately increased the size of our validation dataset beyond the threshold-determining dataset size, about 3000 images. This strategic enhancement was driven by our objective to obtain a more objective and comprehensive evaluation of our experiment. Through this method, we sought to establish a solid foundation for the evaluation of our clustering-based frame matching system. The larger validation dataset enabled us to conduct a thorough analysis, leading to more reliable and generalizable insights about the performance of our system. However, due to the lack of our frame categories, this has caused more frame production. This approach underscored our commitment to ensuring the robustness and applicability of our frame recommendation methodology in real-world scenarios.

## 5.2 Visualization Experiment

In our study, we conducted an experiment to evaluate the efficacy of our frame recommendation system in matching similar images to predefined frames. This experiment was integral in demonstrating the practical application of our similarity-based approach. The experiment commenced with the function *find_most_similar*, which was designed to identify the frame most similar to each image in the dataset. This function iterated through each image, calculating the similarity between the image and each frame using a correlation-based similarity measure. The frame with the highest similarity score exceeding a predefined threshold was selected as the most similar frame for the respective image. In cases where the maximum similarity for an image did not surpass the threshold, the image was deemed to have no sufficiently similar frame. Following the similarity matching, we visualized the results using the *plot_similar_images* function. This function created a series of subplots for each valid image-frame pair, where the valid image and its most similar frame were displayed side by side. This visual representation was pivotal in providing a clear and intuitive understanding of the similarity between the images and the recommended frames. The outcome of this experiment was quantitatively summarized by counting the number of images for which a sufficiently similar frame was identified. This count, along with the visual subplots, offered a comprehensive overview of the system's performance in matching images to frames. It also highlighted instances where the system could not find a frame with a similarity above the threshold, thereby underscoring the limitations and areas for potential improvement in our recommendation algorithm.

## 5.3 Validation of Similarity Threshold

In our study, a threshold validation process was employed to ascertain the accuracy and effectiveness of our frame recommendation system. We increased the threshold from 0 to 1 by 0.1 and proceeded validation according to the increase in the threshold. This process was crucial in evaluating the performance of our image similarity-based recommendation algorithm. To achieve this, we utilized a structured approach to calculate key statistical metrics, including accuracy and the F1 score, which are vital indicators of the system's predictive performance. Initially, we constructed a confusion matrix to assess the true positive (TP) and false positive (FP) rates. This was accomplished by comparing the most similar frames, as determined by our algorithm, against the actual categories of the user-submitted images. For each image-frame pair, we classified the match as a TP if the frame correctly corresponded to the image's category, and as an FP if it did not. This step was essential to understand the algorithm's capability to correctly identify and match frames to the respective image categories. To augment our

| Threshold | Accuracy | F1 Score |
|---|---|---|
| 0.1 | 0.433 | 0.604 |
| 0.2 | 0.433 | 0.604 |
| 0.3 | 0.427 | 0.598 |
| 0.4 | 0.429 | 0.601 |
| **0.5 (Ours)** | **0.407** | **0.579** |
| 0.6 | 0.435 | 0.606 |
| 0.7 | 0.5 | 0.667 |
| 0.8 | 0.55 | 0.709 |
| 0.9 | 0.83 | 0.909 |
| 1.0 | 0 | 0 |

**Table 1. Accuracy for each threshold and F1 score (150 image)**

analysis for F1 score, we extended the computation to include false negative (FN) and true negative (TN) rates, although these metrics were not directly derivable from our dataset due to the nature of the frame recommendation task. In a viewpoint of an image as a whole, we set both TN and FN to 0. The confusion matrix provided a comprehensive view of the system's performance, allowing us to visualize the distribution of TP, FP, FN, and TN outcomes. Subsequently, we calculated the overall accuracy of the system, defined as the proportion of correct predictions (both TP and TN) out of all predictions made. This metric provided a straightforward assessment of the system's effectiveness. Additionally, the F1 score, a harmonic mean of precision and recall, was computed. This score is particularly informative in scenarios where an imbalance between the positive and negative classes exists, as it conveys the balance between the precision (the proportion of TP out of all positive predictions) and recall (the proportion of TP out of actual positive cases).

The combination of these statistical measures—accuracy and the F1 score—offered a robust evaluation of our system. By analyzing these metrics, we were able to critically assess the system's capacity to accurately recommend photo frames, ensuring that the recommendations were not only relevant but also reliable.

## 5.4 Result

Our first result is the number of frame matching for all images. The similarity with the frames we have was calculated for a total of 2,802 photos, and a total of 168,120 image similarity calculations were performed. Among these similarities, the frame with the highest similarity among frames exceeding the threshold is recommended as an appropriate frame, and if no frame exceeds the similarity threshold, the frame is not recommended. As a result, frames were recommended for a total of 1868 images. Only about 67% of the images were recommended for frames, meaning that about 33% of the images had no similar frames. This ratio could be larger if the number of categories was increased.

Our second result is to calculate the TP and FP values for each category and calculate the accuracy from the overall category perspective. In this process, in addition to the threshold of 0.5, we calculated the accuracy for various thresholds. In addition, for the calculation of the F1 score, FN and TN were set to 0. In this process, the accuracy calculated from our 2,802 images was approximately 3%. This seems to be due to the imbalance between the image and the number of frames. Therefore, we reduced the number of images and experimented again on 150 images. This is because if the number of images and frames is balanced, the accuracy will increase. Conversely, this implies that we have to build a large amount of image frame database, and at the same time, it is a limitation of this project. Table 1 shows the frame matching results of 150 images for each threshold. Naturally, as the threshold goes up, the accuracy and F1 score go up. However, this includes a kind of trade-off because as the accuracy increases, the number of images recommended for frames decreases. Therefore, we confirmed that our threshold balances the two well.

## 6. CONCLUSION

Throughout this project, we successfully curated a diverse collection of photos with varied backgrounds, applying image segmentation techniques to reconstruct landscape images. This repository of reconstructed images was instrumental in our aim to enhance user satisfaction. By measuring the similarity between user-input images and our database, we provided tailored frame recommendations, showcasing the practical application of our methodology. However, our project encountered certain limitations that warrant attention for future improvements. The foremost among these is the constraint on the number of images used. Given the vast diversity of real-world photos, the necessity for a more expansive image dataset becomes apparent. This expansion would involve not only increasing the quantity of images but also broadening the range of categories represented. Such an enhancement would allow for more nuanced and varied frame recommendations, catering to a wider array of user preferences. Another significant aspect is the dependency of our frame recommendation system on the specified threshold. While setting a higher threshold could potentially yield more accurate matches, it inversely affects the quantity of recommendations provided. This trade-off between accuracy and choice availability is a critical area for future development. Striking an optimal balance is essential to maintain both the quality of recommendations and the diversity of options available to users.

**Future Research Directions for Enhanced System Efficiency and Performance**

1. Optimizing Clustering Algorithms: A crucial area for future research lies in enhancing the computational efficiency of clustering algorithms used in our system. One promising approach is the implementation of a sliding window technique. Unlike traditional pixel-to-pixel calculations, which can be computationally intensive, a sliding window method processes chunks of the image in each step. This approach can significantly reduce the computational load and accelerate the processing time, especially for high-resolution images. Additionally, exploring other optimization strategies such as dimensionality reduction techniques (e.g., PCA or t-SNE) or more efficient clustering algorithms like DBSCAN or OPTICS, which are known for their scalability and effectiveness in handling large datasets, could further improve the system's performance.

2. Leveraging Deep Learning for Frame Recommendation: The integration of deep learning offers a transformative potential for the frame recommendation process. Convolutional Neural Networks (CNNs), renowned for their prowess in image analysis, could be employed to extract intricate features and patterns from images, surpassing the capabilities of traditional histogram-based methods. For instance, implementing a CNN-based feature extraction followed by a similarity assessment using techniques such as Siamese Networks or Triplet Loss could yield more precise and contextually relevant frame recommendations. Additionally, our initial goal - Generative Adversarial Networks (GANs) could be explored for generating new frame designs that better match user images, thereby expanding the variety of frames available. These deep learning models, equipped with advanced training techniques like transfer learning and data augmentation, can adapt to diverse image styles and compositions, offering a more robust and versatile recommendation system.

Incorporating these advanced methodologies not only promises to overcome the limitations of current threshold-based and clustering approaches but also paves the way for a more sophisticated, accurate, and user-responsive frame recommendation system. Such developments would significantly contribute to the fields of image processing and personalized content recommendation, setting new benchmarks for efficiency and user engagement.

## 7. REFERENCES

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. 2014 CVPR, Rich feature hierarchies for accurate object detection and semantic segmentation, DOI= https://arxiv.org/abs/1311.2524

[2] Yann Lecun, Léon Bottou, Yoshua Bengio, Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86 (11), pp.2278-2324. ff10.1109/5.726791ff. ffhal03926082f

[3] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Neural Networks for Biomedical Image Segmentation, 2015, MICCAI

[4] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.

[5] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[6] D. Deng, "DBSCAN Clustering Algorithm Based on Density," *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, Hefei, China, 2020, pp. 949-953, doi: 10.1109/IFEEA51475.2020.00199.

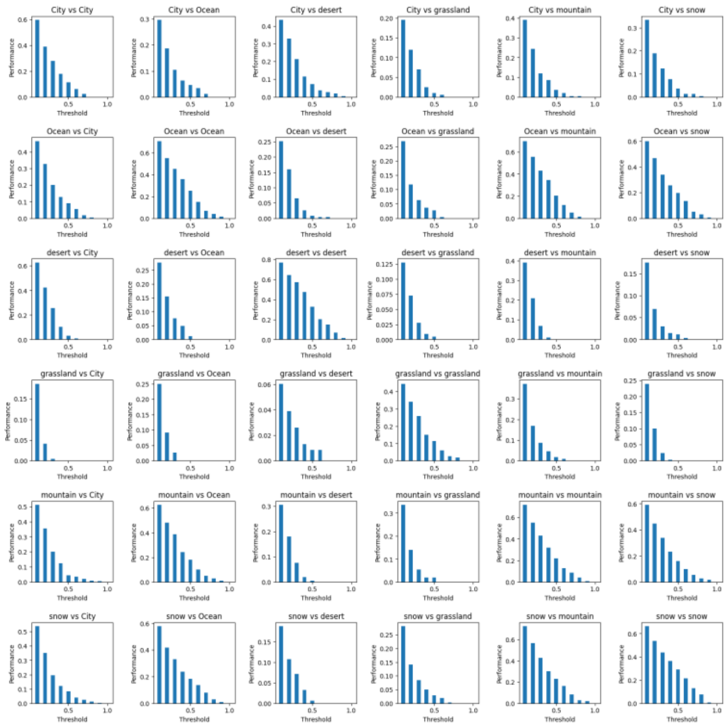
[7] S. Babichev, B. Durnyak, V. Zhydetskyy, I. Pikh and V. Senkivskyy, "Application of Optics Density-Based Clustering Algorithm Using Inductive Methods of Complex System Analysis," *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine, 2019, pp. 169-172, doi: 10.1109/STC-CSIT.2019.8929869.

# APPENDIX A

## Recommended Frame based on similarity

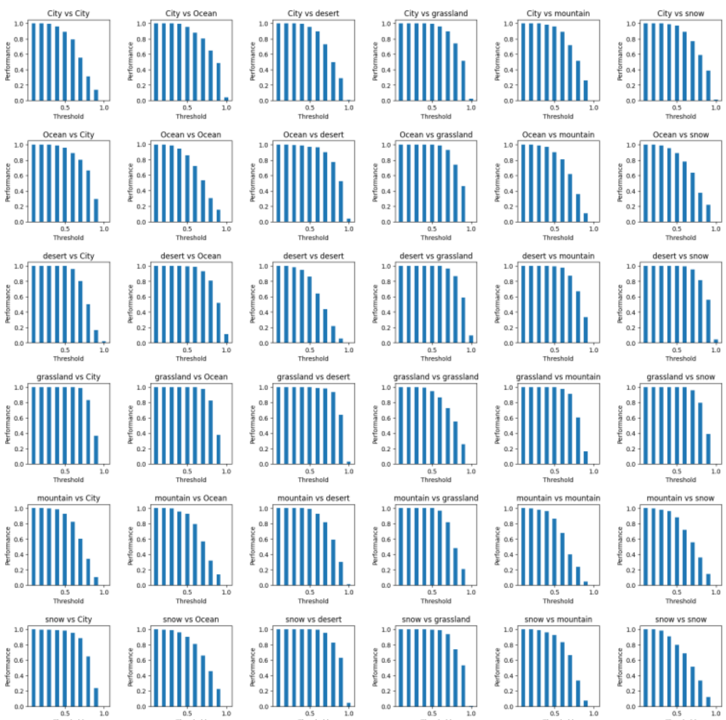
# APPENDIX B

**Performance per threshold between categories based on correlation coefficients (predicted number of times)**



# APPENDIX C

**Performance per threshold between categories based on bhattacharyya distance (predicted number of times)**



# APPENDIX D

**Confusion matrix per threshold (threshold value increases as it goes down to the right.)**