International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I10P115

Volume 72 Issue 10, 94-100, October 2024

Original Article

A Picture is worth a thousand colors - Using Computer Vision to Color Tag E-Commerce Products

Sachin More[1], Jeyvinth Manoj Rayan[2]

[1]*Dollar General Corp, Nashville, TN, USA.*
[2]*KaviGlobal , Chicago, IL,USA.*

[1]*Corresponding Author : sachin.more@ieee.org*

Received: 31 August 2024 Revised: 03 October 2024 Accepted: 19 October 2024 Published: 30 October 2024

Abstract *- Product metadata tagging is one of the core components required for the e-commerce industry to fuel product discoverability, which drives key initiatives like personalized recommendations, search relevancy, pricing, SEO, decreasing customer dissatisfaction, etc. Color Tags of products are highly prominent in the Fashion section, where the optics of the product are given higher precedence. In a marketplace setup, it is difficult to ensure a good quality of color tagging for the products owing to the scale of products, language constraints, manual tagging errors, etc.*

In this framework, we develop models that can extract the dominant colors of the apparel from the Product Images and enable automatic tagging of the products with these colors. Our approach has the following steps:1) U2-Net Image Segmentation algorithm to segment the foreground from the fashion image. 2(a)Extract the Hex color codes and RGB values for the extracted foreground and map the dominant RGB value to the nearest standardized color name using KD Tree-based clustering method. 2(b)Predict the Global color using a classification approach with a fine-tuned EfficientNet Model. We use this pipeline to predict the color label for the test dataset, which has pre-tagged, human-validated color labels that can be used as our golden set for validation. We evaluate the misclassification rate from this golden set against our predictions to assess the model performance. The approach is expected to achieve 26% percent of missing color imputation and suggest 11% percent of mislabeled color tags. This further translates to a 5% percentage of search results showing relevant results, leading to a considerable increase in the conversion owing to relevant results being shown.

Keywords *- Image Segmentation, color clustering, Dominant Color extraction, E-commerce Fashion color extraction.*

1. Introduction

With the ever-increasing volumes of sellers, products, and customers in the e-commerce marketplace, it is even more imperative that e-commerce products also contain the right tagging of key metadata features like Price, Brand, Color, Size, etc. However, in a real-world scenario, it is hard to expect the best quality of these product attributes, being supplied by the sellers when they onboard the products to the marketplace, owing to the volumes of the product catalogue, language barriers, and the lack of manual resources and time. The quality of the product attributes can be impacted by missing values and mislabeled data, which could be a result of assigning the attributes of a single variant of a product to all the other variants, causing erroneous labeling. These missing/improper labeling of the product metadata could have many downstream implications on use cases like search, recommendations, pricing, promotions, etc. The lack of quality of the product metadata could impair the query match in the elastic search index, thus missing out on filtering and selecting many relevant products and reducing the Recall for the search. The quality could also affect some of the popular contextual recommendations like Similar Items, which are crucially dependent on the metadata of the products to arrive at a notion of similarity between the products. Applications like competitive pricing require one to match products from the competitor sites and assign the optimal pricing strategy, which requires rich metadata for the product match. The availability of the metadata itself on the Product Description page could ensure a higher probability of conversion of the product than the lack of it.

In the fashion industry, color plays a pivotal role in influencing customers to visualize the products and aiding them in decision-making. It has been found that about 30% of search keywords in fashion consist of color references. Also, a vast majority of similar item recommendations seem to have higher Click Through Rate (CTR) by showing items of the same color. In a typical marketplace where there are a lot of third-party seller products onboarded, one cannot ensure the quality of these color tags, given the scale of products being uploaded daily, language constraints, lack of resources, etc. While manual color tagging is un-scalable and error-prone, automated color tagging by itself is a non-trivial task, owing to many noisy attributes and a lack of ground truth for any supervised learning framework.

To solve this, we propose a computer vision-based framework that uses the images of the product that is displayed on the Product Description Pages to extract the colors and provide the right color tagging of the product. The color Extraction process on the Images involves the following stages:

- Segmenting and noise removal, such as background, irrelevant/non-focused clothing, accessories, etc.
- Retrieval of the RGB value on the segmented image.
- Consolidating the RGB values to the available color palette

Our method employs a U2-Net segmentation to segment all the non-interesting parts of the image, like skin, background, extra clothes, and accessories (anything that is not advertised by the image). This provides us with just the masked image of the product that is in focus. RGB values for these segmented pixels are extracted, and the dominant color is determined using distribution analysis. This dominant RGB value is then mapped to the nearest standardized catalogue color name using KD Tree.

We also explored converting this color prediction problem into a classification model. For this, CNN architecture-based Efficient net algorithm was used and tested on a fashion dataset that has pre-tagged, human-verified color labels.

2. Related work

Work dedicated to explicitly making use of colors in the fashion industry using AI is limited. The problem of extraction of the right segments to denoise the image was explored in unsupervised methods like [11] which proposed a Gaussian Mixture Model-based approach to segment and extract the colors from the images. [12] suggested an estimation based on Constrained Delaunay Triangulation (CDT) estimations based on graph cuts with foreground and background seeds. [13] proposed methods to resolve the human pose variances and background removal using some unique HOG, LBP, and color features for each body part. [14] discussed the clothing parsing problem using a retrieval-based approach. For a query image, similar styles are found from a large database of tagged fashion images and are then used to recognize clothing items in the query. This approach combines parsing from pre-trained global clothing models and local clothing models learned on the fly, as well as transferred parse masks from retrieved examples. Other works, as in [3], pointed out the advantages the fuzzy-c-means can provide over k-means, even though they aimed at using it for color image segmentation. [15] GrabCut combines the texture information and the edge contrast information to arrive at an iterative graph cut to achieve an enhanced quality of segmentation. Once the right segmentation is achieved, we need to deal with the problem of extraction of the right color.[16] proposed a method for extracting color themes from images using a regression model trained on color themes that were annotated by people. To collect data for their work, the authors asked people to extract themes from a set of 40 images consisting of 20 paintings and 20 photographs. However, such a data-driven approach may suffer from generalization issues because millions of colors exist in the real world. Automatic palette extraction has been the focus of [1, 2], in which a hue histogram segmentation method has been used. The hue histogram segmentation has a disadvantage in that it is affected by the saturation and intensity values, as well as singularities when the saturation values are zero. To consolidate the color gradients into a globally accepted color palette, we explored various color clustering approaches. [17] used the k-means algorithm on an input image to generate a palette consisting of a small set of the most representative colors. An iterative palette partitioning based on cluster validation has been proposed in [18] to generate color palettes.[19] discussed Clothing color attribute representation from Pedestrian images was illustrated using a novel process of combining low-level features (color histograms) with mid-level semantic descriptors (color name distribution).[20] talked about Automatic machine-learning methods are developed to generate color palettes for a fashion show based on runway images. A set of ground-truth data to test the models was constructed based on asking each of the 22 participants to select three colors to represent each of the 48 images from a particular fashion show. The problem with clustering algorithms is to know the number of colors prior. A low number of clusters will result in an incorrect color number if the image has more colors than the one used to build the clustering model. On the other hand, using a high number of clusters has the drawback of extracting several colors and makes it difficult to extract colors accurately.

3. Color Extraction Framework

The colors of clothing items seen in images usually are not accurate because of distortions. Separating the "foreground" and the "background" is a difficult task unless the position of the target object is known in a picture, which requires a huge number of annotated images. Even after the determination of the foreground image, various challenges are involved in extracting the clothes owing to a high level of Noise.

There could be various types of noise presented, such as "Skin color of Model", "Hair Color of Model", "Additional Accessories worn by Model", and "Background of Model". Refer to Figure1(a) for an example. Noise can also be at the Image quality level, such as "Compressed Image", "Illumination", "Filters applied over the image" etc.

Fig. 1(a) Different noises in a fashion image

Further segmentation of the Foreground images is necessary since they may contain varieties of apparel such as T-shirts, Shirt, Sweatshirt, Jacket, Jeans, Tops, etc., and objectifying different types of clothes requires huge data for each of these types of clothes to increase the number of classes or requiring a hierarchical classification.

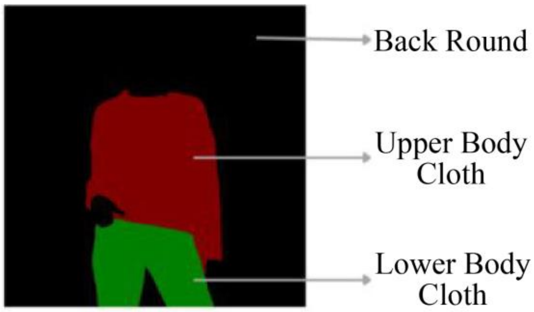

Fig. 1(b) Masked image after image segmentation

To bring in a generalization and restrict the number of classes, we mapped an image into four categories - "Upper Body Cloth", "Lower Body Cloth", "Full Body Cloth" and "Background". Our approach primarily consists of image segmentation, which eliminates the noise and extracts the focus segment of interest, followed by color prediction.

3.1. Image Segmentation

We use Image segmentation to extract the focused object from the image by classifying the segments into four classes, namely - "Upper Body Cloth", "Lower Body Cloth", "Full Body Cloth", and "Background". There have been many studies on Image segmentation techniques, as discussed in the related work section. However, recent advancements have seen some of the deep- learning-based image segmentation methods like MaskRCNN [4], U-Net [5], YOLOACT [21], RetinaMask [22], etc. Most of these Image segmentation algorithms adopt a similar framework, which involves the network learning the feature representation of different classes based on the attribute classification followed by a transfer learning paradigm, which learns to classify the signals from the noise. We used the U2-Net [6] Image segmentation model to segment out the upper body, lower body, full body cloth, and background. U2-Net utilizes Salient Object Detection (SOD), which is a task-based visual attention mechanism in which algorithms aim to explore objects or regions more attentively than the surrounding areas on the scene or images. It was built with the inspiration of U-Net, a novel network called Residual U-block, RSU, to capture intra-stage multi-scale features. The obtained features are down-sampled so that the contextual features get retained, and the result does not get affected by the extracted fine details of the picture, i.e., color, texture, etc. After the features are extracted, they are encoded and decoded after up-sampling. We chose U2-Net over KD-tree because the computation of the model is much less as it focuses on the down-sampled features, and it has been proven to be effective for extracting and separating "Foreground" and "Background" from images effectively.

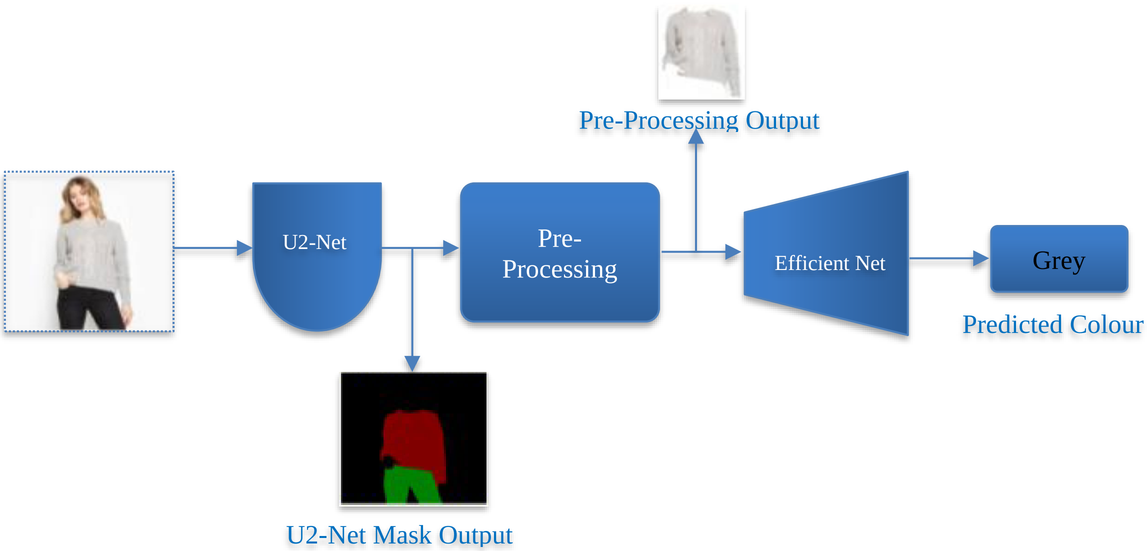

Fig. 2(a) Architecture of proposed approach

The full architecture of the U2-Net contains 6 encoders, 5 decoders, and 1 saliency map fusion module, which contains the sigmoid function and 1 convolution layer of 3*3 dimension. We fine-tuned the U2-Net model on the iMaterialist, and the model learned to segment out the upper body, lower body, full body, and background objects. Finally, it produced the segmentation mask. Figure 1(b) explains the output of U2-Net.

3.1.1. Color Prediction

Once the region of interest (i.e., Upper body, Lower body, and Full Body cloth) was extracted, we experimented with the following approaches to get the colors from the segmented cloth objects:

KD Tree Approach

The colors of any object are often represented as a combination of red, blue, and green values. Each of these color values is an integral value bounded between 0 and 255. To identify the right color space in segmented cloth objects, our experiments addressed the extraction of RGB values from the segmented images and used the KD Tree algorithm [7], which is a tree-based algorithm used for the spatial division of data points and their allocation into certain regions, to cluster the predefined RGB values of CSS colors into space and thereby predicting the nearest RGB value for a given segmented image.

Once the nearest RGB value was extracted, we mapped it to the predefined CSS color using Euclidean distance to provide the color name of the object. The challenges we faced while performing the KD Tree approach were that we didn't have many of the RGB values for different variations of colors, and color mapping for names became difficult for the new variations of RGB values. Hence, we continued to explore our experiment results with a Transfer Learning approach.

Transfer Learning Approach

Extracting the exact color from a Fashion image is always a challenging task. A Global color, like Blue, has a lot of variants like deep sky blue, light steel blue, midnight blue, dark slate blue, etc. The extraction of these variants is difficult compared to the previous approach (KD Tree Approach). Generalizing these variations as a single global color would be a better way to extract the right color from the image. Hence, our next experiment is an upgrade toward learning the variants of different colors and predicting the right global color for images.

Data Augmentation: We used various data augmentation techniques to increase the number of sample images as well as to generate more variation of images, which would lead to the robustness of models like Flip, Transpose, RandomRotate, ShiftScaleRotate, MotionBlur, MedianBlur, Blur, RandomBrightness, RandomContrast, GaussianNoise, OpticalDistortion, GridDistortion.

Modeling: To model the color classification problem from an image, we used SOTA EfficientNet-B3 [8] pre-trained backbone. It is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. Unlike conventional practice that arbitrarily scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients.

With the scaling capability of the EfficientNet model, it can learn and focus on the color in a cloth-segmented image with white background Figure 2(b), and it can identify multiple colors for mixed-pattern or multi-color clothes. The EfficientNet model was initially trained on the ImageNet Dataset. The ImageNet dataset is a very large collection of human-annotated photographs designed by academics for developing computer vision algorithms.

The datasets comprise approximately 1 million images and 1,000 object classes. The datasets used in challenge tasks are sometimes varied (depending on the task) and were released publicly to promote widespread participation from academia and industry. We added a new Feed Forward layer and removed the final layer of EfficentNet. We used SGD for updating the EffcientNet architecture, with a learning rate of 0.01, momentum of 0.95, weight decay of 1e-6, and Nestrov=True with a batch size of 32. For detailed metrics on model convergence, refer to Figure 2(c)

Fig. 2(b) Focus gradients of the EfficientNet model on different apparel images to predict the color of the apparel

With the Transfer Learning approach, we were able to predict the right color for cloth-segmented images and multi-colors for mixed-pattern clothes. Figure 2(a) explains the proposed architecture with image segmentation using U2-Net and color prediction using EfficientNet.

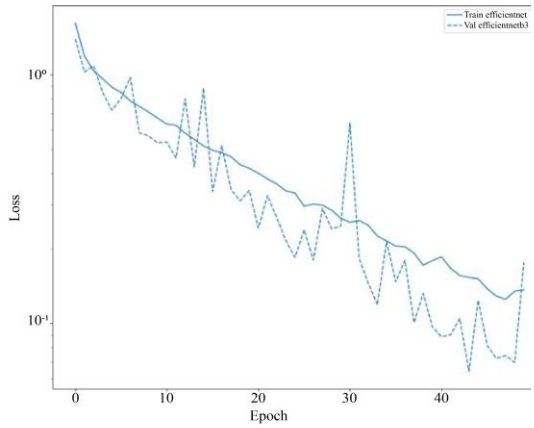

Fig. 2 (c) Graph showing the convergence of the EfficientNet model depicting Training and Validation Loss for each epoch

4. Dataset

For Image segmentation: U2-Net Experiment was fine-tuned on 45k images from the iMaterialist (Fashion) 2019 Dataset (Kaggle.com). The iMaterialist dataset consists of fashion objects and their descriptions obtained by crawling the web. It contains 46 apparel objects (27 main apparel items and 19 apparel parts) and 294 related fine-grained attributes. This dataset also has polygon annotations of clothing images from daily life, celebrity events, and online shopping websites, which domain experts and crowd workers label for fine-grained segmentation.

For color prediction (Transfer Learning approach): We used the Fashion Product Images Dataset and Digikala Fashion dataset, which was available on Kaggle.com (WWW), for color classification. The Fashion Dataset consists of professionally shot high-resolution product images and has multiple label attributes describing the product, which were manually entered while cataloguing. It also has descriptive text that comments on the product characteristics. The Digikala dataset is shared publicly by Digikala (one of the biggest online e-commerce markets in Asia) for an online challenge. We clubbed all available products from these two datasets in a single dataset and separated them into different color folders based on their available metadata. The total sample was 15409 images, out of which 12327 were part of the training dataset and the rest, 3082 were used for the validation set.

Table 1. The distribution of labels in the dataset

Label	Number of samples
Beige	499
Black	843
Blue	1170
Brown	1180
Cream	396
Green	1214
Gray	1179
Maroon	495
Lavender	157

5. Results and Discussion

To evaluate the model, we made use of the following performance metrics.

Table 2. Performance metrics

Validation Metric	Percentage Value
Precision	96.03%
Recall	95.13%
Average Precision	97.98%
AUC	99.76%

Refer to Figure.4(a) and Figure 4(b) for graphs.

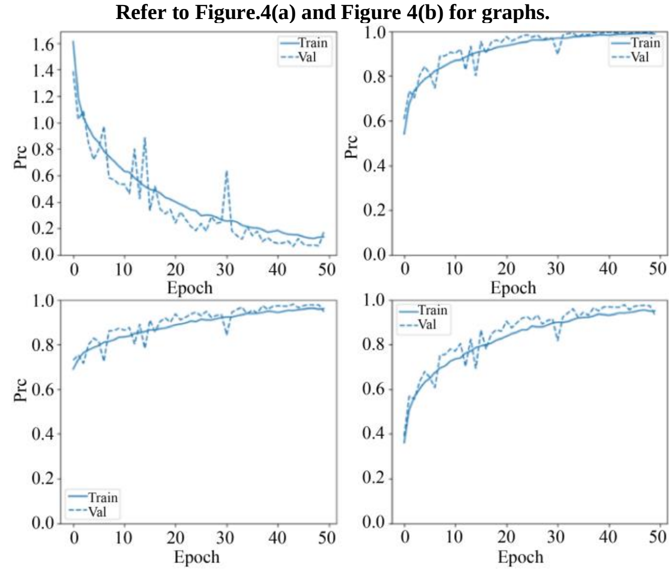

Fig. 4(a) Precision, Recall, Average Precision and AUC for Training and Validation data with augmentation.

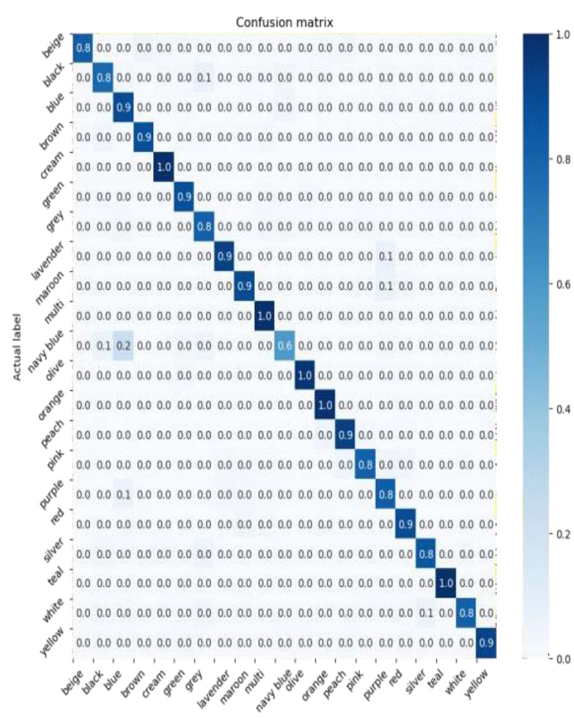

Fig. 4(b) Confusion matrix for each color class

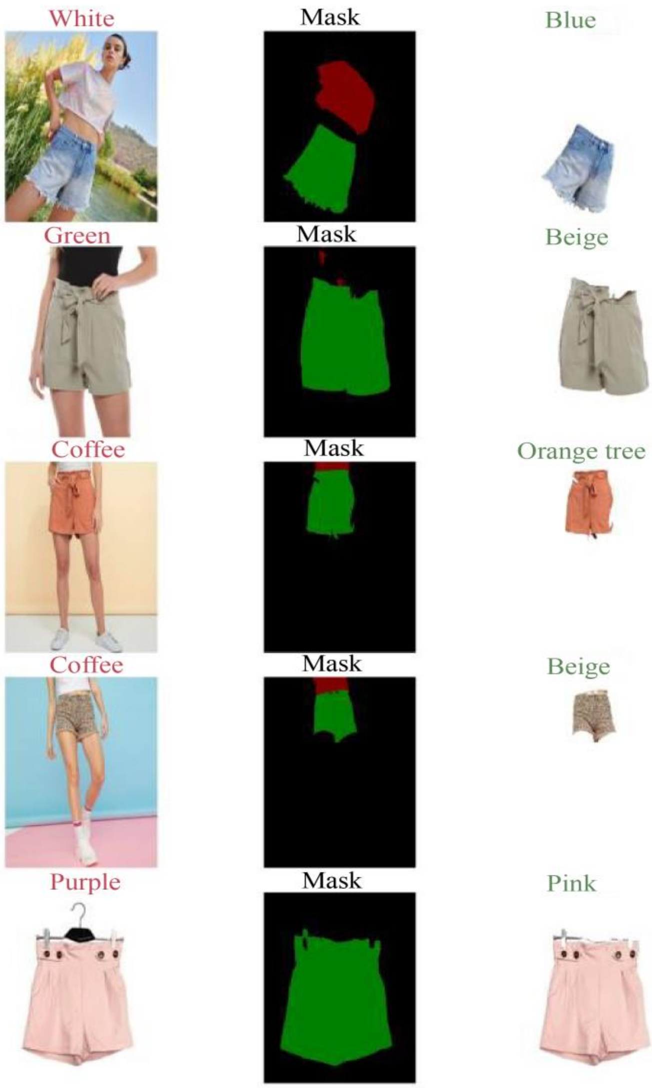

Fig. 4(c) Image segmentation and color prediction results on our dataset

We tested the approach on one of the e-Commerce real-world fashion images. Figure 4(c) highlights the input images (left) with color attributes (red color) mentioned on the website, mask images (middle) from the U2-Net image segmentation model, and Pre-Processed images with predicted color (green color) from EfficientNetB3 trained model. This model not only helped impute the missing colors in our catalog, but also predict the mislabeled color tags, with 95% confidence interval, thereby helping us improve the overall data quality of color attributes in our catalog.

6. Conclusion

The fundamental aim of this paper was to propose a computer vision framework for automating the color tagging of e-commerce products, a crucial aspect of enhancing product discovery and user experience. The framework leverages advanced techniques like U2-Net image segmentation for precise foreground extraction and EfficientNet-based transfer learning for accurate color prediction. This approach effectively addresses challenges such as noise in images, variations in color representation, and the need for scalability. Experimental results demonstrate the efficacy of the proposed framework in accurately predicting color tags, with significant improvements in the imputation of missing colors and correction of mislabeled tags. This translates to enhanced search relevance, personalized recommendations, and, ultimately, increased customer satisfaction and conversion rates. Future research can explore further refinements to the framework, such as incorporating additional features like texture and pattern recognition and addressing challenges related to complex lighting conditions and diverse product categories. Overall, this study contributes to the growing field of computer vision applications in e-commerce, offering a promising solution for automating and optimizing product metadata management.

References

[1] Julie Delon et al., “Automatic Color Palette,” *Proceedings IEEE International Conference on Image Processing 2005*, Genova, Italy, pp. 2-706, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[2] Julie Delon et al., “Automatic Color Palette,” *Inverse Problems and Imaging*, vol. 1, no. 2, pp. 265-287, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hoel Le Capitaine, and Carl Frélicot, “A Fast Fuzzy C- Means Algorithm for Color Image Segmentation,” *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-11)*, Atlantis Press, Aix-les- Bains, France, pp. 1074-1081, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kaiming He et al., “Mask R-CNN,” *Proceedings IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Proceedings International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 234-241, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xuebin Qin et al., “U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection,” *Pattern Recognition*, vol. 106, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Martin Skrodzki, “The KD Tree Data Structure and A Proof for Neighborhood Computation in Expected Logarithmic Time,” *arXiv*, pp. 1-12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mingxing Tan, and Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv*, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[9] Ujjal Kr Dutta et al., "Color Variants Identification in Fashion E- Commerce via Contrastive Self-Supervised Representation Learning," *arXiv*, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[10] Mohammed Al-Rawi, and Joeran Beel, "Probabilistic Color Modelling of Clothing Items," *Proceedings Recommender Systems in Fashion and Retail*, Ireland, pp. 21-40, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[11] Marco Manfredi et al., "A Complete System for Garment Segmentation and Color Classification," *Machine Vision and Applications*, vol. 25, pp. 955-969, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[12] Zhilan Hu, Hong Yan, and Xinggang Lin, "Clothing Segmentation Using Foreground and Background Estimation Based on The Constrained Delaunay Triangulation," *Pattern Recognition*, vol. 41, no. 5, pp. 1581-1592, 2008. [CrossRef] [Google Scholar] [Publisher Link]
[13] Si. Liu et al., "Street-To-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set," *MM '12: Proceedings of the 20th ACM International Conference on Multimedia*, pp. 1335-1336, 2012. [CrossRef] [Google Scholar] [Publisher Link]
[14] Kota Yamaguchi et al., "Retrieving Similar Styles to Parse Clothing," *Proceedings IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, pp. 1028-1040, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[15] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, "GrabCut": Interactive Foreground Extraction using Iterated Graph Cuts," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 309-314, 2004. [CrossRef] [Google Scholar] [Publisher Link]
[16] Sharon Lin, and Pat Hanrahan, "Modeling How People Extract Color Themes from Images," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, New York, USA, pp. 3101-3110, 2013. [CrossRef] [Google Scholar] [Publisher Link]
[17] Qing Zhang et al., "Palette-Based Image Recoloring Using Color Decomposition Optimization," *IEEE Transactions on Image Processing*, vol. 26, pp. 1952-1964, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[18] Ju-Mi Kang, and Youngbae Hwang, "Hierarchical Palette Extraction Based on Local Distinctiveness and Cluster Validation for Image Recoloring," *Proceedings 2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2252-2256, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[19] Mu Gao et al., "A Hybrid Approach to Pedestrian Clothing Color Attribute Extraction," *Proceedings 2015 IAPR International Conference on Machine Vision Applications*, Tokyo, Japan, pp. 18-22, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[20] Peihua Lai, and Stephen Westland, "Machine Learning for Colour Palette Extraction from Fashion Runway Images," *International Journal of Fashion Design, Technology and Education*, vol. 13, no. 3, pp. 334-340, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[21] Daniel Bolya et al., "YOLOACT: Real-time Instance Segmentation," *arXiv*, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[22] Cheng-Yang Fu, Mykhailo Shvets, and Alexander C. Berg, "RetinaMask: Learning to Predict Masks Improves State-of-the- Art Single-Shot Detection for Free," *arXiv*, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[23] Wei-Dong Liu, and Xi-Shui She, "Application of Computer Vision on E-Commerce Platforms and Its Impact on Sales Forecasting," *Journal of Organizational and End User Computing (JOEUC)*, vol. 36, no. 1, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[24] Abon Chaudhuri et al., "A Smart System For Selection Of Optimal Product Images In E-Commerce," *Proceedings 2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[25] Martin Danelljan et al., "Adaptive Color Attributes for Real-Time Visual Tracking," *Proceedings 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1090-1097, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[26] Ivan Kuteynikov, and Marina Yashina, "Computational Complexity Optimization of Vehicle Video-Recognition Algorithm Using the Virtual Detectors Method," *Proceedings 2024 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*, Vyborg, Russian Federation, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[27] Abon Chaudhuri et al., "A Smart System for Selection of Optimal Product Images in E-Commerce," *Proceedings 2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, pp. 1728-1736, 2018. [CrossRef] [Google Scholar] [Publisher Link]