

Content-based image retrieval for fabric images: A survey

Silvester Tena¹, Rudy Hartanto², Igi Ardiyanto³

^{1,2,3}Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia

¹Department of Electrical Engineering, Universitas Nusa Cendana, Indonesia

Article Info

Article history:

Received Jun 3, 2021

Revised Jul 26, 2021

Accepted Aug 4, 2021

Keywords:

CBIR

CNN

Fabric image

Feature extraction

Traditional descriptor

ABSTRACT

In recent years, a great deal of research has been conducted in the area of fabric image retrieval, especially the identification and classification of visual features. One of the challenges associated with the domain of content-based image retrieval (CBIR) is the semantic gap between low-level visual features and high-level human perceptions. Generally, CBIR includes two main components, namely feature extraction and similarity measurement. Therefore, this research aims to determine the content-based image retrieval for fabric using feature extraction techniques grouped into traditional methods and convolutional neural networks (CNN). Traditional descriptors deal with low-level features, while CNN addresses the high-level, called semantic features. Traditional descriptors have the advantage of shorter computation time and reduced system requirements. Meanwhile, CNN descriptors, which handle high-level features tailored to human perceptions, deal with large amounts of data and require a great deal of computation time. In general, the features of a CNN's fully connected layers are used for matching query and database images. In several studies, the extracted features of the CNN's convolutional layer were used for image retrieval. At the end of the CNN layer, hash codes are added to reduce search time.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Silvester Tena

Department of Electrical Engineering and Information Technology

Universitas Gadjah Mada

Grafika 2 St., Kampus UGM, Yogyakarta 55281, Indonesia

Email: siltena@mail.ugm.ac.id

1. INTRODUCTION

Several studies have been carried out on fabric image retrieval, especially on its identification and classification using visual features. Fabric image retrieval is useful for e-commerce, inventory management, and fabric product design. Furthermore, with the rapid development of e-commerce, product retrieval has become increasingly important for online shopping.

Keyword-based search methods are challenging due to the complicated and impractical nature of specific visual characteristics verbally defined using few keywords. A user-provided sample image is another approach used to retrieve context-based textile images from the database visually. Presently, there are numerous existing methods used to extract low-level features, including local binary pattern (LBP), histogram of oriented gradients (HOG), speeded-up robust features (SURF), scale-invariant feature transform (SIFT), fisher vector (FV), gray-level co-occurrence matrices (GLCM), Gabor wavelet, discrete fourier transform (DFT), a bag of words, vector of locally aggregated descriptors (VALAD), and generalized search trees (GIST). Extraction methods based on low-level features are categorized into traditional extraction methods. However, recent studies have led to the development of convolutional neural networks (CNNs), one of the most widely

used models in deep learning. The CNN approach uses semantic knowledge in an image to process the extraction of high-level features.

One challenge in content-based image retrieval (CBIR) is the semantic gap between low-level visual features and high-level human perceptions. Traditional image retrieval methods focus mainly on features like color, texture, shape, and spatial information. Unlike conventional approaches, such as CNN's, which have the potential to learn hierarchical and high-level features. Recently, it has made significant contributions to CBIR feature extraction. High-level features-such as illumination, coarse-to-fine fabrication, sketch, style, and clothing translation-present challenges for image retrieval research. These high-level features are extracted despite the high requisite computational complexity [1]. Generally, CBIR includes two main components: 1) an image extraction algorithm; and 2) the selection of a suitable distance or similarity measurement method [2].

Therefore, this research aims to classify methods capable of extracting low-level and high-level features using traditional and CNN methods, respectively. It also aims to identify the strengths and weaknesses of both traditional and CNN feature extraction methods by presenting a survey of fabric image retrieval and relevant categories, focusing on the last five years.

2. RESEARCH METHOD

The paper is written based on the literature review. The topic chosen relates to image retrieval. Feature extraction methods are categorized into traditional feature extraction methods and CNN methods to facilitate discussion. The feature extraction method chosen is all methods that handle low-level features and high-level features. Discussions about various extraction methods and features used for image retrieval as knowledge insights for further research development. Journal and paper conferences were obtained from several publishers, namely IEEE Xplore, Sage, MDPI, Scopus, and Science Direct. The query used is ["Fabric/batik/textile/clothes/fashion"] AND "retrieval". Journals or conference papers selected according to criteria are strong links to the topic of fabric image retrieval, from reputable publishers and research results of the past five years. The selected papers are grouped based on feature extraction using the traditional feature extraction and CNN method. The most decisive part of image retrieval is the selection of feature extraction methods and similarity measurement. The selection of features and the extraction methods are adapted to the image type. The dataset images used in the cited paper are fabric, batik, fashion, clothes, and textile. Some examples of the dataset used in the paper are cited in Figure 1.

The remaining sections of the research are organized. Section 3 describes traditional feature extraction, while section 4 analyzes CNN-based feature extraction. Traditional and CNN descriptors are compared in section 5. Finally, a conclusion is presented in section 6.



Figure 1. Sample image in dataset; a) fabric image [2], b) batik image [3], c) clothes image [4], and d) fashion image [5]

3. FEATURE EXTRACTION BASED ON TRADITIONAL METHODS

The method of indexing the contents of an image database is known as feature extraction. Mathematically, it is the encoding of an n -dimensional vector called the feature vector, which is calculated using image processing and analysis techniques. It is also used to compare one image with another. Table 1 (Appendix) shows the summary of feature extraction based on traditional methods.

The research on shape-based image retrieval carried out by [6] compared several methods, such as HOG, pyramid HOG (PHOG), and normalized PHOG (NPHOG). The first step was to detect an object's face using the Viola-Jones method, with the center point of the face chosen as the reference for normalization. The grid normalization process is carried out at level-1 with reference to the center point face and is different from conventional grid partitions. The results showed that NPHOG generally outperformed the other methods,

although it had low performance in the long dress category, where the image boundaries were unclear. Therefore, the authors stated that combination with other features tends to improve retrieval accuracy. A study carried out by [4] presented a novel sketch-based clothing image retrieval algorithm. Unlike conventional image recovery strategies focused on keywords and content, sketch-based clothing image recovery offers an intuitive and natural way to help users explain their search requirements. This algorithm also used the conditional random field method to train a dataset with semantic component labels and create a sketch-image pairs. Query images were created by entering clothes sketches using mobile sensors. According to the authors, future work is needed to develop a rapid, parallel implementation for the segmentation of the clothing components. However, the method was not tested on skewed and sparse sketch models.

Image retrieval was performed using texture features [5]-[8]. The research conducted by [7] used a texture feature for batik image retrieval, while the feature extraction methods included Gabor, log-Gabor, GLCM, and LBP. The query dataset was generated from each template based on translation, scale, and rotation with Euclidean distance (ED) used to calculate the similarity. The careful selection of log-Gabor features, GLCM, and Gabor features in this study provided complete coverage of the spatial frequency spectrum for batik motif retrieval. The results showed that the fusion of several feature descriptors provided the highest accuracy compared with individual feature descriptors.

The research carried out by Zhang *et al.* [8] combined LBP and DFT during the texture feature extraction process. The equalization histogram aimed to improve the quality of the fabric image with the benefits of the Fourier and LBP features combined using the product of two distances. The result showed that combining extraction methods improved retrieval accuracy. Furthermore, the research carried out by [9] aimed to determine an effective and efficient texture-based method of solving the problem of fabric image retrieval through mobile devices. Topography morphology and invariant Gaussian-Hermite moments analysis were used for texture feature extraction. Similarity matching was performed using the Kullback-Leibler distance. According to the authors, future research needs to combine color features to display the images of the same fabric close together during the retrieval process. On large datasets, the combination of features improves retrieval accuracy. Tariq *et al.* [10], research was conducted to retrieve the textile image using color and texture features. The feature extraction methods used included HOG, hue, saturation, value (HSV), and LBP, with a support vector machine (SVM) used as a classifier. Similarity measurement was based on the true and false positive per image rate.

A study of wool fabric images using color features was conducted by [11]. The research showed that feature extraction methods included dominant colors (DCs) and color moments (CMs) with diverse, complicated, dark, and blurry wool fabric appearances. The authors recommended the combination of color and texture features in future research to improve retrieval performance. Mutia *et al.* [6] stated that precision is higher for color features than the texture in the image extraction process and lowest in the striped fabric category, where embedded lines cause image similarity.

In a research carried out by [12] using color and spatial information features, CM and GIST were used to extract color characteristics and spatial information, respectively. Furthermore, Gabor filters were also used to obtain spatial information. Similarity matching was performed using ED for each CM and GIST, with the weight accumulated, thereby leading to a comparison of the query and database images. The combination of these two feature extraction methods outperformed other hybrid schemes for printed fabric images. Conversely, the retrieval time is longer than state of the art, therefore, to improve the accuracy of printed image retrieval, the authors combined color, texture, and spatial information features of fabric images. The research focused on fractal image capture, and color features [13] with HSV used to provide color features. Fractal-based texture features of batik image represent fractal dimension and lacunarity. The combination of these two methods increased retrieval accuracy with similarity matching performed using the Canberra distance (CD). The authors aim to investigate more features to be extracted in future research and determine the most efficient method with acceptable accuracy and computing time.

Traditional descriptors for lacy and embroidered fabric image were combined in a research carried out by [14], which combined GIST and SURF for spatial and texture features. The fabric image dataset contained many patterns involving repetition, with an outstanding performance of the GIST feature description for image retrieval with identical patterns. Furthermore, the SURF method was used to treat rotated images, and their combination enhanced the retrieval accuracy for rotated images. The authors stated that fabric wrinkle is a further research challenge.

The combination of the perceptually uniform histogram and motif co-occurrence histogram methods was termed hybrid histogram descriptor (HHD) [15]. Uniform histogram extracted using the information on the color, and edge orientation in perceptual regions produced a motif co-occurrence histogram by computing a pair of motif pattern probabilities. Similarity estimation was based on Canberra distance, while HHD outperformed a variety of other conventional approaches. Both feature extraction methods led to increased dimensions and computation time. Therefore, future research needs to address computation time using the

locality-sensitive hashing (LSH) method. The research conducted by [16] was on fabric image retrieval and classification using an SVM. Furthermore, traditional descriptors were used to extract color, shape, and texture features. The feature extraction methods include color histogram, HSV, color auto-correlogram, CM, Gabor wavelet, and wavelet moments. A total of 190 dimensions were obtained after combined feature extraction with the SVM method to retrieve more relevant fabric images in the database. According to the authors, the training time needed for large datasets is quite large. Research on women's clothing using the HOG method and region of interest (ROI) was conducted by [17] with the Viola-Jones algorithm used to detect the ROI, which increased retrieval performance. The results showed that the proposed method had a higher precision value than PHOG. Further research was carried out on larger datasets using combinations with other features. A study combining PHOG with Moore-neighbor tracing (MNT) was carried out to retrieve images of Islamic women's clothing features [18]. The combination of these two methods was termed PHOG cross-feature. The MNT method made sketches of shape features in the blouse-pant, long dress, and tunic categories with the highest accuracy exhibited in the blouse-pant category, followed by blouse-pant and tunic. The results showed that both sketch and image-based retrieval had low accuracy in the tunic category, which was difficult to distinguish boundaries due to the mixture of long dress and blouse-pant. Therefore, to improve accuracy, it needs to be combined with color and texture.

Dewi *et al.* [5] compared several color features, namely red, blue, yellow, green, and brown to retrieve images of Muslim women's fashion. The feature extraction methods used were scalable color descriptor (SCD) and dominant color descriptor (DCD). DCD elicited higher retrieval accuracy than SCD because its results were in line with human perceptions. Muslim fashion, which typically consists of several colors, also makes the DCD a more appropriate feature used in the Muslim clothing search system. Furthermore, the retrieval accuracy was measured using the *F*-measure. This study only compared the methods used against human perception, and due to the limited dataset, there was no treatment to test the robustness of the methods.

The combined color, shape, and texture features retrieved 200 images of fabric and 50 images of typical clothing fabric [4]. Feature extraction methods included red, green, blue (RGB) color histogram, LBP, bag of features (BoF), and joint equal contribution (JEQ)-based feature merge. The highest accuracy was achieved with complex-simple impressions using fusion JEQ, which homogenizes the extraction dimensions of the three methods. Furthermore, the support vector regression (SVR) method was used to map functions between generic impressions and image features, while similarity was measured using Pearson correlation coefficient (PCC) and Spearman's correlation coefficient. The developments of this research are applicable to natural image and painting datasets.

Research on the classification of batik images identified seven classes based on texture and shape features [19]. The research compared the individual and combinational use of features. The extraction method for texture and shape features was GLCM and Otsu methods, with artificial neural networks used for classification. The results showed that the combination of the two features had higher accuracy than the shape alone, while the lowest was achieved on the "mega mendung" batik image, because it contained more colors. Research on the retrieval of eight categories of batik images was conducted by [20] using color and texture feature extraction methods such as color auto-correlogram (HSV) and GLCM. Therefore, further research needs to be carried out to incorporate all three features to enhance retrieval and classification accuracy.

The research carried out by [21] designed an automated system to detect defective woolen fleece fabric by reducing the size of the texture features after the extraction process. Principal component analysis (PCA) is the most common and widely used technique for reducing dimensions in linear transformation. The center symmetric local binary pattern approach was used to extract texture characteristics, while *k*-nearest neighbors and Naive Bayes determined similarity. The experiments indicated that combined work has a significant amount of classification accuracy, while the Naive Bayes was ineffective on high-dimensional feature vectors. Classification accuracy decreased when testing multiple classes, therefore to improve it and define similar defects, the authors proposed using a binary-level feature extraction process.

Research carried out by [22] extracted texture and color features using the maximum run length method from LBP and the sine-cosine algorithm. The similarity was measured by applying optimum similarity weights (OSW) to the LBP. The application of OSW improved retrieval accuracy and low feature dimensions, while the combination of color and texture features increased precision. The color feature affected accuracy because the batik images contained many colors in addition to the texture pattern. A study carried out by [23] used CM and the perceptual hash algorithm to retrieve yarn-dried fabric images. Color and spatial features were extracted using both CM and the perceptual hash algorithm. Meanwhile, similarity matching was performed using the Euclidean distance and Hamming distance (HD). The results showed that combining the two extraction methods yielded higher accuracy and retrieval rates than state-of-the-art, although it required additional computation time.

Research was carried out by [24] to create a database of fabric features based on the color and shape of repeating dot patterns. In this study, the second-order wavelet transform (WT) method was used to reduce

the image size to 1/16. The binary image of each pattern was obtained by the Sobel method and the morphological operation with the Fabric images represented by a repeating color image pattern and template. The proposed method was suitable for monochrome or polychrome images with small and large repeating patterns of the dot-printed fabric. According to the author, the fabric database is applicable to image retrieval.

Research on image retrieval batik based on the statistical model of the magnitude filter Gabor was carried out by [25] using Gaussian and Rayleigh probability distribution. Both types of distribution function for feature extraction and perform well in textural image retrieval systems, particularly in grayscale formats. The result showed that the rich colour images can be used to improve the performance of the Gabor feature.

Research on retrieval of women's clothing by displaying pictures and captions was carried out by [26], with the images classified according to length and sleeves. The feature extraction method was Gabor filter and color histogram for texture and color, respectively. The histogram approach determined the length of the dress sleeves, which are categorized as short, medium, and long. Furthermore, the binarization process was carried out to determine the upper and lower limits of the image as well as to calculate the length of the arm from the limit using RGB, CIELAB, and HSV color formats. The pre-processing process used the Viola-jones algorithm to remove the head of the clothing image. The results showed that the Gabor and the CIELAB color space methods outperformed other methods. Future research needs to be conducted on developing other women's accessories such as jewelry and footwear.

Zhang *et al.* [27], research was conducted to retrieve lace fabric images with variants of texture pattern size, point of view, and texture density. The texture features were extracted using the Multi-scale and Rotation Invariant LBP (MRI-LBP) methods, while feature extraction was carried out at four different image size scales. Furthermore, local and global feature information was generated from small and large lace fabric image sizes, respectively. The result showed that the selection of image scale and sampling radius produced better retrieval outcomes, while the combination of multi-scale features provided higher retrieval precision than state-of-the-art. Future research needs to optimize feature extraction by automatically selecting the image scale and sampling radius.

The traditional descriptor approach is still widely used for research purposes. The strength of traditional descriptors such as SIFT, VALAD, LBP, GLCM, HSV, HOG, CM, GIST, and FV are due to their robustness to illumination and geometric changes. Generally, the most dominant color feature used affects retrieval performance for various datasets. For instance, some fabric, patterned, texture images play a critical role in enhancing retrieval accuracy. Color features are particularly sensitive to illumination, while texture and shape features are sensitive to geometrical shifts. The selection of conventional descriptor methods and features that are suitable for and tailored to the image form need should yield good results. The combination of color, texture, and shape features can improve retrieval accuracy, although it requires additional computation time.

4. FEATURE EXTRACTION BASED ON CNN METHODS

This section explains the feature extraction process using deep learning methods, such as CNNs, one of the most applied image retrieval methods. Table 2 shows the summary of feature extraction based on CNN methods. The results of the modified CNN were named compact-CNN [29], with an automatic assessment process used to smoothen the fabric appearance. Fabric images pre-processed as CNN input, and smoothness was labeled for supervised training. Furthermore, both processes effectively improved the performance of the model and illumination stability. The test dataset was limited to light and solid-colored fabrics and did not address multi-color fabric images or input variations. The proposed method needs to be tested for effectiveness and stability on multi-color images and large datasets. CNN feature extraction needs to be modified to produce more resistant features and improve retrieval performance [30]. Meanwhile, image features were obtained from a classic CNN and visual attention model (VAM). The VAM generates image attention maps and global features from classic CNN. It also handles background images to improve retrieval accuracy and is preferable in distinguishing clothing regions from the background when performing retrieval tasks.

Features from several CNN layers were combined with metric similarity in fabric image obtained from the fully connected (FC) layers [31]. Similarity matching was based on cosine similarity (CS) with an increase in retrieval performance using a combination of inactivated three layers of CNN features. Mean average precision (MAP) was enhanced by 5.27% on the fabric132 dataset, while the activation function process reduces information on image features. The proposed method is feasible for overcoming the challenge of a lack of high-resolution training data labeled in the same domain for the retrieval task.

According to [32] the focus ranking method can be integrated with CNN. Furthermore, the Alex-Net architecture is optimized by adding an auxiliary layer at the end of the FC to calculate the loss function and study the representation and metrics of images in fine-grained fabric image retrieval. The result showed that the focus ranking method shortens the computation time for matching query and database images. A large-scale fabric image retrieval dataset with approximately 25,000 images of 4,300 fabrics for training and testing was built. Fabric image variations were based on the camera angle, photograph conditions, and field angle. The

focus ranking method was optimized using stochastic gradient descent (SGD), with the dimensional features stabilizing at 256. The focus ranking method maintained better performance than others despite increasing the number of images in the dataset.

A research was carried out by [1] to determine the best combination of parameters for a multi-layer CNN (MCNN) with modeling and optimization method (MAOM) architecture. The result showed that MAOM architecture incorporates uniform experimental design (UED), multiple regression, and optimization. Meanwhile, UED is well suited for solving various factor problems at numerous levels. Furthermore, MAOM is a structural approach used to determine the best mix of parameters for an MCNN to classify images in average error rates. The optimization results demonstrated a better average error rate than a classical MCNN.

The CNN and hashing methods were combined in a research conducted by [2], which aimed to retrieve factory images from the WFID based on coarse-level and fine-level features. Hash coding was most successful at the coarse level, while the CNN extracted fine-level features from the fabric image was inputted through the stages of the equalization histogram process. Binary encoding was performed using the hamming distance (HD) with the dataset labeled and included in supervised training. One of the advantages of using these methods is that they reduced the computation time and feature dimensions. However, the weakness was that it produced a high error rate for images with many patterns (i.e., printed fabrics). Multi-task learning (MTL) combines a modified CNN with deep hash coding to retrieve fabric images [33]. MTL aims to improve the prediction accuracy and learning efficiency of each task in comparison with training a separate model. A single image typically extracts thousands or even millions of features with the high-dimensional ones directly used to retrieve and increase computational cost, thereby reducing retrieval efficiency. Furthermore, to achieve compact image representation, high-dimensional features are quantized to a fixed-length vector, called hashing code.

The dataset newly constructed by [34] consisted of 46,656 fabric images of 972 subjects, with multiple images from the front and back sides. The dataset was tested using BoF and three types of deep learning methods, namely LeNet, AlexNet, and VGGNet. Retrieval performance indicators included the cumulative match characteristic curve and MAP. The results showed that the dataset is also used to test other methods. The CNN made many errors in recognizing sophisticated textures and patterned fabrics. Tests showed that the texture features of the fabric image affected the retrieval performance.

The research conducted by [35] determined specific clothing features from a series of videos and a collection of online shopping images. The result showed that the combination of CNN-based feature extraction and color filters tends to improve retrieval accuracy. The color filter function removes the search candidates that have significant color differences with the query image. The method worked adequately during retrieval and when the clothing of interest was blocked by other objects, such as hands, hair, accessories, and people. Retrieval test results showed that the higher the top k , the better the system's performance. The research carried out by [36] combined global and local features to retrieve clothing images, using adaptive spatial attention unit (ASAU), multiple granularity network, spatial constraint network, and long short-term memory network techniques. Furthermore, the ASAU architecture was used to extract more local information. These methods were useful in handling images with background interference and local attention with increased computation time due to the parallel ASAU process.

Xiong *et al.* [37] compared the retrieval of the shop and street photos with triplet and Siamese network models based on sharing and non-sharing parameters. Partial-sharing parameters and the triplet model elicited the best results in the skirts category, thereby reducing the feature parameters by 27.5%. Cosine distance and HD were used for similarity matching. The result showed that the partial-sharing parameters scheme significantly reduced redundant parameters. The authors stated that the results of these studies can be used as a reference for conducting cross-domain experiments in retrieval clothing applications.

The study conducted by [38] added a loss function at the end of the fully connected layer to classify and retrieve fashion images, thereby increasing retrieval accuracy. This study used two types of loss, namely classification, and similarity. Both functions support the learning of feature representation among various fashion categories. The fully connected layer of each model included the adoption of two-loss paths and an additional relaxation module. The loss functions included object-based classification (OC), object category-based similarity (OS), and instance-based similarity (IS). The best accuracy was found when top k was set to 5 for the loss function combination DenseNet121 (OC + IS) on the shop image dataset.

The research combined low-level semantic information on critical points with the image's high-level features to retrieve clothing images [39]. A cascade function was used for CNN-based feature extraction, with the top 5 used to achieve better accuracy than state of the art. The CNN's accuracy was low when used to cloth complex shape changes, while the combination of global and local features improved retrieval performance. Future research focuses on improving the performance of the clothing identification algorithm caused by the body and other factors.

Robust retrieval modeling was designed for large-scale real and automatic data relationships with incomplete labeling used to achieve a strong retrieval representation. Global feature extraction was performed

using a CNN, and SGD was used to measure similarity [40]. Multi-weight testing effectively addresses issues related to incorrect and imbalanced data in large-scale, real-world clothing data. Multi-weight testing was conducted in 24 categories, with the best results found in the gender category. The authors suggested that future research needs to be conducted to modify multi-weight neural networks for cross-domain clothing image retrieval. The R-CNN method was used by [41] as pre-processing of input images, while the R-CNN results were used as an input to deep neural networks called StyleNet. The results showed that the input image initiated by the R-CNN process can enhance the performance of classical CNN.

The study carried out by [42] compared several schema crops images of cloth for retrieval testing using the CNN method. Softmax cross-entropy and centre loss architectures were combined to improve performance and time efficiency. The results showed that the sampling area of the original image had more influencing performance than the depth of the CNN network. These features do not hold up when there are blank areas or labels on the fabric image.

The research conducted in [43], using hierarchical hash coding on retrieval clothing, has the ability to speed up retrieval time. This research was conducted using Alex-net architecture for feature extraction with modifications made between the FC7, FC8, and latent layers. The retrieval clothing image results are faster than the base Alex-net architecture and used to test the similarity of images in the candidate pool at the fine level to shorten the search time.

CNN's extraction features globally require large amounts of data, therefore, training it on large datasets provides a sufficient knowledge base to identify objects. CNNs are more sensitive to color features, with the main advantage of automatically identifying essential features without human supervision [34]. CNNs have the ability to reduce the semantic gap in image representation [44]. Another benefit is that they allow for image retrieval on large datasets and use a deep layer to obtain more specific feature parameters, which are then used in similarity matching. In general, the features of fully connected CNN layers are used to match query and database images [2], [33], [45]. In several studies, the features extracted in the convolutional layer were used for image retrieval [31], [46]. One of the weaknesses of this method is found in image retrieval using complex textures [2]. The drawbacks of CNNs include memory complexity, high processing time, and large feature dimensions. Deep neural networks tend to quickly encounter overfitting on small image datasets, hence they are unable to analyze the data effectively.

Table 2. A summary of feature extraction based on CNN method

References	Dataset	Features/Extraction and Matching Method	Dim/t (s)	Metrics
Wang <i>et al.</i> [29]	Graded fabric	Texture/ CNN, Compact-CNN	-	P= 84.0%, 95.38%, 100% for error 0 ⁰ , 0.5 ⁰ , 1 ⁰ respectively
Wang <i>et al.</i> [30]	Fashionista Deep Fashion Street2 Shop	Global features/CNN+VAM/ED	-	P= 0.923
Cai <i>et al.</i> [31]	Fabric1000 Fabric-132	Global features/ CNN/CS	4096 & 1000	mAP increased by 5.27%
Deng <i>et al.</i> [32]	FIRD	Global features/ CNN/SGD	256	R= 93.9%, mAP= 0.75
Chou <i>et al.</i> [1]	Fashion-MNIST	Global features/ M-CNN, UED, MR	-	P= 96.53% (training), P= 91.68% (test)
Xiang <i>et al.</i> [2]	WFID	Global features/ DCNN+Hash/ HD, ED	1024/t=15.3	Rank k= 10, P= 98.56% error rate= 10.08
Xiang <i>et al.</i> [33]	WFID	Global features/CNN, DH/	128-bit	mAP= 0.947 at top= 20
Shen <i>et al.</i> [34]	Fabric	Global & texture features/CNN, BoF/ED	-	CMC= 91.14% at rank=1
Xiao & Yichao [35]	clothes	Global, color/CNN, color filter/CS	1024	Top k=20, R= 0.45
Luo <i>et al.</i> [36]	clothes	Global, local/ASAU, MGN & SCN, LSTM	-	Top k= 5, R= 0.5916
Xiong <i>et al.</i> [37]	clothes	Global/CNN/ ED, CS, HD	-	Triplet (Non-share), P= 0.736
Park <i>et al.</i> [38]	Deep Fashion	Global/CNN+ loss Function (OC, IS, OS)/ Top k accuracy	-	Top k=5 for DenseNet121 (OC+IS), P= 0.90
Wang <i>et al.</i> [39]	Deep Fashion	Global, local/CNN	-	Top 5, P= 91.83%
Li <i>et al.</i> [40]	e-Clothing.com, taobao.com, meilishuo.com	Global/ CNN/ SGD	-	mAP= 99.52% on gender categories
Yan <i>et al.</i> [41]	Fashionista	Local features/R-CNN/DCC	-	StyleNet, P = 84.3%
Zhu <i>et al.</i> [42]	fabric	Global/CNN	t= 0.03s	P= 99.89%
Lin <i>et al.</i> [43]	Clothing	Global/CNN/ HD	t= 0.6s	-

5. COMPARISON OF TRADITIONAL AND CNN FEATURE EXTRACTION

A survey of feature extraction using both traditional and CNN methods indicates that each approach has certain strengths and weaknesses. For instance, traditional descriptors have shorter computing time and

fewer system requirements. However, their retrieval performance decreases on large datasets, while CNNs handle large amounts of data. Furthermore, improving the quality of input images can improve retrieval performance. Pre-processing techniques like histogram equalization and ROI improve recognition of input image features. The methods selected for combination with a CNN need to be adjusted according to the dataset characteristics with the fabric image variations based on the camera angle, photograph conditions, and field angle. The accuracy of the method needs to be tested with data variations using geometric changes and image illumination.

Some of the fabric image characteristics described in the reviewed studies are. First, fabric images comprise many latticed and striped fabrics [2], [8]. Second, it contains complex textures that are more critical than color features for fabric image retrieval [34]. Third, it has repetitive patterns and creamy textures [9], [14]. Fourth, fabric images with indistinct embedded lines decreased the average precision [11]. Fifth, printed fabric images contain color, edge, contour, and contrast features [12]. Sixth, it includes various traditional patterns [13], [19]. Seventh, fabric images include both coarse-grained and fine-grained features [29], [32].

Feature selection and extraction based on fabric image characteristics is necessary to achieve optimal results. Future research developments are suggested to modify CNN methods, which comprises global features capable of integrating with local feature descriptors, such as images with intricate texture patterns [39]. Integrating CNNs and hash encoding aims to reduce feature dimensions and computation time [2], [33], [43], [47]. Furthermore, its features are strengthened by the color filter process to produce color features that are more resistant to illumination and background [35].

According to various literature reviews, CNN methods achieved the best image retrieval performance. The choice of extraction method specifies the retrieval precision according to the image pattern of the fabric. The combination of extraction methods reduces the semantic gap, computation time, and feature dimensions. The metrics comparison of the traditional and CNN methods is shown in Table 3.

Table 4 compares the characteristics of traditional and CNN feature extraction methods in the context of fabric image retrieval. The characteristics of the image are required to choose the appropriate feature extraction method for image retrieval. In future research, both methods should undergo more thorough testing using geometric image changes, illumination changes, and background interference. Geometric changes may include translation, scaling, flipping, and rotation. Light changes include differences in the amount and intensity of the light as well as variations in day versus night and indoor versus outdoor illumination. Background interference tests the method's ability to distinguish retrieval objects. Tests should also be carried out involving wrinkling fabric to capture pictures of material with different degrees of creasing. For example, hash coding can be used to reduce feature dimensions of CNN output and computation time [2], [33], [49]. Pruning neural networks can be done to reduce the computational load on the CNN method. Pruning is divided into two categories, namely weight and filter pruning [50]-[52]. A combination of techniques that can handle texture and color features is needed to improve accuracy for image types with complex textures.

Table 3. Comparison of traditional and convolutional neural network (CNN) methods

References	Extraction Method	Metrics
Chou <i>et al.</i> [1]	Matlab Example	$P_{\text{test}}=0.9080$
	MAOM MCNN	$P_{\text{test}}=0.9168$
Xiang <i>et al.</i> [2]	MS+LBP	Error rate=42.5%
	DSH	Error rate=16.8%
	DCNN+Hash	Error rate=10.1%
Wang <i>et al.</i> [29]	Fourier+SVM	$P=79.22\%$
	Inception v3	$P=80\%$
Deng <i>et al.</i> [32]	Compact CNN (pre-processing & label smoothing)	$P=84\%$
	Alexnet	MAP=50%
	Softmax	MAP=60%
	Triplet	MAP=64%
Shen <i>et al.</i> [34]	Alexnet+SGD	MAP=74%
	BoF	MAP=44.63%
	LeNet	MAP=48.57%
	AlexNet	MAP=50.42%
Xiao & Yichao [35]	VGGNet	MAP=51.82%
	CNN (GoogleNet)	$R=0.38$
	CNN+Color Filter (CF)	$R=0.43$
Wang <i>et al.</i> [39]	HOG+SVM	$P=70.35\%$
	AlexNet	$P=89.38\%$
	Fashionnet	$P=90.17\%$
	Modify CNN	$P=91.83\%$
Li <i>et al.</i> [48]	SURF	$P=0.8867$
	CNN	$P=0.9167$
	CNN-SURF	$P=0.9285$

Table 4. Characteristics of traditional descriptor and convolutional neural network (CNN) methods

Feature Extraction Method	Features	Advantages	Weaknesses
Traditional descriptor methods (e.g., LBP, DFT, HOG, SIFT, SURF, GLCM, Gabor wavelet, GIST, DWT, CM, SCD, DCD)	Color, texture, shape, spatial information	Combining features improves retrieval accuracy [7], [8], [10], [11], [13], [15], [19] Small feature dimensions [8], [15], [16], [21] Faster computation time [8], [9], [12], [15] Can handle background interference and scene rotation [14], [53]	Not suitable for complex fabric images [7], [9], [11] Use of a single feature has low retrieval accuracy [6], [8], [11], [12], [17], [18], [28]. Retrieval accuracy is lower than with CNNs
CNN method	Global features	Suitable for large datasets and both supervised and unsupervised learning [8], [10], [11], [13], [15], [16], [19] Reduces the semantic gap in comparison with traditional methods [8], [15], [16], [21] Can adapt to illumination and background interference [29], [36]	Not suitable for sketch or grayscale images [28] Low accuracy for images with many patterns, such as printed fabric or clothing with complex shapes [2], [34], [39] Large feature dimensions [48], [54] High computation time [31], [36], [41]

6. CONCLUSION

Feature selection and extraction methods are critical aspects of fabric image retrieval. Feature extraction using traditional methods produces low-level features, such as color, texture, shape, and spatial information. Meanwhile, CNNs elicit high-level features with the ability to overcome the semantic gap. However, both methods have various advantages and drawbacks. For instance, feature selection and extraction methods affect retrieval accuracy, feature dimensions, and computation time. In addition, the use of mixed methods for feature extraction tends to improve the retrieval performance. Traditional descriptors, which produce low-level features, overcome geometric and illumination changes, and reduce feature size. Furthermore, CNNs, which handle high-level features that are tailored to human perceptions, have huge feature dimensions, and require a great deal of computation time. CNNs produce global image features, and modifying it reduces feature size and computation time. Pruning neural networks can be done to reduce the computational load on the CNN method. Pruning is divided into two categories, namely weight and filter pruning. At the end of the CNN layer, hashing code is added to reduce search time. In general, the features of fully connected layers are used to match query and database images. Several studies have shown that the features extracted in the CNN’s convolutional layer were used for image retrieval.

APPENDIX

Table 1. A summary of feature extraction based on traditional methods

References	Dataset	Features/Extraction and Matching Method	Dim/t [s]	Metrics
Mutia <i>et al.</i> [6]	Islamic women apparel	Shape/HOG, PHOG, NPHOG/ED	189	Average P=5% above PHOG & P=3% above HOG
Lei <i>et al.</i> [28]	Amazon.com, tmall.com	Sketch/CRF, AED, CE/CD	-	P=93%, R=57%
Nurhaida <i>et al.</i> [7]	Batik	Texture/Gabor, GLCM, LBP/CD	-	P=84.54%
Zhang <i>et al.</i> [8]	WFID	Texture/DFT+LBP/PCC, ED	DFT+LBP=512/ t=1.4	P=84%, R=54%
Li <i>et al.</i> [9]	FID	Texture/TM, IG-HM/KLD	t<1	P=83%, R=92%
Tariq <i>et al.</i> [10]	INRIA, MITD	Color, texture/ HOG, HSV, LBP, SVM/TPR, FPPI	HSV=256 LBP=8192	TPR=65%, HSV>90%
Zhang <i>et al.</i> [11]	WFID	Color/ DC, CM/ED	t=0.2	P=87%, R=44%
Jing <i>et al.</i> [12]	Printed-FID	Color, SI/CM, GIST/ ED	t=9,051	P=86.3%, R=53.3%
Suciati <i>et al.</i> [13]	Fabric batik & common	Color, texture/HSV, Fractal/ CD	1x5/1x72	R>90%
Yao & Ke [14]	Fabric images	Texture, spatial information/ GIST, SURF/ ED	-	P>80%, rotate rate: ± 90%
Feng <i>et al.</i> [15]	Outex-00014 (include fabrics)	Color, shape/ PUH, MCH/CD	PUH=445, MCH=229, t=0.72	APR:84.21% ARR: 42.10% for Outex-00013
Bi <i>et al.</i> [16]	Textile	Color, texture, shape/CH, HSV, CAC, CM, GW, WM/ SVM	Total dim=190	Fabric type: P=0.62, R=0.67

Table 1. A summary of feature extraction based on traditional methods (continue)

References	Dataset	Features/Extraction and Matching Method	Dim/t [s]	Metrics
Aulia <i>et al.</i> [17]	Women clothing	Shape/HOG, ROI/ED	-	P increased to 16%
Nasita <i>et al.</i> [18]	Islamic women clothing	Shape/PHOG, PHOGXF/ED	-	P=90.4% (long-dress), P=52.1% (tunic)
Dewi <i>et al.</i> [5]	Muslim women's clothing	Color/SCD, DCD/ED	-	F-score=60% in blue categories
Chen <i>et al.</i> [4]	Fabric	Color, texture, shape/CH, LBP, BOF, JEQ/PR, SCC	-	P=0.89, on impression complex-simple type
Kasim <i>et al.</i> [19]	Batik	Texture, shape/GLCM, Otsu/ANN	-	P=90.48%
Wahyuningrum <i>et al.</i> [20]	Batik	Color, texture/ CAC, GLCM	-	P=87%
Yildiz[21]	Woolen fleece fabric	Texture/ CSLBP, PCA/KNN, Naive's Bayes	-	Accuracy= 0.9545 for kNN with correlation
Prasetyo & Simatupang [22]	Batik	Color, texture / modified LBP, SCA/OSW	LBP= 25	P=91.05%
Li <i>et al.</i> [23]	Yarn-Dyed Fabric	Color, spatial/CM, PHA/ED, HD	t= 1.563	AR=85.30%, RR=53.51%
Kuo <i>et al.</i> [24]	Printed fabric	Color, texture/WT, Sobel, Morphology	-	-
Prasetyo and Winarno [25]	Batik	Gabor/texture/Manhattan, ED, modified CD	144	P=96.55% (modified CD)
Gupta <i>et al.</i> [26]	Woman clothing	Texture, color/Gabor, CH/ Chroma difference	-	P=89.25%, R=87%
L. Zhang [27]	Lace Fabric	Texture/MRI-LBP/CD	324	MAP=81.6%

ACKNOWLEDGEMENTS

This work was supported by Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan Indonesia, ref. 201911212115638).

REFERENCES

- [1] F. I. Chou, Y. K. Tsai, Y. M. Chen, J. T. Tsai, and C. C. Kuo, "Optimizing Parameters of Multi-Layer Convolutional Neural Network by Modeling and Optimization Method," *IEEE Access*, vol. 7, pp. 68316-68330, 2019, doi: 10.1109/ACCESS.2019.2918563.
- [2] J. Xiang, N. Zhang, R. Pan, and W. Gao, "Fabric Image Retrieval System Using Hierarchical Search Based on Deep Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 35405-35417, 2019, doi: 10.1109/ACCESS.2019.2898906.
- [3] H. Prasetyo and B. A. Putra Akardihas, "Batik image retrieval using convolutional neural network," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, no. 6, pp. 3010-3018, 2019, doi: 10.12928/TELKOMNIKA.v17i6.12701.
- [4] Y. W. Chen, X. Huang, D. Chen, and X. H. Han, "Generic and Specific Impressions Estimation and Their Application to KANSEI-Based Clothing Fabric Image Retrieval," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 10, pp. 1-18, 2018, doi: 10.1142/S0218001418540241.
- [5] A. F. Dewi, F. Arnia, and R. Muharar, "Effectiveness of MPEG-7 color features in clothing retrieval," *Bull. Electr. Eng. Informatics (BEEI)*, vol. 6, no. 2, pp. 166-173, 2017, doi: 10.11591/eei.v6i2.619.
- [6] C. Mutia, F. Arnia, and R. Muharar, "Improving the performance of CBIR on islamic women apparels using normalized PHOG," *Bull. Electr. Eng. Informatics (BEEI)*, vol. 6, no. 3, pp. 271-280, 2017, doi: 10.11591/eei.v6i3.657.
- [7] I. Nurhaida, H. Wei, R. A. M. Zen, R. Manurung, and A. M. Arymurthy, "Texture fusion for batik motif retrieval system," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 6, no. 6, pp. 3174-3187, 2016, doi: 10.11591/ijece.v6i6.12049.
- [8] N. Zhang, J. Xiang, L. Wang, W. Gao, and R. Pan, "Image retrieval of wool fabric. Part I: Based on low-level texture features," *Text. Res. J.*, vol. 89, no. 19-20, pp. 4195-4207, 2019, doi: 10.1177/0040517519829003.
- [9] Y. Li, J. Zhang, M. Chen, H. Lei, G. Luo, and Y. Huang, "Shape based local affine invariant texture characteristics for fabric image retrieval," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15433-15453, 2019, doi: 10.1007/s11042-018-6936-y.
- [10] H. Tariq, S. Sheikh, U. A. Mohsin Ali, and A. Ali, "Domain Specific Content Based Image Retrieval (CBIR) for Feminine Textile Designs," *IJCSNS*, vol. 19, no. 2, p. 88, 2019.
- [11] N. Zhang, J. Xiang, L. Wang, N. Xiong, W. Gao, and R. Pan, "Image retrieval of wool fabric. Part II: based on low-level color features," *Text. Res. J.*, vol. 90, no. 7-8, pp. 797-808, 2020, doi: 10.1177/0040517519881819.
- [12] J. Jing, Q. Li, P. Li, and L. Zhang, "A new method of printed fabric image retrieval based on color moments and gist feature description," *Text. Res. J.*, vol. 86, no. 11, pp. 1137-1150, 2016, doi: 10.1177/0040517515606378.
- [13] N. Suciati, D. Herumurti, and A. Y. Wijaya, "Fractal-based texture and HSV color features for fabric image retrieval," *Proc. - 5th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2015*, no. November, 2016, pp. 178-182, doi: 10.1109/ICCSCE.2015.7482180.
- [14] L. Yao and H. Ke, "Robust image retrieval for lacy and embroidered fabric," *Text. Res. J.*, vol. 89, no. 13, pp. 2616-2625, 2019, doi: 10.1177/0040517518798648.
- [15] Q. Feng, Q. Hao, Y. Chen, Y. Yi, Y. Wei, and J. Dai, "Hybrid histogram descriptor: A fusion feature representation for image retrieval," *Sensors (Switzerland)*, vol. 18, no. 6, 2018, doi: 10.3390/s18061943.
- [16] Y. Bi, R. Bhatia, and S. Kapoor, "Intelligent Systems and Applications," *Proceedings of the 2019 Intelligent Systems Conference*, vol. 1, 2019.

- [17] N. Aulia, F. Arnia, and K. Munadi, "HOG of region of interest for improving clothing retrieval performance," *Proc. Cybern. 2019 - 2019 IEEE Int. Conf. Cybern. Comput. Intell. Towar. a Smart Human-Centered Cyber World*, 2019, pp. 7-12, doi: 10.1109/CYBERNETICSCOM.2019.8875636.
- [18] I. Nasita, K. Muchtar, K. Saddami, and F. Arnia, "Cross-domain clothing image retrieval based on hybrid shape feature," *Proc. Cybern. 2019 - 2019 IEEE Int. Conf. Cybern. Comput. Intell. Towar. a Smart Human-Centered Cyber World*, 2019, pp. 98-102, doi: 10.1109/CYBERNETICSCOM.2019.8875676.
- [19] A. A. Kasim, R. Wardoyo, and A. Harjoko, "Batik classification with artificial neural network based on texture-shape feature of main ornament," *Int. J. Intell. Syst. Appl.*, vol. 9, no. 6, pp. 55-65, 2017, doi: 10.5815/ijisa.2017.06.06.
- [20] R. T. Wahyuningrum and I. A. Siradjuddin, "An efficient batik image retrieval system based on color and texture features," *J. Theor. Appl. Inf. Technol.*, vol. 81, no. 2, pp. 349-354, 2015.
- [21] K. Yildiz, "Dimensionality reduction-based feature extraction and classification on fleece fabric images," *Signal, Image Video Process.*, vol. 11, no. 2, pp. 317-323, 2017, doi: 10.1007/s11760-016-0939-9.
- [22] H. Prasetyo and J. W. Simatupang, "Batik Image Retrieval Using Maximum Run Length LBP and Sine-Cosine Optimizer," *ICSECC 2019 - Int. Conf. Sustain. Eng. Creat. Comput. New Idea, New Innov. Proc.*, 2019, pp. 265-269, doi: 10.1109/ICSECC.2019.8907190.
- [23] Z. Li, J. Xiang, L. Wang, N. Zhang, R. Pan, and W. Gao, "Yarn-dyed fabric image retrieval using colour moments and the perceptual hash algorithm," *Fibres Text. East. Eur.*, vol. 137, no. 5, pp. 39-46, 2019, doi: 10.5604/01.3001.0013.2900.
- [24] C. F. J. Kuo, C. L. Lee, and C. Y. Shih, "Image database of printed fabric with repeating dot patterns part (I) - image archiving," *Text. Res. J.*, vol. 87, no. 17, pp. 2089-2105, 2017, doi: 10.1177/0040517516663160.
- [25] W. Heri Prasetyo and Winarno, "Statistical modeling of gabor filtered magnitude for batik image retrieval," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 2-4, pp. 85-89, 2018.
- [26] M. Gupta, C. Bhatnagar, and A. S. Jalal, "Clothing Image Retrieval Based on Multiple Features for Smarter Shopping," *Procedia Comput. Sci.*, vol. 125, 2018, pp. 143-148, doi: 10.1016/j.procs.2017.12.020.
- [27] L. Zhang, X. Liu, Z. Lu, F. Liu, and R. Hong, "Lace fabric image retrieval based on multi-scale and rotation invariant LBP," *ACM Int. Conf. Proceeding Ser.*, vol. 2015-Augus, 2015, pp. 257-261, doi: 10.1145/2808492.2808567.
- [28] H. Lei, Y. Yi, Y. Li, G. Luo, and M. Wang, "A new clothing image retrieval algorithm based on sketch component segmentation in mobile visual sensors," *Int. J. Distrib. Sens. Networks*, vol. 14, no. 11, 2018, doi: 10.1177/1550147718815627.
- [29] J. Wang *et al.*, "Automatic Assessment of Fabric Smoothness Appearance Based on a Compact Convolutional Neural Network with Label Smoothing," *IEEE Access*, vol. 8, pp. 26966-26974, 2020, doi: 10.1109/ACCESS.2020.2971506.
- [30] Z. Wang, Y. Gu, Y. Zhang, J. Zhou, and X. Gu, "Clothing retrieval with visual attention model," *2017 IEEE Vis. Commun. Image Process. VCIP 2017*, vol. 2018-Janua, pp. 1-4, 2018, doi: 10.1109/VCIP.2017.8305144.
- [31] Z. Cai, W. Gao, Z. Yu, J. Huang, and Z. Cai, "Feature extraction with triplet convolutional neural network for content-based image retrieval," *Proc. 2017 12th IEEE Conf. Ind. Electron. Appl. ICIEA 2017*, vol. 2018-Febru, 2018, pp. 337-342, doi: 10.1109/ICIEA.2017.8282867.
- [32] D. Deng, R. Wang, H. Wu, H. He, Q. Li, and X. Luo, "Learning deep similarity models with focus ranking for fabric image retrieval," *Image Vis. Comput.*, vol. 70, pp. 11-20, 2018, doi: 10.1016/j.imavis.2017.12.005.
- [33] J. Xiang, N. Zhang, R. Pan, and W. Gao, "Fabric Retrieval Based on Multi-Task Learning," *IEEE Trans. Image Process.*, vol. 30, no. 2, pp. 1570-1582, 2021, doi: 10.1109/TIP.2020.3043877.
- [34] F. Shen *et al.*, "A Large Benchmark for Fabric Image Retrieval," pp. 247-251, 2020, doi: 10.1109/icivc47709.2019.8981065.
- [35] L. Xiao and X. Yichao, "Exact clothing retrieval approach based on deep neural network," *Proc. 2016 IEEE Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2016*, no. 1, 2016, pp. 396-400, doi: 10.1109/ITNEC.2016.7560389.
- [36] Z. Luo, J. Yuan, J. Yang, and W. Wen, "Spatial Constraint Multiple Granularity Attention Network for Jiahui Yuan Jie Yang Wei Wen Samsung Research China-Beijing (SRC-B), Beijing, China Institute of Automation, Chinese Academy of Sciences," *2019 IEEE Int. Conf. Image Process.*, 2019, pp. 859-863, doi: 10.1109/ICIP.2019.8802938.
- [37] Y. Xiong, N. Liu, Z. Xu, and Y. Zhang, "A parameter partial-sharing CNN architecture for cross-domain clothing retrieval," *VCIP 2016 - 30th Anniv. Vis. Commun. Image Process.*, 2017, pp. 1-4, doi: 10.1109/VCIP.2016.7805463.
- [38] S. Park, M. Shin, S. Ham, S. Choe, and Y. Kang, "Study on fashion image retrieval methods for efficient fashion visual search," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, 2019, pp. 316-319, doi: 10.1109/CVPRW.2019.00042.
- [39] X. Wang, Z. Xie, and S. Hao, "Clothing identification based on fused key points," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1481, 2019, pp. 116-120, doi: 10.1145/3319921.3319949.
- [40] R. Li, F. Feng, I. Ahmad, and X. Wang, "Retrieving real world clothing images via multi-weight deep convolutional neural networks," *Cluster Comput.*, vol. 22, no. s3, pp. 7123-7134, 2019, doi: 10.1007/s10586-017-1052-8.
- [41] C. Yan, L. Zhou, and Y. Wan, "A Multi-Task Learning Model for Better Representation of Clothing Images," *IEEE Access*, vol. 7, pp. 34499-34507, 2019, doi: 10.1109/ACCESS.2019.2904536.
- [42] J. Zhu, Y. Yang, J. Cao, E. Chak, and F. Mei, "Artificial Intelligence on Fashion and Textiles," *AITA: International Conference on Artificial Intelligence on Textile and Apparel*, vol. 849, no. January, 2019, pp. 121-128, doi: 10.1007/978-3-319-99695-0.
- [43] K. Lin, H. F. Yang, K. H. Liu, J. H. Hsiao, and C. S. Chen, "Rapid clothing retrieval via deep learning of binary codes and hierarchical search," *ICMR 2015 - Proc. 2015 ACM Int. Conf. Multimed. Retr.*, 2015, pp. 499-502, doi: 10.1145/2671188.2749318.

- [44] X. Li, J. Yang, and J. Ma, "Recent developments of content-based image retrieval (CBIR)," *Neurocomputing*, vol. 452, pp. 675-689, 2021, doi: 10.1016/j.neucom.2020.07.139.
- [45] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8689 LNCS, no. Part 1, pp. 584-599, 2014, doi: 10.1007/978-3-319-10590-1_38.
- [46] A. Jose, R. D. Lopez, I. Heisterklaus, and M. Wien, "Pyramid Pooling of Convolutional Feature Maps for Image Retrieval," *Proc. - Int. Conf. Image Process. ICIP*, 2018, pp. 480-484, doi: 10.1109/ICIP.2018.8451361.
- [47] J. Sun, X. Ding, L. Du, Q. Li, and F. Zou, "Research progress of fabric image feature extraction and retrieval based on convolutional neural network," *Fangzhi Xuebao/Journal Text. Res.*, vol. 40, no. 12, 2019, pp. 146-151, doi: 10.13475/j.fzxb.20190105306.
- [48] X. Li, J. Yang, and J. Ma, "Large Scale Category-Structured Image Retrieval for Object Identification through Supervised Learning of CNN and SURF-Based Matching," *IEEE Access*, vol. 8, pp. 57796-57809, 2020, doi: 10.1109/ACCESS.2020.2982560.
- [49] L. Zhou, Z. Zhou, and L. Z. B, "Deep Part-Based Image Feature for Clothing Retrieval," *International Conference on Neural Information Processing ICONIP 2017: Neural Information Processing*, pp. 340-347, 2017, doi: 10.1007/978-3-319-70090-8_35.
- [50] J. H. Luo, H. Zhang, H. Y. Zhou, C. W. Xie, J. Wu, and W. Lin, "ThiNet: Pruning CNN Filters for a Thinner Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2525-2538, 2019, doi: 10.1109/TPAMI.2018.2858232.
- [51] J. H. Luo, J. Wu, and W. Lin, "ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, 2017, pp. 5068-5076, doi: 10.1109/ICCV.2017.541.
- [52] C. Sándor, S. Pável, and L. Csató, "Pruning cnns with linear filter ensembles," *Frontiers in Artificial Intelligence and Applications*, pp. 1435-1442, 2020, doi: 10.3233/FAIA200249.
- [53] I. C. Amitha and N. K. Narayanan, "Object Retrieval in Images using SIFT and R-CNN," *2020 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 2020*, 2020, doi: 10.1109/ICITIIT49094.2020.9071557.
- [54] A. O. Tarasenko, Y. V. Yakimov, and V. N. Soloviev, "Convolutional neural networks for image classification," *CEUR Workshop Proc.*, vol. 2546, 2019, pp. 101-114.

BIOGRAPHIES OF AUTHORS



Silvester Tena received B.Eng. degree from the Department of Electrical Engineering, Universitas Nusa Cendana, Indonesia, in 2000. He obtained a Master of Engineering in electrical engineering from Universitas Gadjah Mada (UGM), Indonesia, in 2005. Currently, as a PhD candidate at the Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada. He started to join as teaching staff in the Department of Electrical Engineering, Faculty Science and Engineering, Universitas Nusa Cendana in 2000. His current research interests include Image Processing, Image Retrieval and Machine Learning.



Rudy Hartanto received a bachelor's degree (Ir.) in electrical engineering from Universitas Gadjah Mada (UGM), Indonesia, in 1989. He received a Master of Engineering degree (M.T.) in electrical engineering from Universitas Gadjah Mada (UGM), Indonesia, in 1998. In 2015, he received his Doctoral degree in Information Technology from Universitas Gadjah Mada (UGM), Indonesia. He started to join as teaching staff in the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM), in 1989. Currently, he is an Associate Professor and a Head of the Master of Information Technology Program in the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM). His current research interests include Human- Computer Interaction, Computer Vision, Multimedia, and Information Systems.



Igi Ardiyanto received B. Eng. degree from the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM), Indonesia. M. Eng. and Dr. Eng. degrees from the Department of Computer Science and Engineering, Toyohashi University of Technology (TUT), Japan. He is an Associate Professor at the Department of Electrical and Information Technology, UGM, Indonesia. He is interested in Robotics, Intelligent Vehicles, Embedded Systems, Computer Vision, and Parallel Computing.