IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 9, 2022, accepted March 15, 2022. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.3161510

Imbalanced Classification via Feature Dictionary-Based Minority Oversampling

MINHO PARK, HWA JEON SONG, AND DONG-OH KANG
Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

Corresponding author: Minho Park (roger618@etri.re.kr)

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [22ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System].

ABSTRACT Image classification research is one of the fields continuously studied in the computer vision domain, and several related studies have been actively conducted until recently. However, a limit exists regarding the prediction performance of real-world datasets due to the data imbalance problem between classes. Data augmentation through artificial sample generation for minority classes is one of the methods used to overcome this limitation. Among the various oversampling methods, we propose the feature dictionary-based generative model for the oversampling method. Feature dictionaries are built through the pretrained feature extractor, and the proposed generative model synthesizes artificial samples based on the dictionary. Class-to-class balanced training can be conducted by fine-tuning the classifier as additional data for the minority class. We experiment by applying the proposed framework to the fashion dataset, which has an extreme class imbalance. The experimental results demonstrate that the proposed model achieved the highest top-1 performance on various public fashion datasets. In addition, we analyze the number of samples in the dictionary and test the effectiveness of the elements that comprise the proposed model using various ablation studies.

INDEX TERMS Deep learning, imbalanced classification, generative adversarial network.

I. INTRODUCTION

Image classification is one of the fields that has long been studied in the computer vision domain [1]–[3]. With the recent development of deep learning and the availability of various and large-scale datasets, image classification performance has dramatically increased. However, the real-world datasets have a disadvantage of a large class imbalance due to labeling and time costs. Even public datasets, which are universally used in many studies [4], [5] suffer from data imbalances between classes. Class imbalances are especially severe in datasets used for specific domains. For example, in the case of fashion datasets, the difference in the number of data between the categories of clothes that people primarily wear and those that they do not is very significant. In extreme cases, the ratio of the number of images between the majority and minority classes exceeds 3000 times [6]. This is one of the most significant reasons for failing to improve image classification performance [7]. Because of the class imbalance, the prediction model is easily biased by majority classes. The biased model is optimized to favor the majority classes and fails to learn the fine discriminant features for the minority classes. Therefore, the prediction model has poor classification performance for minor classes than major classes [8]–[12]. Due to the limitations of image classification models with long-tailed datasets, quality results are not achieved in several applications in the various industries based on biased prediction models.

Studies have been conducted on imbalanced classification to overcome the limitations. There are two main methods: reweighting and resampling. Reweighting is a method of adjusting the weight of the objective function by class so that it is inversely proportional to the class frequency. Resampling is a method to balance the amount of data between classes by adjusting the number of data, and it is divided into ''undersampling,'' which reduces the majority class, or ''oversampling,'' which increases the minority class. Recently, oversampling has gained popularity due to the disadvantage of data loss caused by intentionally removing majority class data for data balancing in undersampling. The use of generative models to generate artificial samples for a minority class is one of several oversampling methods. Generative

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Cusano.

models that synthesize minority data based on a generative adversarial network (GAN) have been successfully applied to imbalanced classification [13], [14]. Mullick *et al*. [13] proposed an adversarial oversampling framework with a convex generator, discriminator, and classifier. The proposed generator synthesizes fake minority features by combining real features within the convex hull of the minor class. Kim *et al*. [14] transmitted and used the diversity of majority information to enrich minority samples. The proposed model translates the majority samples to the target minority class using a classifier trained independently on the given imbalanced dataset. However, the existing methods are used after they have been rendered unbalanced by artificially adjusting the number of data in a dataset where the imbalance between classes is insignificant. In addition, the resolution of the images in the dataset is small and uniform. Therefore, the existing methods have limitations in applying it to real-world datasets with a large degree of imbalance and nonuniform image ratios and resolutions. In this paper, we propose a deep network-based framework for resolving imbalance of real-world data. We verify the proposed model using the fashion dataset, which is one of the datasets with extreme data imbalance and large image size.

Recently, a method using feature extraction based on a convolutional neural network (CNN) has been primarily used for fashion category classification. In particular, previous studies focused on extracting structural features based on the key points of the clothing image. Based on the landmark detection of the clothes, category classification performance is improved using attention maps or shape-based features that focus on important points of the clothes. Wang *et al*. [15] proposed a fashion landmark-aware attention mechanism to force the deep model to focus on the functional parts of clothing images. Then, it learns representations centered on domain knowledge. Zhang *et al*. [16] suggested a two-stream clothing classification network, which is the extraction of the texture-stream and shape-stream features from clothing images. The performance was greatly improved by jointly learning texture-biased streams focused on clothing texture information and shape-biased streams focused on landmark information. However, this method did not resolve the issue of the label long-tail distribution in fashion image datasets. Recently, studies on generating raw fashion images using generative models to supplement the data for minority classes have been lacking in long-tailed datasets [17], [18]. The methods in [17] and [18] proposed networks that change poses while maintaining the appearance of a fashion image by exploiting GANs [19]. However, fashion images synthesized using these methods retain their shape and pose relatively naturally, but details such as the texture and color of clothing disappear or are transformed. In addition, since the memory cost is very high when training or using the image generative model, the augmentation of the training data itself takes a lot of time. Therefore, employing generated clothing images for the fashion category classification task is challenging.

By complementing existing methods with shortcomings, we propose an oversampling framework through the generative model to solve imbalances between major and minor classes of clothing data. The generative model assists training of classifier for minority classes by synthesizing artificial data through adversarial training. The proposed model devises feature dictionaries containing information on the shape and texture of clothing. Further, we propose the feature generation framework to avoid the loss of detail in clothes when directly generating raw images. The overall framework is trained in two steps. The first step is training feature extractors and classifiers with a real-world clothing dataset. The extractors obtain features, such as texture and shape from the clothes. Then we construct feature dictionaries for minority classes through the trained feature extractor. The second step fine-tunes the classifier through a minimax game algorithm with a generator and discriminator. Artificial samples are synthesized through convex combinations of convex weights generated by the generator and features sampled from the feature dictionary. The samples mimic the distribution of real features through adversarial training with the discriminator and classifier. The generator synthesizes fake data based on feature dictionaries through adversarial learning, and the classifier is fine-tuned with the generated feature. The classifier can be trained in a balanced manner for entire classes by obtaining additional data for the minority class through the generator.

We conduct various experiments to evaluate the proposed method using several public datasets, including DeepFashion [6] and DeepFashion2 [20], which have numerous data but a significant inter-class imbalance. The experimental results reveal that the proposed network outperforms the state-of-the-art methods. In particular, the classification performance for minority classes was much higher than that of the existing model. We also performed ablation studies regarding the usefulness of the minority oversampling via adversarial learning. Last, the change in performance was analyzed for each loss depending on the number of features sampled from the dictionary. The main contributions of this paper are summarized as follows:

- We propose a novel deep network for imbalanced classification that employs minority oversampling with adversarial learning. We also propose the construction of a feature dictionary for clothing and feature generation based on the dictionaries.
- We propose a two-stage training framework. First, the feature extractor and classifier are pretrained. Second, artificial minority samples are synthesized to assist the classifier in fine-tuning.
- We construct comprehensive evaluations for public fashion datasets and confirm that the proposed model outperforms the other methods. Specifically, we confirm that the proposed model works more effectively for minority classes. We also demonstrate the effectiveness of each component of the proposed model and analyze the number of samples from the feature dictionary.

The remainder of the paper is organized as follows. We introduce work related to this research in Section II. In Section III, we explain the proposed deep network for imbalanced classification, consisting of two phases of learning. In Section IV, we quantitatively compare the state-of-the-art model and proposed method and present several ablation studies. Finally, Section V presents the conclusions.

II. RELATED WORK

A. DEEP IMBALANCED CLASSIFICATION

With the recent development of deep learning, the importance of large-scale datasets is increasing. However, datasets are likely to have long-tailed label distribution because of high-priced data acquisition processes and labeling costs in the real world. Such class-imbalance problems cause classification performance degradation, especially for minority class labels [21]. Studies on imbalanced classification have continued in order to address this problem. There are two representative approaches for bypassing class-imbalance problems: *re-weighting* and *re-sampling*.

For *re-weighting*, the losses are reweighted at the category class level by multiplying the weight inversely proportional to the data distribution to extend the influence of minority class training samples. Artificially increasing the proportion of loss to the minority class leads to balanced learning between major and minor classes. Huang *et al.* [7] proposed quintuplet sampling methods and an associated triple-header loss that maintains cluster-wide locality and discrimination between classes. In work by Khan *et al.* [22], the cost-sensitive deep network was proposed, which can automatically extract robust feature representation for the minority class. For *re-sampling*, the given imbalanced dataset is balanced during training either by undersampling the majority classes or oversampling the minority classes. Undersampling the majority class is relatively simple, but the drawback is that the overall performance suffers significantly as the number of data available for learning is artificially reduced [21]. For this reason, oversampling, a method of artificially generating data for the minority and balancing classes, is widely employed [13], [14], [23]. Mullick *et al.* [13] proposed an oversampling scheme using the generative model with adversarial learning manner was proposed. A three-player adversarial game was employed with a generator, a discriminator, and a classifier. In this paper, inspired by this framework, we proposed an imbalanced classification network for clothing images that applies the oversampling framework.

B. FASHION RECOGNITION

In earlier studies, fashion recognition has typically employed traditional image analysis methods based on handcrafted features, such as SIFT [24], HOG [25]. However, these methods limit the extraction of useful features for fashion image analysis, leading to performance degradation. In recent studies, fashion image recognition has performed well due to the ease of use of richly annotated, large-scale clothing datasets [6], [20], [26], [27] and the advances in deep learning models based on convolutional neural networks (CNNs). Todays, deep CNN-based backbone networks (e.g. VGG [1], ResNet [3], and GoogLeNet [2]) trained with the large-scale datasets are primarily used for feature extraction for fashion image recognition. Bhatnagar *et al.* [28] demonstrated that, in fashion classification, the CNN outperforms traditional methods(e.g., the support vector classifier).

Several studies have found that training the classification model while detecting clothing landmarks helps predict clothing categories. As demonstrated in FashionNet [6], the deep model jointly predicts clothing category attributes and landmarks. The learned features from landmark detection are employed to train clothing category prediction. In TS-FashionNet [16], texture and shape biased deep networks trained jointly are proposed. The authors demonstrated that jointly using shape and texture features is necessary for fashion tasks. In this paper, we construct a baseline model centered on these insights.

III. PROPOSED METHOD

A. OVERVIEW

An overview of the proposed imbalanced classification framework for the long-tailed fashion dataset is presented in Fig. 1. The proposed network consists of four types of modules, two feature extractors, two classifiers, a generator, and a discriminator. Furthermore, they were trained in two phases (the upper and lower parts of Fig. 1 illustrate Phases 1 and 2, respectively). In the first phase, two feature extractors and two classifiers were trained with a long-tailed fashion dataset. Then, feature dictionaries were constructed through the trained extractors. In the second phase, while parameters of the feature extractors were fixed, one of the classifiers, the main classifier, was fine-tuned with artificially synthesized minority data. The generator obtained the weights for the weighted summation of features sampled from the dictionaries and generated minority samples. Through a minimax game algorithm between the generator, classifier, and discriminator, the synthesized samples become hard samples close to the distribution of real features.

Zhang *et al.* [16] exhibited the usefulness of fashion landmark detection for shape feature extraction and the effectiveness of joint learning using shape and texture features from fashion images. Therefore, we propose two streams of feature extractors: texture-biased and shape-biased feature extractors. A landmark detection module (LDM) helps the shape-stream extractor extract shape features in the shape stream. Additionally, one of the classifiers, the auxiliary classifier, helps the shape-stream features contain fashion category label knowledge. The shape features are concatenated with texture features to predict clothing categories through the main classifier. Because of the long-tailed problem of the fashion dataset, we fine-tuned the main classifier with artificially generated minority samples. To synthesize the samples,

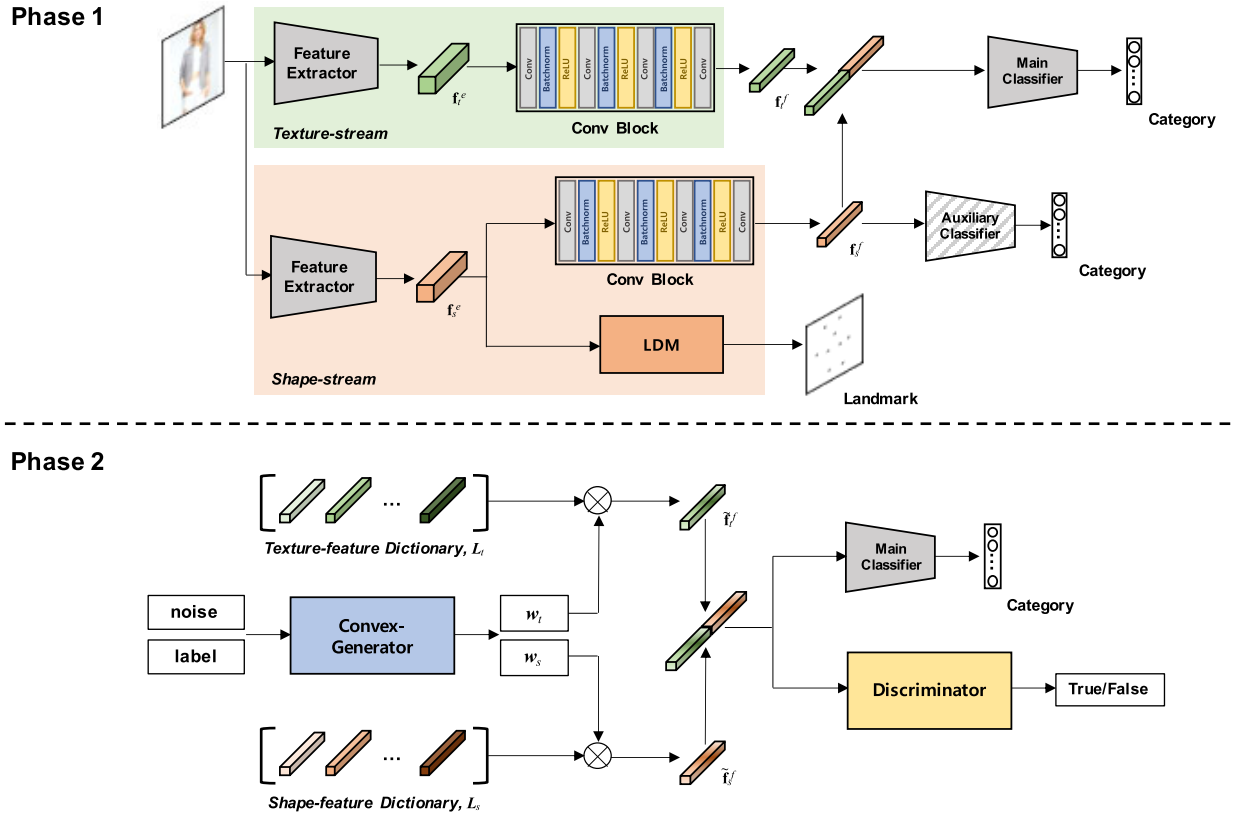


FIGURE 1. The overall architecture of the proposed network. The upper and lower parts show the network trained in the first and second phases, respectively. In Phase 1, the proposed network consists of feature extractors, convolutional blocks (Conv Block), landmark detection module (LDM), auxiliary classifier, and main classifier. A clothing image is fed to texture-stream and shape-stream feature extractors. Extracted features of shape-stream are employed to train the auxiliary classifier and get a landmark map through LDM. Also, the feature is concatenated with the texture-biased feature is eventually employed to train the main classifier. In Phase 2, artificial samples for the minority class are generated using features obtained in Phase 1. The generated features are subjected to adversarial training among classifier, generator, and discriminator to mimic the data distribution of real minority data.

we constructed feature dictionaries with the output features of trained texture-stream and shape-stream feature extractors (the same as the input of the main classifier). Then, the artificial sample was synthesized using a convex combination of features from the dictionaries and convex weights from the generator. Using this method, we augmented minority samples to be inversely proportional to the actual number of data. The details for each phase are described in the following subsections.

B. PHASE 1: CONSTRUCTING FEATURE DICTIONARIES

In Phase 1, the proposed feature extractors, auxiliary classifier, and main classifier were pre-trained with a long-tailed fashion dataset. In addition, F_t and F_s denote the texture-stream and shape-stream feature extractors, respectively. The obtained features from F_t and F_s are $\mathbf{f}_t^e$ and $\mathbf{f}_s^e$, respectively, and f_t and f_s denote the convolutional layer block of the texture and shape streams, respectively. In the texture-stream, $\mathbf{f}_t^e$ is flattened by f_t so that the flattened feature of texture-stream, $\mathbf{f}_t^f$ is obtained. In the shape-stream, $\mathbf{f}_s^f$ is acquired through f_s via the same method as the texture stream. In addition, LDM, which predicts landmark positions of clothing from $\mathbf{f}_s^e$. $\mathbf{f}_s^f$ is employed to train the auxiliary classifier. Moreover, $\mathbf{f}_s^f$ is concatenated with $\mathbf{f}_t^f$, and the concatenated features are employed to train the main classifier. Further, $\mathbf{f}_s^f$ represents the shape of the clothing image associated with the fashion categories via auxiliary classifier and LDM. In addition, category prediction can be more effective for the main classifier training by employing the concatenated feature, $[\mathbf{f}_t^f; \mathbf{f}_s^f]$, containing both the texture and shape presentations of the clothing image.

Based on previous works [6], [16], we used VGG16 [1] as the structure of F_t and F_s. Furthermore, we used the output of the layer `conv5_3` as $\mathbf{f}_t^e$ and $\mathbf{f}_s^e$. We used an ImageNet pre-trained model for both F_t and F_s to make the model efficiently converge. The structure of LDM is based on the landmark processing module presented by Zhang *et al.* [16]. The outputs of the LDM produce a landmark heatmap of $28 \times 28 \times K$, with landmark visibiliy and a vector of length K. We generated the ground-truth of the landmark heatmap by adding a Gaussian filter at the corresponding ground-truth landmark position. In addition, f_t and f_s consist of convolutional layers, batch normalization, and activation functions. The details of the structure of f_t and f_s are depicted in the convolutional layer block in Fig. 1.

We used the cross-entropy loss as an objective function of the auxiliary and main classifiers for training to predict clothing categories. In addition, we employed a weighted squared error to learn the landmark location and cross-entropy loss to predict the visibility of each landmark. In the case of clothing images, because the visibility of landmarks varies depending on the pose, only landmarks visible through weighted summation are included in the loss. The visibility weighted reconstruction loss for landmark prediction can be represented as follows:

$$\mathcal{L}_{land} = \sum_{n=1}^{N_l} v_n \sum_{i,j} \|\mathbf{M}_{n,i,j} - \hat{\mathbf{M}}_{n,i,j}\|_2^2, \tag{1}$$

where N_l and v are the number of landmarks and visibility, respectively, and $\mathbf{M}$ and $\hat{\mathbf{M}}$ denote ground-truth and predicted landmark heatmaps, respectively.

For the end of training in Phase 1, we fixed the weights of F_t, F_s, f_t and f_s. Then, texture and shape feature dictionaries, L_t and L_s were constructed with the flattened features, and $\mathbf{f}_t^f$ and $\mathbf{f}_s^f$ were obtained for the entire images of the fashion dataset. The feature dictionaries for the c-class can be represented by the following:

$$\begin{aligned} L_{t,c} &= [\mathbf{f}_{t,c}^{f,1}; \mathbf{f}_{t,c}^{f,2}; \cdots; \mathbf{f}_{t,c}^{f,N_c}], \\ L_{s,c} &= [\mathbf{f}_{s,c}^{f,1}; \mathbf{f}_{s,c}^{f,2}; \cdots; \mathbf{f}_{s,c}^{f,N_c}], \end{aligned} \tag{2}$$

where N_c is the number of c-class samples.

C. PHASE 2: MINORITY OVERSAMPLING

In Phase 2, the main classifier is fine-tuned with dictionary-based artificial minority samples. Inspired by the concept of minority oversampling via generative adversarial learning by Mullick *et al.* [13], we propose the convex generator and discriminator. The generator, G takes noise, $\mathbf{z} \sim P_N$ and labels $c \in C_m$ as inputs; P_N is the standard normal distribution and C_m is the class set except for one majority class with the largest number of data. The G generates the convex weights, w_t and w_s for the convex combination of data points from L_t and L_s. Generated samples, $\tilde{\mathbf{f}}_t^f$ and $\tilde{\mathbf{f}}_s^f$ increasingly mimic the data distribution of the minority samples through adversarial learning with the classifier and discriminator, D. In addition, G and D consist of three fully connected layers. The architecture details are listed in Table 1. Furthermore, FC and N_C indicate fully-connected layers and the number of category labels, respectively. Additionally, N_s is a hyper-parameter for the number to sample from the dictionary. The generated samples, $\tilde{\mathbf{f}}_t^f$ and $\tilde{\mathbf{f}}_s^f$ can be expressed as follows:

$$\begin{aligned} \tilde{\mathbf{f}}_{t,c}^f &= G(\mathbf{z}, c)_t \cdot \mathbf{f}_{t,c}^f = \sum_{i=1}^{N_s} w(\mathbf{z}, c)_t^i \mathbf{f}_{t,c}^{f,i}, \\ \tilde{\mathbf{f}}_{s,c}^f &= G(\mathbf{z}, c)_s \cdot \mathbf{f}_{s,c}^f = \sum_{i=1}^{N_s} w(\mathbf{z}, c)_s^i \mathbf{f}_{s,c}^{f,i}, \end{aligned} \tag{3}$$

where $\sum w(\mathbf{z}, c)_t = w_t$ and $\sum w(\mathbf{z}, c)_s = w_s$. We set N_s to an integer value smaller than the minimum value of the number of data per class. For every training iteration, the N_s features for each class are randomly sampled from the dictionary.

TABLE 1. The architecture of the proposed generator and discriminator. N_C indicates the number of category labels.

Module	Layer	Filter / Stride	Output Size
Generator	FC1	multiplication : $(256 + N_C)$ * 512	512
	FC2	multiplication : 512 * 1024	1024
	FC3	multiplication : 1024 * $(4096 \times N_s)$	$4096 \times N_s$
Discriminator	FC1	multiplication : $(2048 + N_C)$ * 1024	1024
	FC2	multiplication : 1024 * 512	512
	FC3	multiplication : 512 * 1	1

When training the generator, the label c entered as input is selected with a probability that is inversely proportional to the number of data in class c compared to the total data. The smaller the number of actual data, the greater the number of generated artificial data. The procedure details for generating the class were presented by Mullick *et al.* [13]. We adopted the GAN loss to optimize the generator, discriminator, and classifier to generate realistic artificial points. We used the least-squares formulation of the GAN loss to prevent the gradient vanishing problem [13], [29]. With a minimax game between three modules, the generator G synthesizes artificial data, deceiving the discriminator D and making classifier Q have difficulty predicting labels. The objective function of G is defined as follows:

$$\begin{aligned} \mathcal{L}_G = \; & \mathbb{E}_{\tilde{\mathbf{f}}_{t,c}^f, \tilde{\mathbf{f}}_{s,c}^f \sim p_c^g}[(Q_c([\tilde{\mathbf{f}}_{t,c}^f; \tilde{\mathbf{f}}_{s,c}^f]))^2] \\ & + \sum_{j \in C_m \setminus \{c\}} \mathbb{E}_{\tilde{\mathbf{f}}_{t,j}^f, \tilde{\mathbf{f}}_{s,j}^f \sim p_c^g}[(1 - Q_j([\tilde{\mathbf{f}}_{t,j}^f; \tilde{\mathbf{f}}_{s,j}^f]))^2] \\ & + \lambda \, \mathbb{E}_{\tilde{\mathbf{f}}_{t,c}^f, \tilde{\mathbf{f}}_{s,c}^f \sim p_c^g}[(1 - D([\tilde{\mathbf{f}}_{t,c}^f; \tilde{\mathbf{f}}_{s,c}^f]))^2], \end{aligned} \tag{4}$$

where Q_c denotes the c-th output value of classifier Q, and λ is a hyperparameter for adjusting the relative weights. The loss function of D is defined as follows:

$$\begin{aligned} \mathcal{L}_D = \; & \mathbb{E}_{\mathbf{f}_{t,c}^f, \mathbf{f}_{s,c}^f \sim p_c^d}[(1 - D([\mathbf{f}_{t,c}^f; \mathbf{f}_{s,c}^f]))^2] \\ & + \mathbb{E}_{\tilde{\mathbf{f}}_{t,c}^f, \tilde{\mathbf{f}}_{s,c}^f \sim p_c^g}[(D([\tilde{\mathbf{f}}_{t,c}^f; \tilde{\mathbf{f}}_{s,c}^f]))^2], \end{aligned} \tag{5}$$

where p_c^d and p_c^g are the conditional probability distribution of the c-th real and generated classes, respectively. We trained G and D using the adversarial loss functions and fine-tune Q simultaneously, and we updated the parameters of Q using the mean squared error (MSE) as a baseline. To solve the imbalanced problem in clothing data, we trained Q using re-weighting loss(e.g., Focal [30] and LDAM [31]) and the re-balancing of data through minority oversampling.

The details of the algorithm of the training framework in Phase 2 are described in Algorithm 1. The class was selected with a probability proportional to $(N_{d,c_m} - N_{d,c})$ at every iteration to generate artificial data in order to generate more samples as the number of data of the corresponding class is small. In addition, $N_{d,c}$ and c_m indicate the number of data of class c and the class with the most significant number of data, respectively. Moreover, $\|$ refers to concatenation.

Algorithm 1 Minority Oversampling With Adversarial Learning in Phase 2

Require: feature dictionaries: L_t and L_s, pre-trained classifier: Q
Require: the number of sample: N_s, the number of data of each classes: $N_{d,1}, N_{d,2}, \ldots, N_{d,C}$
Note: the class set: $\mathcal{C}$, the class with the largest number of data in $\mathcal{C}$: c_m
Note: the one hot vector of class c: Y_c, the ones' complement of Y_c: $\overline{Y_c}$

while not converged **do**
  for w steps **do**

    1. Update Q and D
    Select $c_r \in \mathcal{C}$ with equal probability.
    Sample $B_{t,c_r} = \{f_{t,c_r}^{f,1}, f_{t,c_r}^{f,2}, \ldots, f_{t,c_r}^{f,N_s}\}$ and $B_{s,c_r} = \{f_{s,c_r}^{f,1}, f_{s,c_r}^{f,2}, \ldots, f_{s,c_r}^{f,N_s}\}$ from L_t and L_s, with class label c_r.
    Update Q and D by respective gradient descent on $(Q(f_{t,c_r}^{f,i}||f_{s,c_r}^{f,j}), Y_{c_r})$ and $(D((f_{t,c_r}^{f,i}||f_{s,c_r}^{f,j})|Y_{c_r}), 1)$ with random index i and $j \in [1, N_s]$.
    Generate $w(\mathbf{z}, c_r)_t$ and $w(\mathbf{z}, c_r)_s$ with noise $\mathbf{z}$ which has standard normal distribution.
    Generate $\tilde{\mathbf{f}}_{t,c_r}^{f}$ and $\tilde{\mathbf{f}}_{s,c_r}^{f}$ with Eq.(3)
    Update Q and D by respective gradient descent on $(Q(\tilde{\mathbf{f}}_{t,c_r}^{f}||\tilde{\mathbf{f}}_{s,c_r}^{f}), Y_{c_r})$ and $(D((\tilde{\mathbf{f}}_{t,c_r}^{f}||\tilde{\mathbf{f}}_{s,c_r}^{f})|Y_{c_r}), 0)$, keeping G fixed.

    2. Update G
    Select c_g with probability $\propto (N_{d,c_m} - N_{d,c_g})\forall i \in \mathcal{C}\backslash\{c_g\}$.
    Sample $B_{t,c_g} = \{f_{t,c_g}^{f,1}, f_{t,c_g}^{f,2}, \ldots, f_{t,c_g}^{f,N_s}\}$ and $B_{t,c_g} = \{f_{s,c_g}^{f,1}, f_{s,c_g}^{f,2}, \ldots, f_{s,c_g}^{f,N_s}\}$ from L_t and L_s, with class label c_g.
    Generate $w(\mathbf{z}, c_g)_t$ and $w(\mathbf{z}, c_g)_s$ with noise $\mathbf{z}$ which has standard normal distribution.
    Generate $\tilde{\mathbf{f}}_{t,c_g}^{f}$ and $\tilde{\mathbf{f}}_{s,c_g}^{f}$ with Eq.(3)
    Update G by respective gradient descent on $(Q(\tilde{\mathbf{f}}_{t,c_g}^{f}||\tilde{\mathbf{f}}_{s,c_g}^{f}), \overline{Y_{c_g}})$, keeping Q fixed.
    Update G by respective gradient descent on $(D((\tilde{\mathbf{f}}_{t,c_g}^{f}||\tilde{\mathbf{f}}_{s,c_g}^{f})|Y_{c_g}), 1)$, keeping D fixed.

  end for
end while

IV. EXPERIMENTS

A. DATASETS AND EVALUATION METRICS

To verify the effectiveness of the proposed model, we used two public fashion datasets, the DeepFashion-C [6] dataset and DeepFashion2 [20] datasets. DeepFashion-C consists of 50 grained categories. Within the category class, the difference in the number of data between the class with the most and the class with the least is approximately 4000 times different. Each image in this dataset has coordinates for a bounding box of target clothing and eight landmarks. The image size and proportion are different; therefore, we first cropped each image using a bounding box. Then, we resized each image to fit the height or width of 224 while maintaining the proportion and filled the remaining part with a black background so that the final image was 224 × 224 in size. We used 209,222 of the 289,222 images in the data set for training, 40,000 for validation, and the remaining 40,000 for testing. DeepFashion2 consists of 13 categories, and the difference between the maximum and minimum values of the number of data for each class is approximately 100 times. Each category has a different number of landmarks, up to 39. Images in this dataset were also resized to 224 × 224 size in the same manner. We used 312,184 of the total 364,674 images in the data set for training and the remaining 52,490 for testing.

We used the standard top-k classification accuracy to evaluate the overall classifier performance. In addition, we employed two indices that are unbiased to a specific class to measure the robustness against imbalanced data: the average class-specific accuracy (ACSA) [7], [13] and geometric mean (GM) [32]. Assuming that a confusion matrix exists for n classes and each component is a_{ij}, ACSA and GM were calculated as follows:

$$ACSA = \frac{1}{n}\sum_{i}^{n}\left(\frac{a_{i,i}}{\sum_{j}^{n} a_{i,j}}\right),$$
$$GM = \left(\prod_{i}^{n}\frac{a_{i,i}}{\sum_{j}^{n} a_{i,j}}\right)^{\frac{1}{n}}. \quad (6)$$

B. IMPLEMENTATION DETAILS AND NETWORK TRAINING

The proposed model was implemented and trained in Pytorch [33]. We applied an Adam optimizer [34] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. In each iteration, we set the batch size to 256. The learning rate was initialized at 5e-5 and decreased to 1e-5 after three epochs. The weight parameter λ was set

TABLE 2. Comparison of classification performance of loss variants on the DeepFashion dataset.

Algorithm	Loss	Top-1(%)	ACSA
Zhang *et al*	CE	71.65	0.3343
Ours w/o Convex-GAN	CE	73.95	0.3237
Ours w/ Convex-GAN	MSE	74.27	0.3109
	Focal	74.36	0.3430
	LDAM	**74.38**	**0.3828**

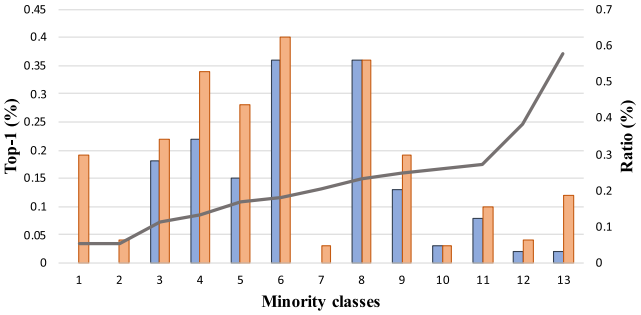


FIGURE 2. Results of top-1 accuracy on minority classes in DeepFashion dataset. The blue and orange bars indicate the results of ours without and with Convex-GAN, respectively. The black line refers the proportion of each class in the entire data set.

to 10^3. Moreover, F_t and F_s of the proposed model loaded the weights that are pre-trained on ImageNet. We froze the pretrained weights till `conv4_3` layer to prevent overfitting. In Phase 1, we pre-trained the shape stream, including F_s, f_s, LDM, and the auxiliary classifier for three epochs. Afterward, the entire network was trained for 15 epochs. When composing feature dictionaries after Phase 1, we included up to 2000 features per class. In the case of classes with less than 2000 total data among the classes, the entire number was included. For example, if the number of data is 1500, all 1500 data are included. In Phase 2, G, D, and the main classifier were trained for 10 epochs. The weight parameter of D was updated every five iterations to stabilize adversarial learning and prevent the discriminator performance from becoming dominant compared to other modules. Three modules, G, D, and Q, were trained in three steps. First, Q and D were trained as real samples in feature dictionaries. In the next step, artificial samples were generated by fixing the weight parameters of G and generating convex weights. Additionally, Q and D were trained with the generated data. In the final step, we update the weighting parameters of G with the error values obtained after generating new artificial samples and fixing the weighting parameters of Q and D.

C. COMPARISON WITH THE STATE-OF-THE-ART METHODS

1) DeepFashion

To verify the effectiveness of the proposed method, we compared its performance to that of the existing state-of-the-art method of the DeepFashion dataset from Zhang *et al.* [16]. Recently, the model has reported the best performance in the fashion category classification recently.

TABLE 3. Comparison of classification performance of loss variants on the DeepFashion2 dataset.

Algorithm	Loss	Top-1(%)	ACSA	GM
Zhang *et al*	CE	81.89	0.6949	0.6673
Ours w/o Convex-GAN	CE	83.65	0.7185	0.6870
Ours w/ Convex-GAN	MSE	**83.88**	0.7084	0.6418
	Focal	83.81	0.7205	0.6922
	LDAM	83.80	**0.7291**	**0.7086**

Table 2 compares the results between the state-of-the-art model from Zhang *et al.* and the proposed model with loss-type variants of the top-1 and ACSA on the DeepFashion dataset. "Ours w/o Convex-GAN" indicates the proposed model trained only up to Phase 1, and without training through minority oversampling using the convex generator. "Ours w/ Convex-GAN" refers to the fully trained proposed model. The top-1 accuracy of the proposed model without the convex generator is 2.3% higher than that found by Zhang *et al*. In the case of ACSA, the proposed model without the convex generator was lower than that found by Zhang *et al*. However, the model trained with the convex-generator outperformed the model of Zhang *et al*. Furthermore, the results demonstrate that the performance is higher when reweighting losses and when Focal and LDAM are used. Compared to the baseline MSE loss, when we used Focal and LDAM losses, the top-1 accuracy of 0.09% and 0.11% increased, and the ACSA of 0.321 and 0.719 increased, respectively.

Figure 2 presents the results of the top-1 accuracy for the minority classes whose count of data is in the bottom 25% of all classes in the DeepFashion dataset. These classes represent less than 1% of the total data. The blue and orange bars indicate the results of the proposed model without and with Convex-GAN, respectively. The black line refers to the proportion of each class in the entire data set. Figure 2 illustrates that oversampling through Convex-GAN can effectively improve the performance of minority classes. Labels (labels 1, 2, 7 in Figure 2) that were not predicted at all in the model without Convex-GAN because the number of data was too small demonstrate substantial performance improvement through Convex-GAN-based oversampling.

2) DeepFashion2

To verify the effectiveness of the proposed method on different imbalanced clothing datasets, we compared the performance with the existing state-of-the-art method on the DeepFashion2 dataset. Table 3 lists the comparison results between the model by Zhang *et al.* and the proposed model with loss-type variants of top-1, ACSA, and GM on the DeepFashion2 dataset. As with the results on DeepFashion, the proposed model, which used minority oversampling via a convex generator with adversarial training, outperformed the state-of-the-art method in terms of all metrics. For the top-1 performance, the model trained using the MSE loss displayed

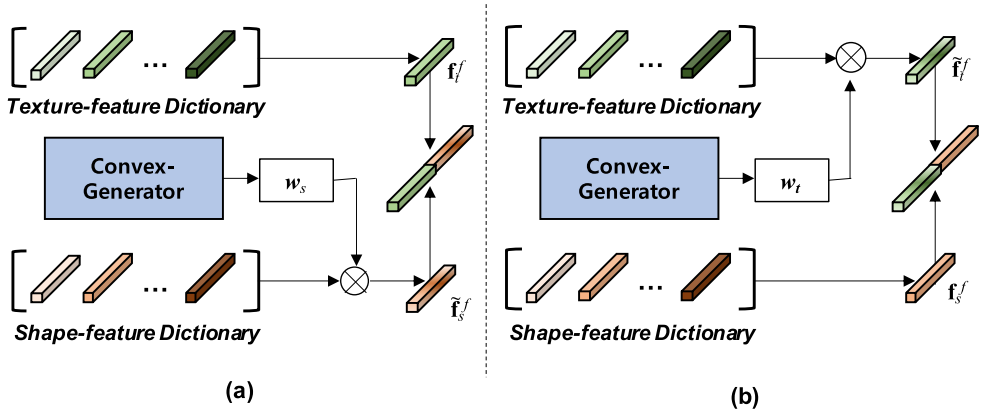


FIGURE 3. Two network structures of convex combination for only shape-stream and only texture-stream. (a) Only shape-stream feature generation. (b) Only texture-stream feature generation.

TABLE 4. Performance for analyzing the contribution of auxiliary classifier on the DeepFashion dataset.

Algorithm	Top-1(%)	ACSA
Ours w/o Aux-classifier	71.51	0.3090
Ours w/ Aux-classifier	73.95	0.3237
Ours w/o Aux-classifier + Convex-GAN	70.58	0.3778
Ours w/ Aux-classifier + Convex-GAN	**74.38**	**0.3828**

TABLE 5. Performance for analyzing the contribution of convex-generator on the DeepFashion dataset.

Algorithm	Top-1(%)	ACSA
Ours w/o Convex-GAN	73.95	0.3237
Ours w/ Convex-GAN (only shape stream)	72.43	0.3692
Ours w/ Convex-GAN (only texture stream)	72.88	0.3422
Ours w/ Convex-GAN	**74.38**	**0.3828**

the highest performance, and in ACSA and GM, the model using the LDAM loss also exhibited the highest performance. Compared with the model by Zhang *et al.*, the respective performance increased by 1.99%, 0.0342, and 0.0413 for the top-1, ACSA, and GM, respectively.

D. ABLATION STUDIES

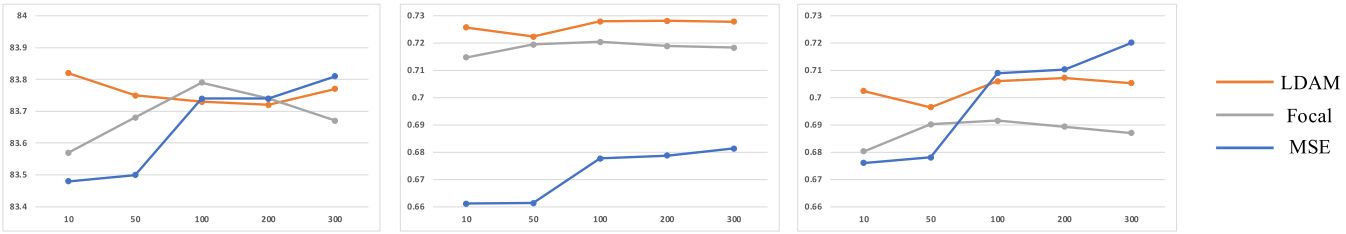
To understand how the proposed component affects imbalanced classification performance, we performed an ablation study on the DeepFashion dataset. We conducted experiments with and without each element, including the auxiliary classifier and convex generator for the texture/shape stream, and added and analyzed the effectiveness of each component based on the performance differences.

1) AUXILIARY CLASSIFIER

To verify the effectiveness of the auxiliary classifier, we performed four experiments. The first experiment trained only feature extractors and classifiers. Except for the auxiliary classifier, the model architecture was similar to the network used in Phase 1. The second experiment was trained with a model with an auxiliary classifier added, used in the first experiment. The third and fourth models were fine-tuned by oversampling through a convex generator for the first and second models, respectively. The results of the top-1 accuracy and ACSA are provided in Table 4. Compared with the base model without the auxiliary classifier, the performance of the added model improves by 2.44% for top-1 and 0.0147 for ACSA. When we applied adversarial training through the convex generator to a model without the auxiliary classifier (the third experiment), ACSA increased, but the top-1 accuracy decreased. This result demonstrates that adversarial training is more stable because the shape-stream feature learned category labels through the auxiliary classifier. As a result, the proposed model (fourth experiment) increased by 3.8% for top-1 and 0.005 for ACSA compared to the model without the auxiliary classifier.

2) CONVEX GENERATOR

To verify the effectiveness of the convex generator of each stream in the proposed model, we performed four experiments. The model of the first experiment was the proposed model without the convex generator, the same as that trained in Phase 1. The second model used artificial samples synthesized through the convex generator only for shape-stream features and real-world samples for texture-stream features. Conversely, the third model used only texture-stream synthetic and real-world samples for shape-stream features. Fig. 3 displays the procedures for artificial sample generation for both cases. The fourth experiment was the fully trained model. The results are provided in Table 5. When we fine-tuned the classifier by generating only shape-stream features, it increased by 0.0455 for ACSA but decreased by 1.52% for top-1 accuracy. In addition, for texture-stream-only generation, the performance of ACSA improved by 0.0185, but top-1 accuracy was reduced by 1.07%. These results reveal that separately generating two-stream features



(a) Top-1 (%) (b) ACSA (c) GM

FIGURE 4. Results of top-1 accuracy, ACSA and GM obtained by loss type and number of samples on DeepFashion2 dataset. (a), (b) and (c) refer to the graph of results of top-1 (%), ACSA and GM, respectively. The orange, gray and blue line indicate LDAM, Focal and MSE, respectively. The values on the x-axis refer to the number of features sampled in the dictionary.

is ineffective and has the potential to interfere with the learning of the classifier. When two-stream features are synthesized together, we can expect effective performance improvement through adversarial training of the generator, discriminator, and classifier.

E. ANALYSIS FOR THE NUMBER OF SAMPLES

We set the number of samples tested as a hyperparameter before the convex combination of the features sampled from the convex-weight was obtained through the generator and the feature dictionary. The higher the number of features to be sampled, the more real-world samples were referenced when synthesizing artificial samples. However, the generator was trained while being constrained to generating points within the convex hull formed by the real data of the class of interest; thus, it may become difficult to train the generator if excessively large numbers of real-world samples are referenced. In addition, the training time also increases in proportion to the number of features to be sampled. Therefore, it is important to set the hyperparameter to an appropriate number of samples. In this section, we analyze the trend of category classification performance changes while altering the number of features retrieved from the dictionary. Figure 4 presents the results of the performance of top-1, ACSA, and GM by loss type and number of samples. We used the MSE, Focal, and LDAM loss and five different numbers of samples (10, 50, 100, 200, and 300) for the experiments. We employed the DeepFashion2 dataset where the number of samples of the minority class is sufficient and the number of categories is moderate to effectively test the results. In Fig. 4, the top-1(%), ACSA, and GM results obtained by loss type and the number of samples on DeepFashion2 dataset are provided in order from left to right. The results for the three losses exhibit different trends. First, in the case of LDAM loss, the overall performance for all metrics is high, and the amount of variation depending on the number of samples is insignificant. Second, the medium number of samples (50, 100) for focal loss exhibits the best performance. Third, the results of the MSE demonstrate higher performance as the number of samples increases. According to these results, the number of real sample features used for generation is approximately 100, which delivers effective performance.

V. CONCLUSION

In this paper, we proposed a novel imbalanced classification framework for fashion datasets using feature dictionary-based minority oversampling. We devised a deep network for clothing data through two-phase training schedules in the proposed method. We trained two-stream networks to construct feature dictionaries in the first phase and pretrained the classifier. In addition, we built feature dictionaries using the trained feature extractor. In the second phase, we synthesized artificial samples by employing the convex generator and feature dictionaries. Artificial minority samples were generated by the multiplication of features sampled from dictionaries and convex weights obtained from the generator. We trained the convex generator by playing a minimax game with the discriminator and classifier. Through minority oversampling using the generative model, the classifier was fine-tuned. As a result, we obtained a robust model for imbalanced fashion datasets. We validated the effectiveness of the proposed method using a public fashion dataset with a large inter-class imbalance. The results revealed that the proposed model outperforms state-of-the-art models in several metrics, including ACSA and GM. In particular, the results indicated that the proposed method is more effective because the frequency of the dataset is lower in the class. In addition, we verified the effectiveness of various components constituting the proposed model. The experimental results demonstrated a change in performance depending on the presence or absence of an auxiliary classifier, and oversampling for both shape and texture features effectively improves performance. Finally, we analyzed the performance change trend that varies depending on how many features are sampled from dictionaries to generate artificial data, confirming the optimal number of samples according to the loss.

REFERENCES

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

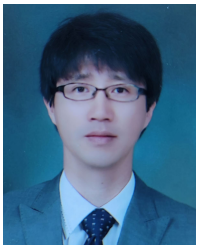
[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[4] S.-A. Rebuffi, H. Bilen, and A. Vedaldi, "Learning multiple visual domains with residual adapters," 2017, *arXiv:1705.08045*.

[5] O. Russakovsky, J. Deng, H. Su, and J. Krause, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[6] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "DeepFashion: Powering robust clothes recognition and retrieval with rich annotations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1096–1104.
[7] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5375–5384.
[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 28, pp. 321–357, Jun. 2006.
[9] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SmoteBoost: Improving prediction of the minority class in boosting," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*. Springer, 2003, pp. 107–119.
[10] M. Lin, K. Tang, and X. Yao, "Dynamic sampling approach to training neural networks for multiclass imbalance classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 4, pp. 647–660, Apr. 2013.
[11] Y. Yan, M. Chen, M.-L. Shyu, and S.-C. Chen, "Deep learning for imbalanced multimedia data classification," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 483–488.
[12] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.
[13] S. S. Mullick, S. Datta, and S. Das, "Generative adversarial minority oversampling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1695–1704.
[14] J. Kim, J. Jeong, and J. Shin, "M2M: Imbalanced classification via major-to-minor translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2020, pp. 13896–13905.
[15] W. Wang, Y. Xu, J. Shen, and S.-C. Zhu, "Attentive fashion grammar network for fashion landmark detection and clothing category classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2018, pp. 4271–4280.
[16] Y. Zhang, P. Zhang, C. Yuan, and Z. Wang, "Texture and shape biased two-stream networks for clothing classification and attribute recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13538–13547.
[17] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," 2017, *arXiv:1705.09368*.
[18] M. Park, H. G. Kim, and Y. Man Ro, "Generative guiding block: Synthesizing realistic looking variants capable of even large change demands," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4444–4448.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[20] Y. Ge, R. Zhang, X. Wang, X. Tang, and P. Luo, "DeepFashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5332–5340.
[21] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[22] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3573–3587, Aug. 2017.
[23] Y. Cui, Y. Song, C. Sun, A. Howard, and S. Belongie, "Large scale fine-grained categorization and domain-specific transfer learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4109–4118.
[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Mar. 2004.
[25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.
[26] S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu, "ModaNet: A large-scale street fashion dataset with polygon annotations," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1670–1678.
[27] M. Jia, M. Shi, M. Sirotenko, Y. Cui, C. Cardie, B. Hariharan, H. Adam, and S. Belongie, "Fashionpedia: Ontology, segmentation, and an attribute localization dataset," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2020, pp. 316–332.
[28] S. Bhatnagar, D. Ghosal, and M. H. Kolekar, "Classification of fashion article images using convolutional neural networks," in *Proc. 4th Int. Conf. Image Inf. Process. (ICIIP)*, Dec. 2017, pp. 1–6.
[29] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[31] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," 2019, *arXiv:1906.07413*.
[32] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. 4th Int. Conf. Mach. Learn.*, vol. 97, Nashville, TN, USA, Jul. 1997, pp. 179–186.
[33] A. Paszke, S. Gross, F. Massa, and A. Lerer, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, Vancouver, BC, Canada, Dec. 2019, pp. 8026–8037.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

MINHO PARK received the B.S. and M.S. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2018 and 2020, respectively. He is currently working with the Artificial Intelligence Research Laboratory, Visual Intelligence Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon. His research interests include machine learning, generative model, and image/video analysis.



HWA JEON SONG received the B.S., M.S., and Ph.D. degrees in electronics engineering from Pusan National University, South Korea, in 1993, 1995, and 2005, respectively. From 1995 to 2001, he was a Researcher at Hyundai Motor Company. Since 2010, he has been a Principal Researcher with the Electronics and Telecommunications Research Institute (ETRI). His research interests include speech recognition, multi-modal representation, and artificial general intelligence (AGI).

DONG-OH KANG received the B.S. degree from Yonsei University, South Korea, in 1994, and the M.S. degree and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1996 and 2001, respectively, all in electronic engineering. Since 2001, he has been working at the Electronics and Telecommunications Research Institute (ETRI), where he is currently a Principal Researcher with the Visual Intelligence Research Section. He is developing artificial intelligence technologies for self-growing multimodal knowledge graphs at ETRI. His research interests include multimodal knowledge representation, knowledge graph completion, and self-growing intelligent agents.