Fashion and Textiles

RESEARCH Open Access

Unsupervised generation of fashion editorials using deep generative model

Minjoo Kang[1], Jongsun Kim[2] and Sungmin Kim[3*]

*Correspondence:
sungmin0922@snu.ac.kr

[1] Department of Fashion and Textiles, Seoul National University, Seoul, South Korea
[2] Department of Fashion Design, Suwon Women's University, Suwon, Gyeonggi-do, South Korea
[3] Research Institute of Human Ecology, Seoul National University, Seoul, South Korea

Abstract

This research intended to establish a new fashion-related artificial intelligence research topic concerning fashion editorials which could induce streams of further studies. A new fashion editorial dataset, which is a prerequisite in training an AI model, has been established in this study to meet the research purpose. A total of over 150K fashion editorials were initially collected and processed to satisfy necessary dataset conditions. A novel dataset of fashion editorials consisting of approximately 60K editorials is proposed through the process. In order to prove the adequacy of the new dataset, data distribution was analyzed and a generative model was selected and trained to attest that new fashion editorials can be created with the proposed editorial dataset. The results generated by the trained model were qualitatively investigated. The model has shown to have learned various features that compose editorials with the dataset, successfully generating fashion editorials. Quantitative evaluation with FID scores was conducted to support the selection of the generative model used for the qualitative assessment.

Keywords: Fashion editorial, Fashion editorial dataset, Unsupervised learning, DCGAN, Fashion GAN

Introduction

Fashion editorial is an advertisement with photographs usually published on fashion magazines or, these days, online. Showing the latest trend in fashion in a visually creative and pioneering way, it covers a wide range of topics and presents various objects that are carefully planned and chosen for the scene. It started gaining its reputation with the advent of style magazines such as i-D and came to take a crucial part in fashion through integration with art (Williams, 2008). Editorial today is treated as piece of art by some and has come to take unneglectable part in not only fashion advertisement but also aesthetics of fashion. Parallel with the industry, it is also extensively used in numerous academic analysis such as synchronic or diachronic analysis of the fashion trend in apparel studies.

Generative adversarial network (GAN), first introduced by Goodfellow et al. (2014), is a generative artificial intelligence model that has remarkably expanded research boundaries in many image related fields. It is a deep learning model constituting of a generator which tracks down a feature distribution of a sample and a discriminator which tracks down whether the given sample is real from a dataset or fake made by the generator. By playing a mini-max game with loss values, generator learns to create fake samples good enough to deceive discriminator, while discriminator follows up to learn to distinguish the theoretically ever-progressing generator from the dataset until reached equilibrium. From image generation to reconstruction of 3D object from point cloud, it has walked a long way throughout the last decade enabling many of industrial sectors to apply this technology to the problems they had difficulty solving with previous existing methods. With its ability to produce what has never been before, many have explored the creative aspect of GAN, training the model to create face of a real or a fake person (Lin et al., 2022; Marriott et al., 2021; Radford et al., 2016), draw art works (Jones, 2017; Robbiebarrat, 2017), sceneries (Jones, 2017; Radford et al., 2016; Robbiebarrat, 2017) and even, to generate 3D objects (Lin et al., 2022; Marriott et al., 2021; Wu et al., 2016).

Unlike the stream of advanced GAN studies, however, in terms of fashion where images such as photos or illustrations are essential, academics generally focuses on two broad topics. One is the generation of apparel product images and the other is the image-based virtual fitting of those items. These studies concerning fashion product photos started with the advent of Fashion-MNIST dataset consisting of 70K apparel images published by Xiao et al. (2017) (Fig. 1). Thereafter, engineering academics endeavored to supplement apparel attributes or categories in order to advance GAN into generating more detailed and realistic apparel images (An et al., 2023; Choi et al., 2023; Kumar & Gupta, 2019; Lee & Lee, 2019; Lin et al., 2021; Pernus et al., 2023; Ping et al., 2019; Rostamzadeh et al., 2018), and also furthered their research into image-based fitting simulation which suggested possible GAN application in online virtual try-on in the future (Jetchev & Bergmann, 2017; Lang et al., 2020; Liu et al., 2019; Pandey & Savakis, 2020).

Fashion relevant GAN, however, fails to diverge from these two topics until today. This is primarily because there exist no other tangible big datasets other than that of the apparel

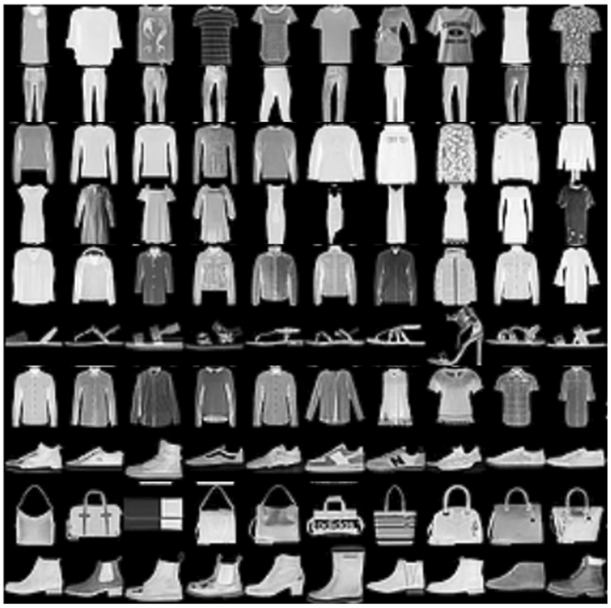


Fig. 1 Fashion MNIST (Xiao et al., 2017) Note. Adapted from Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, by Xiao et al., (2017). Copyright 2017 by the Zalando SE

product images, like Fashion-MNIST, which can easily be scraped from online shopping malls, and secondarily because these GAN studies are held by the engineering academics who have knowledge and ability to collect necessary data and adopt the technology but lack in ideas of fashion related topics. Generating a detailed and realistic fashion item is surely an important task for GAN because photo-realistic level of accuracy is demanded to be undoubtingly accepted by the fashion industry. In order to do this, a proper large-scale fashion product dataset is a prerequisite that follows the improvements in algorithms and there still is a long way to go to formulate the absolute dataset.

Due to this absence, many other studies focused on developing methods to naturally transfer, for instance, a black blouse worn by someone in T-pose or possibly just a product image not worn by a person, to someone posing as, say, the Statue of Liberty. In spite of how practical the usage of GAN would be for a virtual try-on in future, however, fashion is not solely about practicality as seen by the engineering sector. It is a field where creativity and originality are essential. There is no doubt in saying that this is especially true for fashion design for, strictly speaking, high fashion. General everyday clothes are not designed to stand out significantly. They rather look quite similar to one another. Thus, in order to reach the public and be the needle found in this haystack to raise a sale, fashion advertisement, usually in a form of photograph, becomes relatively more creative than fashion design itself to catch the eyes of the consumers. These picture form of advertisements are called, fashion editorials. Therefore, taking into mind how familiar and widespread editorials are in both the industry and the academia from the fashion field perspective, a novel research topic that encompasses both practicality and creativity has been proposed in this study—an artificial intelligence (AI) that concerns fashion editorials. The foremost priority to actualize this research idea is not in developing a generative AI model. The first and foremost task is to prepare an adequate dataset that could be used to serve this purpose. Because there exists no large enough fashion related dataset other than fashion product images, it is necessary to establish a new fashion editorial dataset which could be used to train a model that can generate creative fashion editorials.

By establishing a new fashion relevant dataset, a collection of fashion editorials which can be used to train any generative models in an unsupervised way, this research fills a gap in between the state-of-the-art generative artificial intelligence technology and a void of its application in the fashion field. It not only broadens the boundaries of fashion associated AI research beyond limited topics of realistic and detailed product image generations and virtual try-on held by the engineering field, but also initiates a novel research topic which can be followed by a stream of further studies of expanding the size of the dataset, various tagging works on different specification levels, categorization or classification of fashion editorials in the dataset for a more specific unsupervised generation of editorials, implementation of supervised learning of editorials, supervised high-quality editorials generation with diffusion model or VQ-VAE, and so on.

Literature Review

Supervised learning

Type of learning conducted to artificial intelligence heavily depends on a dataset it is trained upon. Existence of labels in a dataset determines whether the learning is to be unsupervised or supervised. Fashion-MNIST dataset, even though they are not presented in Fig. 1, contains not only the fashion item images but also 10 corresponding labels—t-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, ankle boot. Labeled dataset is capable of both learnings. Labels can be excluded from training to perform unsupervised learning as opposed to supervised learning performed with the labels. In supervised learning, because there is (or are) designated name, a tag, namely a label, paired with the data, there is a certain expectation upon users as to what to expect as a result. For instance, in case of using supervised Fashion-MNIST trained GAN, when given a label "sandal", it is expected of this particular GAN to draw images of sandals and no others such as trousers or t-shirts.

Supervised learning is often conducted to generative models because labels are useful in guiding the model to generate a desired image result in general. Many a deep generative model recently proposed such as Stable Diffusion and Dall·E 2 have been trained on a tremendous amount of data with labels. Stable Diffusion is known to be trained on 5 billion data (Beaumont, 2022), and because more than one label could be paired with a single data depending on the level of specificity of the dataset, total number of tags may easily exceed 10 billion. Hence, a tremendous tagging work is required. Especially, considering the level of specification the users need to provide to these trending generative models to achieve an image that meets their needs, it is assumed that an exceeding number of text labels are present in these datasets. For example, in case of a fashion editorial where a female model holding a white flower poses in front of a dark burgundy colored studio wall, a simple singular tag 'editorial' could be given. However, when you go into describing the scene specifically, tags may go from a somewhat detailed 5 tags such as 'female', 'holding a flower', 'white flower', 'burgundy wall', 'dark' to a very precise description about the atmosphere of the scene, her clothes, posture, facial expression, the way she is holding the flower, how the flower looks like and what kind of flower she is holding, and it could go on and on. If given 5 labels each, the sole number of tags (and tags only) would be 25 billion for a dataset of 5 billion images, consisting of up to 30 billion information in total. Moreover, because misleading labels can deviate learning of a model, labels must be tagged manually with discretion in order to provide adequate information that precisely describe the data. Like so, implementation of supervised learning could be very expensive, especially when it is trained upon a large-scale dataset with at least more than a billion labeled data. Considering the number of images abovementioned large-scale dataset (5 billion images), many must have had gone through years of labeling labor for conformable supervised learning to reach the level the trending generative models are in now. With the absence of these massively labeled datasets, such generative models would not have reached where they are at present.

Unsupervised learning

Unsupervised learning, conversely, does not require such expensive human tagging labor because it is trained without labels and what is to be learned is determined by the model. It is known that during unsupervised training GAN learns certain concepts itself in a way that is unknown to the users from the dataset. Examining the latent filters of an unsupervised GAN trained on LSUN bedroom has shown to have learned bed and windows. This made possible the synthesis of the desired features through filter (vector) arithmetic. A subtraction of man without glasses from man with glasses and then an addition of woman without glasses sums up to generating a woman with glasses. In case of Fashion-MNIST GAN that went through unsupervised training, it automatically learns not to attach sleeves or a collar to a sandal or a boot by examining the data laid in the dataset without any further information, such as labels, provided by human. This implies that the model may not know what sleeve, collar or sandal is per se, but knows that they are not the same and that they are something different, and that a sandal does not consist of a sleeve or a collar. GAN can not only generate, for instance, sandals that are present in the dataset, but also it has ability to create new sandal designs using the aspects of a sandal it learned from training, because it has learned what makes a sandal, to be a sandal, regardless of the data labels.

Diffusion/VQ-VAE model versus GAN

The famous generative models mentioned earlier, Stable Diffusion and Dall·E 2 are built using a diffusion model (Ho et al., 2020; Sohl-Diskstein et al., 2015) and a VQ-VAE (vector quantized variational autoencoder) model (Van Den Oord et al., 2017), respectively. Both models are similar in a way that they are trained to generate image that is most likely to a given data by dataset. Diffusion model learns to output a restore of a noised input image and VQ-VAE learns to minimize discrete vectors' distances between an input and an output (result). They try to mimic the input data, thus when trained in unsupervised fashion, they are unable to create, for example, a new style of sandal like GAN, but rather train themselves on how to copy the shown sandal well. When given a picture of a sandal with sleeves attach on both sides during unsupervised training, a well-trained GAN knows this is a bizarre case because sandals do not normally have sleeves on them and considers it as an outlier, whereas diffusion or VQ-VAE try their best to create the most look-alike of this unusual sandal. Image must be learned with its labels by the latter models in order to generate creative results they show today due to this property. To be more concrete, whereas GAN can create an orange strap sandal with a white buckle even though this particular design does not exist in the dataset in an unsupervised learning, diffusion model and VQ-VAE have to be given the precise description of a sandal—"An orange sandal with a white buckle"—in a supervised fashion in order to generate something creative. They require a immense and tagged dataset to be trained upon to learn objects and distinguish them in a given scene, and be articulately guided to generate a creative or an original result. Accordingly, diffusion model and VQ-VAE were not considered suitable in testing the validity of the gathered editorial dataset due to their replicative feature in an unsupervised training. Their aptitude for replicating could not only disrupt editorials' creative and original features, but also cause plagiarism issues. Additionally, no matter how realistic and well-functioning, implementing a supervised diffusion model or a VQ-VAE is too costly in this initial stage of dataset preparation. Hence, an unsupervised learning of GAN was chosen to be conducted with the editorial dataset of 60K images.

Some may assert in using already well proven models like Stable Diffusion or Dall·E 2 to generate new editorials. They are not, however, suitable for fashion editorial generation. In order to generate editorial with these models, a highly detailed description must be provided to the prompt (Fig. 3a). This limits user's imagination in the generation and construction of a specific description itself could be a difficult task unless the user has a definite image he or she wants produce in mind. Also, because these models only draw what they are requested, it is only possible to examine the most possibly similar editorial in the user's head and not its variations. This suggests that these two models are inadequate in being a creative and original image reference source in the fashion field. Additionally, for they are not specifically trained on images from fashion domain, a simple description of a floral dress, for example, produces daily-wear-like or product-like images rather than what fashion majors would want to find—trendy fashion shoots or editorials that present a floral dress (Fig. 2b). These aspects could hold back fashion fields' interest in further exploring the generative AI technology and hinders its adoption to the field. Moreover, deployed models like Stable Diffusion and Dall·E 2 could not be used with the new editorial dataset because the information necessary to continue the training the models underwent is inaccessible. Thus, these already deployed models cannot personally be train to our taste. On the other hand, a GAN, built from scratch and supervised, can not only always continue its training from the last state, but also has ability to generate variations of what was requested by, for instance, adding various visual or prop details. When trained on editorials, it will be able to create trendy fashion scenes with a simple tag. Even when unsupervised, it can generate original editorials unlike aforementioned generative models.

As explained through the previous two paragraphs, GAN was selected as the generative model over diffusion/VQ-VAE model or Stable Diffusion/Dall·E 2 to confirm the newly proposed editorial dataset. To abridge, data labeling of supervised learning is costly and taxing compared to unsupervised learning which does not require data labels. However, diffusion and VQ-VAE models have replicative feature, so they must be supervised in order to create non-dataset-existent (creative) image, whereas GAN is able to create non-dataset-existent image in both unsupervised and supervised fashion which makes it an adequate tool for an initial study like this study. A new fashion editorial dataset of 60K editorials has been established and an unsupervised learning of GAN was conducted to confirm adequacy of the proposed dataset. Specifically, considering the fact that the training of GAN could be highly unstable, a well-known limitation of GAN which led to the development of deep convolutional GAN (DCGAN) (Radford et al., 2016), a DCGAN model was employed for unsupervised learning of the fashion editorials. DCGAN has shown to have learned different features of paintings in different genre like abstract, landscape, portrait and so forth through an unsupervised learning (Fig. 3). In Fig. 3a, although not so clear, it has learned to draw a figure that resembles a naked human body for nude portraits and a human figure with a face for portraits. The accuracy of a general human figure, however, is a resolvable problem. If possible, potentially more training with a large-scale dataset for each genre of painting can easily overcome this limitation. Like so, it was concluded that DCGAN is capable of generating various scenes with a human-like figure, thus a DCGAN was constructed to check whether it is able to learn how to draw a human model, whole or partial, in an editorial like composition. Approximately 150K online fashion editorials were gathered and manually processed, cropping unnecessary blank spaces along the borders, or attached articles, to form a valid editorial dataset of 60K data for the DCGAN to be trained upon.

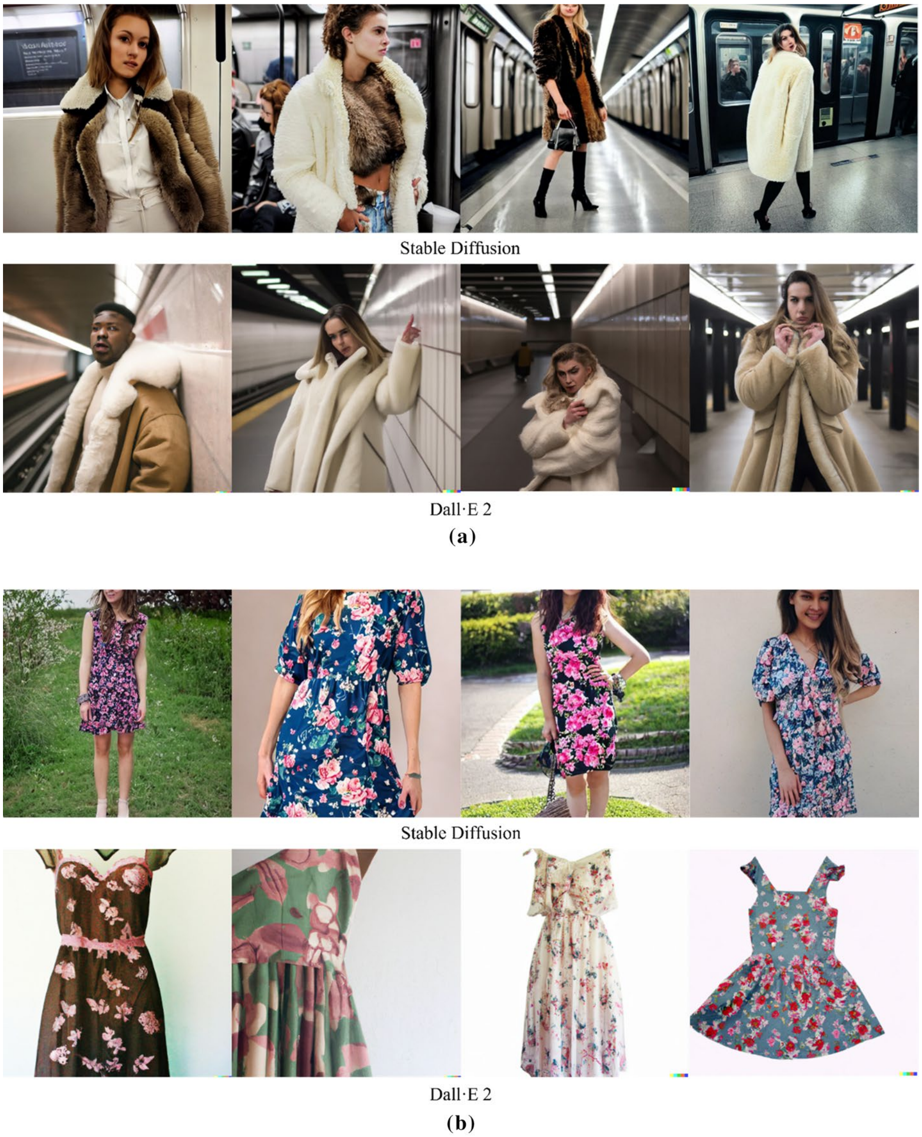


Fig. 2 Results from Stable Diffusion and Dall·E2. **a** Fashion editorials prompted with "A model wearing a furry ivory jacket posing in a subway". **b** A floral dress prompted with "floral dress" Note. Adapted from Stable Diffusion, by Stability.ai, 2023 (https://clipdrop.co/stable-diffusion). In the public domain. Adapted from Dall.E, by OpenAI, 2023 (https://labs.openai.com/). In the public domain

Methods

Preparation of dataset

A total number of over 150K fashion editorials were retrieved both manually and automatically via web crawler from online websites. Collected image data underwent inspection with discretion. Images were either eliminated or processed to prevent DCGAN from learning improper or unnecessary things. The purpose of the inspection is to focus the generative model's attention to the desired learning features such as human models, props, objects, proper color combinations and other various visual features within the scene. Examples of processing include cropping unnecessary blank spaces that were appended while being downloaded through web crawler, cropping unnecessary texts

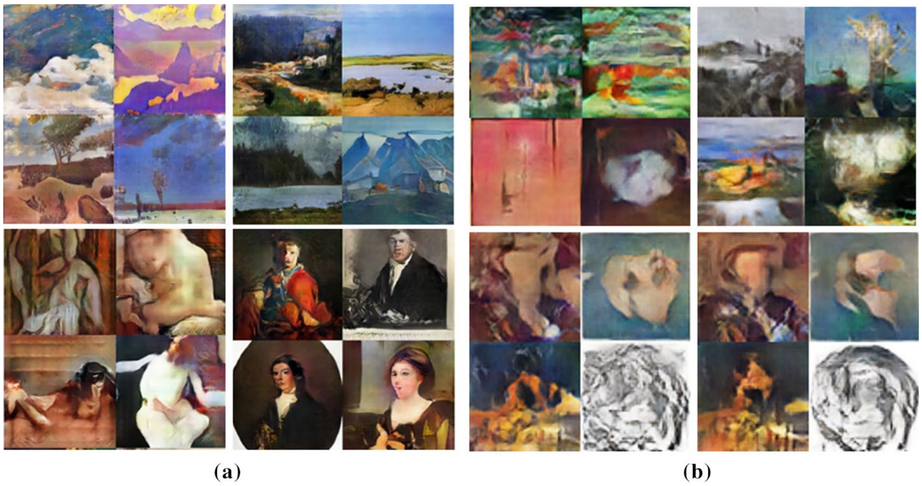
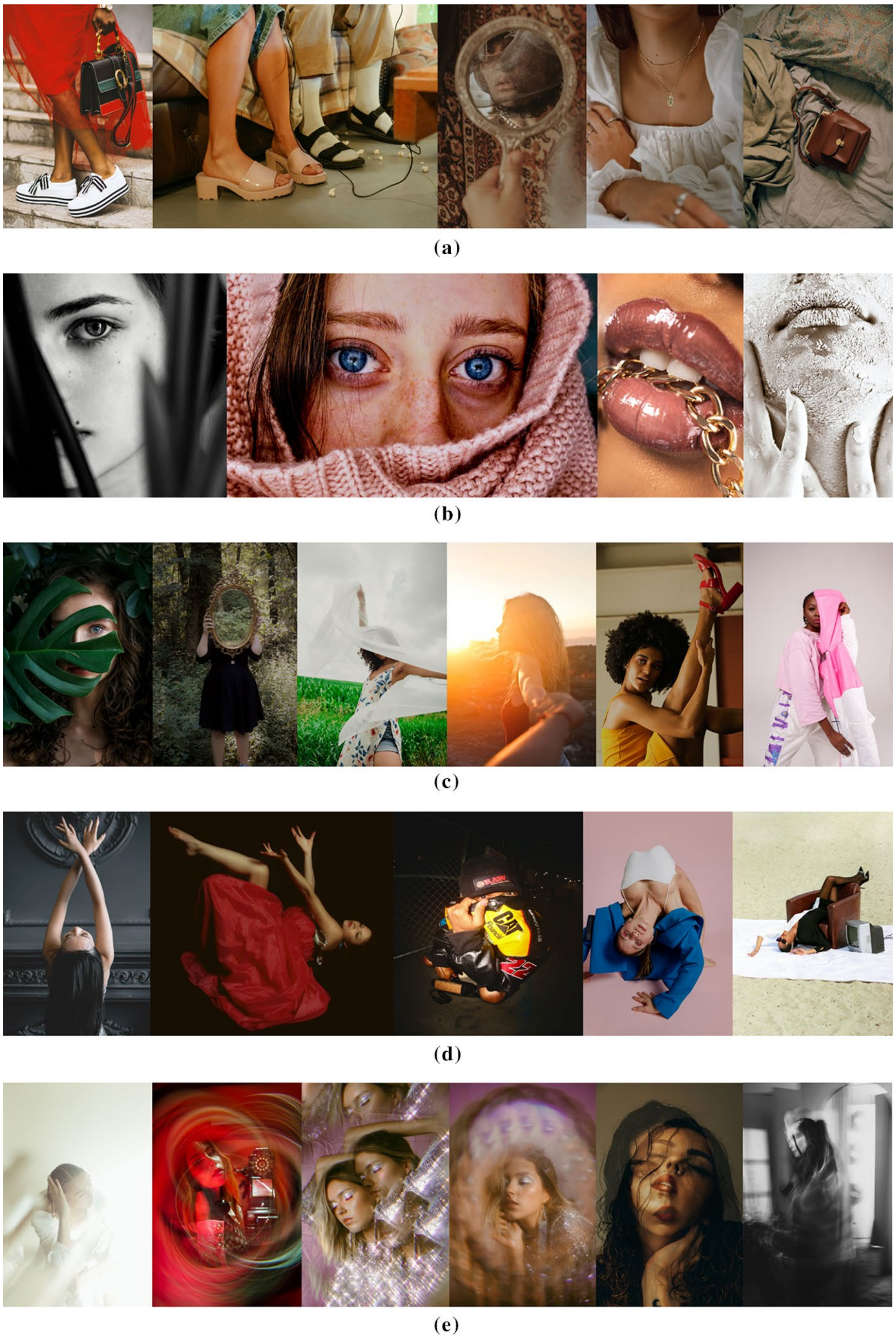


Fig. 3 Genre paintings generated by GANs. **a** Art-DCGAN. (From left to right, top to bottom) Abstract landscape, landscape, nude-portrait, portrait. **b** GAN-Gogh. (From left to right, top to bottom) Abstract, landscape, nude, portrait **a** Note. From art-DCGAN, by Robbiebarrat, 2017 (https://github.com/robbiebarrat/art-DCGAN?tab=readme-ov-file). In the public domain; **b** Note. From GANGogh: Creating Art with GANs, by K. Jones, 2017 (https://towardsdatascience.com/gangogh-creating-art-with-gans-8d087d8f74a1). In the public domain

which are part of an article published with an editorial, splitting multiple photos into single images, etc. Some editorials with multiple photos were left as they are, only when series of photos within are displayed in a chronicle manner, for it was reflected as one of a kind. Since it was desired of the initial DCGAN to specifically learn human features, editorials that do not or only partially display body parts such as legs or arms were removed from the collection except for face close-ups (Fig. 4a). They were considered as a type of an editorial, a portrait kind. Quite a few eye(s) and lip close-ups were found, so they were also considered as one of a sort (Fig. 4b). If over half the proportion of the body (or a face) is shown, waist-up, waist-down or a vertically splitted half of a torso with an arm and a leg, the image was not removed from the dataset because it was considered adequate to learn human body construct (Fig. 4c). The ones with excessively eccentric poses or ones with excessively dynamic postures were eliminated to some extent for opposing reason, but not entirely, because the dynamicity was thought as one of the features of the editorials (Fig. 4d). Editorials that are too blurry, dark or has visual effects that are thought to be entangling natural human features were also removed (Fig. 4e). Ultimately out of 150K, a fashion editorial dataset with 60,499 images was established. Images in Fig. 5 are free to use under Pexels license to present visual aid of what data have or have not been removed.

Construction of DCGAN

Then, DCGAN, a composite model of two sub-models, a generator and a discriminator, is built following the practice of Brownlee (2019). Each sub-model consists of five convolutional layers, four of which up- or down-sampling data with 2 × 2 stride, followed by LeakyReLU (Leaky Rectified Linear Unit) activation function of slope 0.2. The discriminator is activated on sigmoid while the generator is activated on hyperbolic tangent function in the output layer. Both of the sub-models and the composite model are



(a)

(b)

(c)

(d)

(e)

Fig. 4 Editorial evaluation examples for the dataset. **a** Eliminated examples where human parts are not or only partially displayed. **b** Remained examples of eye(s) and lip close-ups. **c** Remained examples displaying partial human body. **d** Eliminated examples with overly dynamic postures or composition. **e** Eliminated examples with visual effects where human features are unclear

compiled using optimizer of Adam with a learning rate of 0.0002. Unlike the practice suggested by Brownlee (2019), filters of the 4 convolutional layers of the generator were set to 256 as opposed to the 128 filters of the discriminator, in order to advance generator's ability to generate more objects considering the fact that an editorial contains many

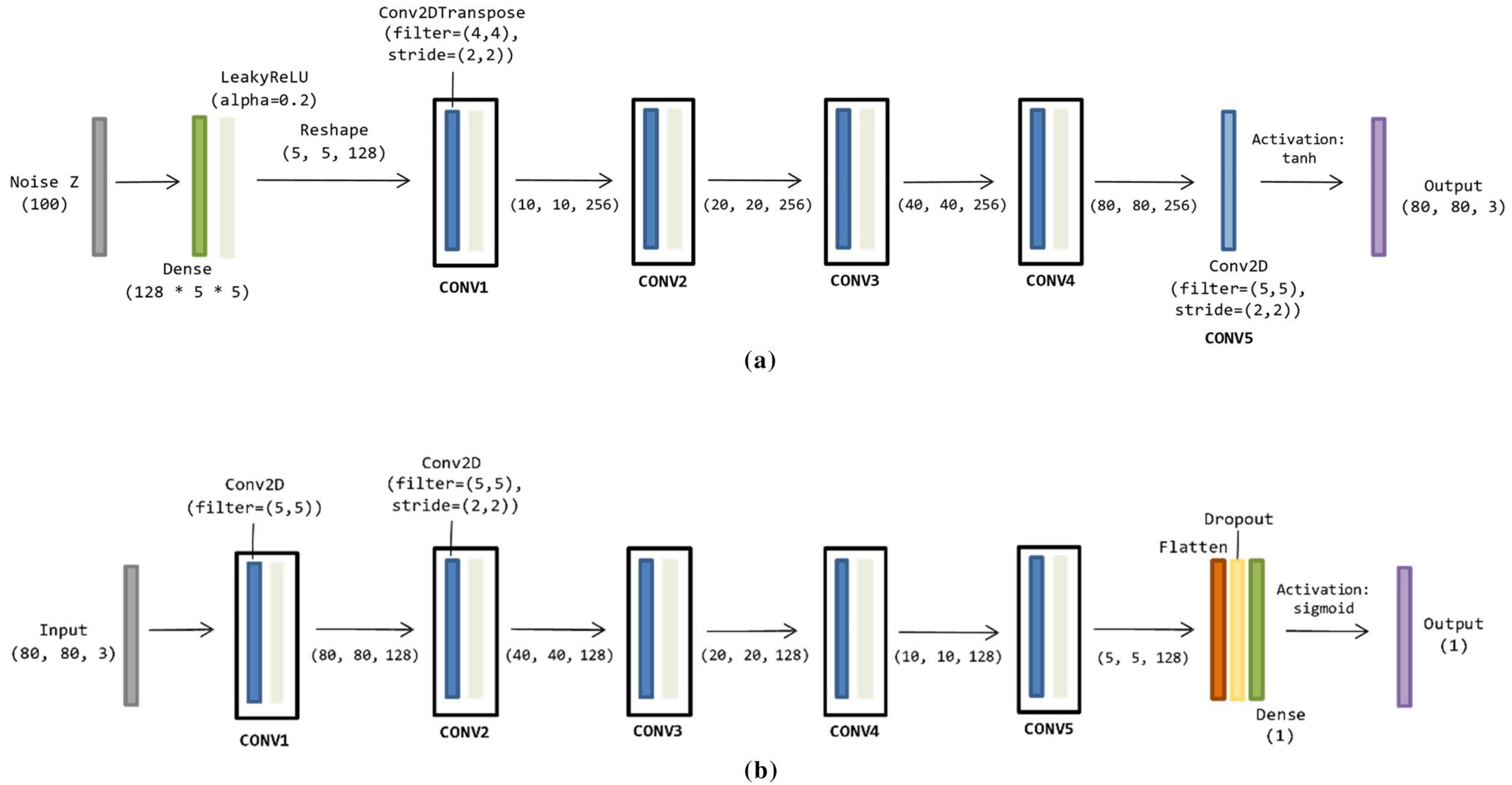


Fig. 5 Structure of the editorial DCGAN. **a** Generator. **b** Discriminator

visible features in a single scene. Experiments have shown that the generator presented better, more sophisticated results with the filter size of 256 compared to the results when it was set to 128. The loss of the generator has also improved from around being around 4–10 to 1–4, which implies that the generator is doing a better job in fooling the discriminator than before. The overview of the editorial DCGAN is as in Fig. 5.

The training dataset was not augmented in anyway but resized to 80 × 80 pixels. Because images were not cropped to a square in the dataset preparation stage, those with different proportions were, not cropped, but compressed to meet the 1:1 square size. Random cropping and center cropping of the data to fit the square were tested, but since many of the editorials in the dataset do not necessarily have human models placed in the center, these types of cropping led to emptying the scene of a human or leaving only partials of a body which would then make the image in question inadequate for the training—the subject of elimination. To avoid this from happening, images were resized, though maybe distorted, to preserve models' body features. The training was conducted on 10G memory with GeForce RTX 3060 graphics card on Ubuntu 22.04 operating system. The model was written in Python 3.10.6 using Tensorflow version of 2.11 and trained for total of 1600 epochs with mini-batch of 128 which took 12 min per 10 epochs. The two sub-models were saved at the end of every 10 epochs, a checkpoint, for safe-keeping.

Results and Discussion

Data distribution

Data distribution of the editorial dataset was analyzed along-side with training actual generative models with it. Editorials of 80 × 80 used for the training was examined. Mean R, B, G value in each pixel is as shown in Fig. 6. Oscillation over pixels in each color channel is examined. This means that color fluctuates over the scene which indicates diversity and complexity of the editorial scene. This implies the inconsistent, or in other words, creative feature of fashion editorials which is expected, but could potentially lead to inconsistency and difficulty in training. This further indicates that classification of the proposed dataset which could reduce visual diversity in a scene will improve overall dataset quality and is a priority that follows.

Color frequency histogram of the dataset is as shown in Fig. 7. Each channel parallels bimodal distribution, the utmost peak leaning towards the darkest color. This reveals

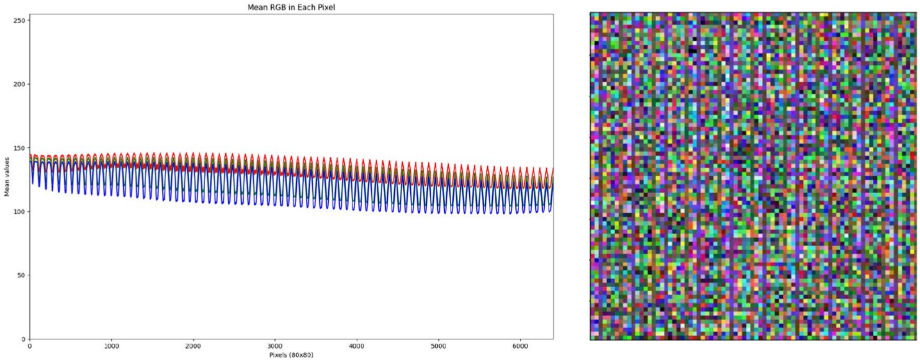


Fig. 6 Mean value of R, G, B in each pixel and its corresponding image

general tendency of dark tones in editorials to convey luxurious and elegant atmosphere. However, editorials collected for the dataset are not limited to editorials by high-end luxurious fashion brands and visual inspection of the dataset shows numerous editorials with bright colors, especially in their backgrounds. The second highest peak in the distribution is assumed to be skin of human models because the dataset was processed to always contain human figures and because skin colors generally have the highest red value, followed by green and blue as shown in the graph. This indicates that with proper manipulation, a generative model will be able to classify human figures and separate them from editorials in the dataset. After subtracting human models out of the scene, color frequency distributions will closely follow log normal distribution. Further investigation with the new distribution could tell what are learnt from the editorial which could reveal underlying reason behind the frequency discrepancies in the dark and bright colors.

Qualitative evaluation

Qualitative evaluation must be held for the generated results because the training of GANs tend to be highly unstable and there exists no absolute method to quantitively evaluate the model. Hence, close inspection of its state at every checkpoint of training is crucial. Some editorials generated by the DCGAN at the 1600th checkpoint of the training is as in Fig. 8. The scrutiny over the results revealed that the generator is overfitted. This implies that the training was excessive and that the generator has learned to draw several specific images that can certainly fool the discriminator as in images highlighted in a red box in Fig. 8.

Rolling back to previous checkpoints, the generator at 1000th epoch suggested the best practice without any indication of the overfitting among 4000 editorial results. However, results created by the 800–1000th-epochs-trained-generators are presented

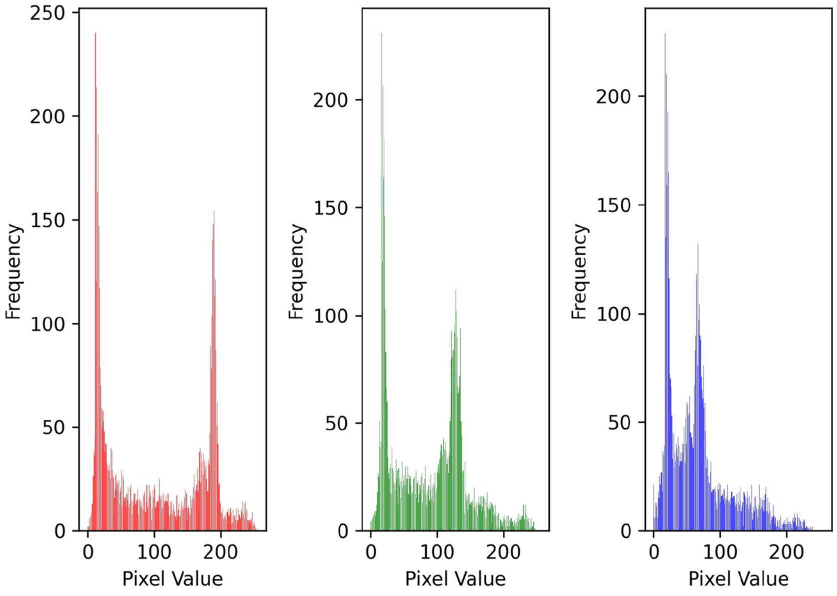


Fig. 7 RGB frequency

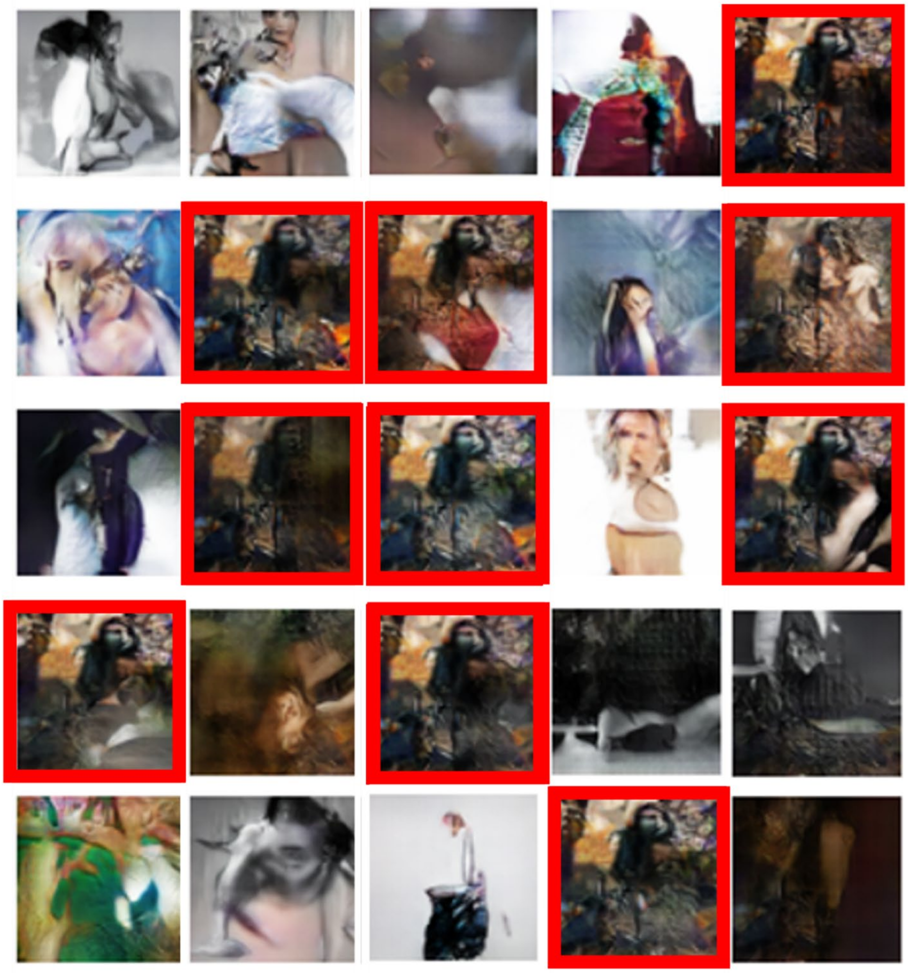


Fig. 8 Editorials from 1600th checkpoint. Results highlighted in a red box reveals the alleged overfitting of the generator

in this study for the 1000th-epoch model may have other glitches that are unnoticed. Inspection of these general results revealed that the generative model has successfully learned certain features that compose editorials from the proposed dataset (Fig. 9). However, it is still too early to assert that the model has fully learned the construct of a human figure, thus whether or not it has intentionally generated a face of a body with visual effects like reflection or duplication is uncertain at this stage (Fig. 10c). Skin colors that data distribution seems to have revealed is yet to be confirmed which demands further analysis in future studies, and not the scope of this study. Even without additional manipulation process, however, as stated in the introduction, this is a matter of time and will eventually be solved as the research continues. As the dataset expands, the generator will be able to naturally learn what human figure is clearly in a purely unsupervised fashion. The generator's alleged failure in learning to precisely generate human without supervision is assumed to be highly due to the amount of information an editorial contains. Collected editorials in the dataset are processed to always have at least a partial of a human, but this does not mean that every editorial is

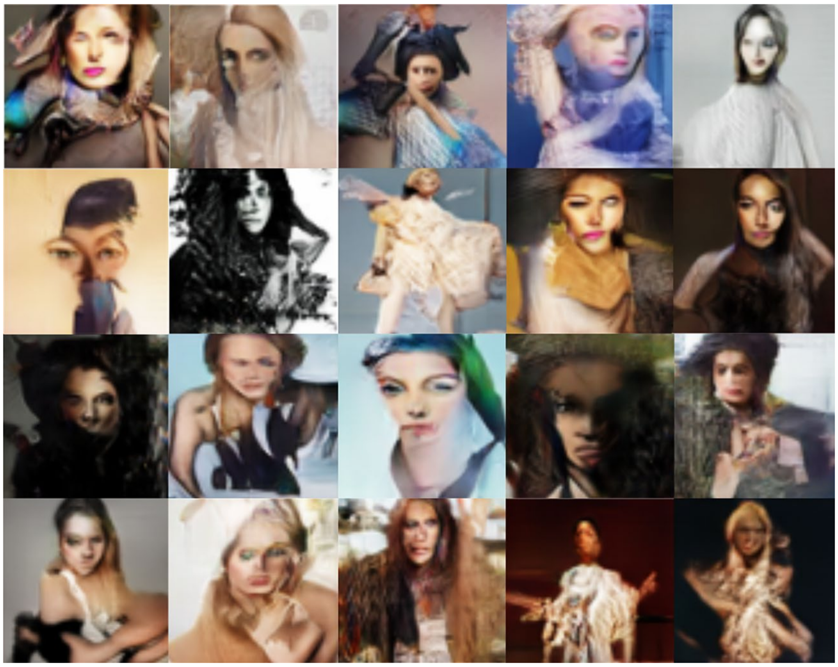
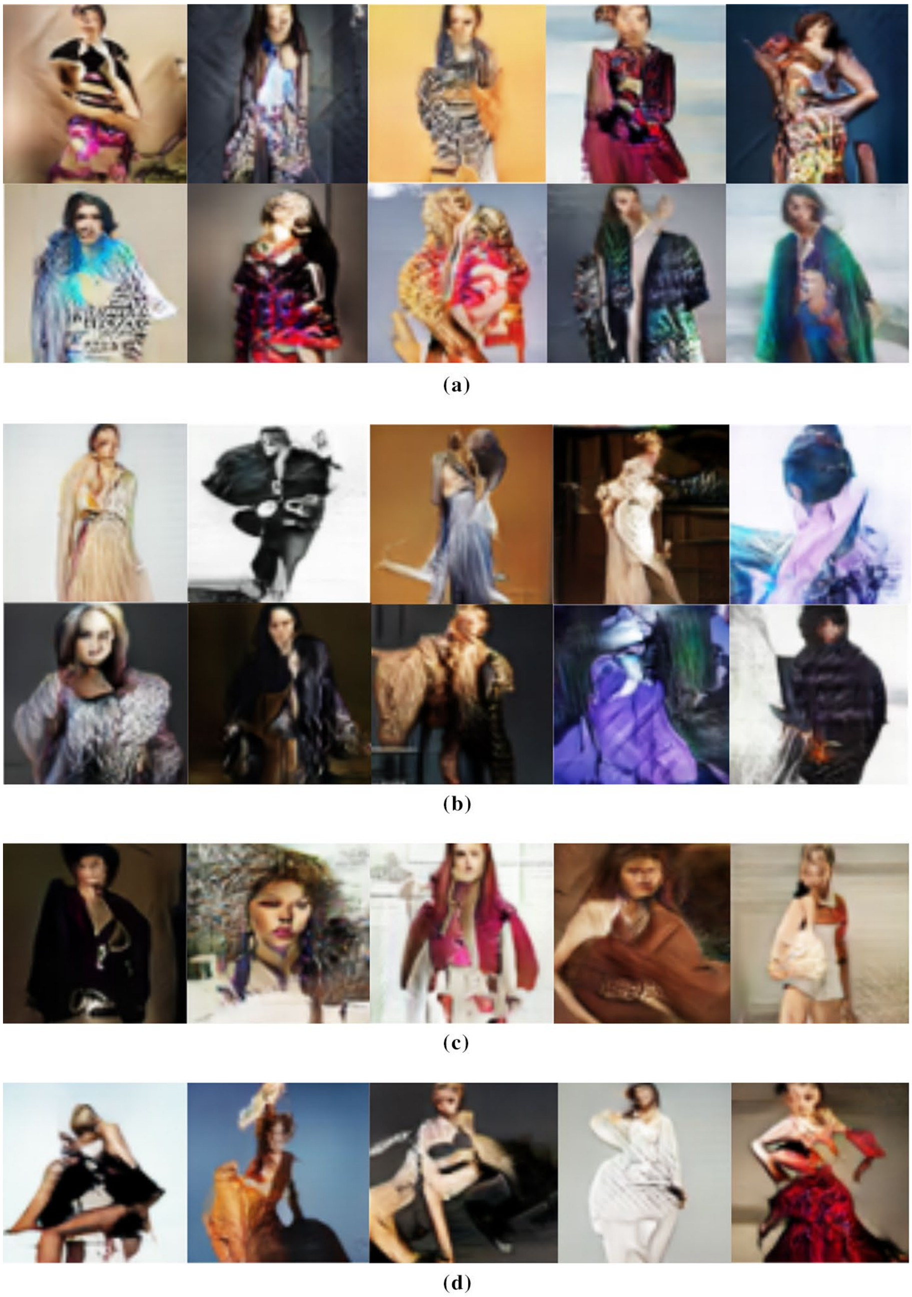


Fig. 9 Generated editorials

focused on a human figure. Not all editorials are shot in front of a solid-colored wall with its focus on a human model. They could be shot anywhere, indoor or outdoor, could display many an object or a prop, and a human figure(s) could be placed far-off, merely visible, in a very complex scenery. In terms of human figure, the generator must learn to draw their faces, hairstyles, make-ups, body construct and various postures with the learnt body construct. In terms of other non-humans, the generator has to learn to draw different fashion items such as hats, hairbands, scarfs, belts, shoes, bags or accessories, many different clothing shapes, fabric textures, patterns, color combinations, backgrounds from solid-colored studio to a very complex scene shot in a midst of a city, and visual aspects such as lightings that cast shadows, different shooting effects by camera like optic angle, focal, shutter speed, light exposition, and different photoshopped effects like noir, sepia, blur, smudge, dot and so on. In a small image of size 80 × 80, while it is difficult to distinguish objects clearly, there is too much to learn. So, with only the 60K dataset, a vague level of generation was to be expected. With larger dataset, more accurate and clearer editorials in high quality can be obtained. Furthermore, with improvements in hardwares and additional generative models for enlarging the image, bigger sized editorials can be used for the training and can be generated in higher resolution.

Even with the proposed editorial dataset of 60K images, however, the DCGAN has shown to have learned different aspects that editorials consist. It has learned to draw different types of clothing, creative patterns, color schemes, different materials like leather, fur, padding, and fabric drapes that are consistent with these materials (Fig. 10a, b). It further reveals to have learned to generate fashion items such as hat, hood, belt, or bags (Fig. 10c) and various postures human models hold as in non-portrait type editorials do (Fig. 10d).



(a)

(b)

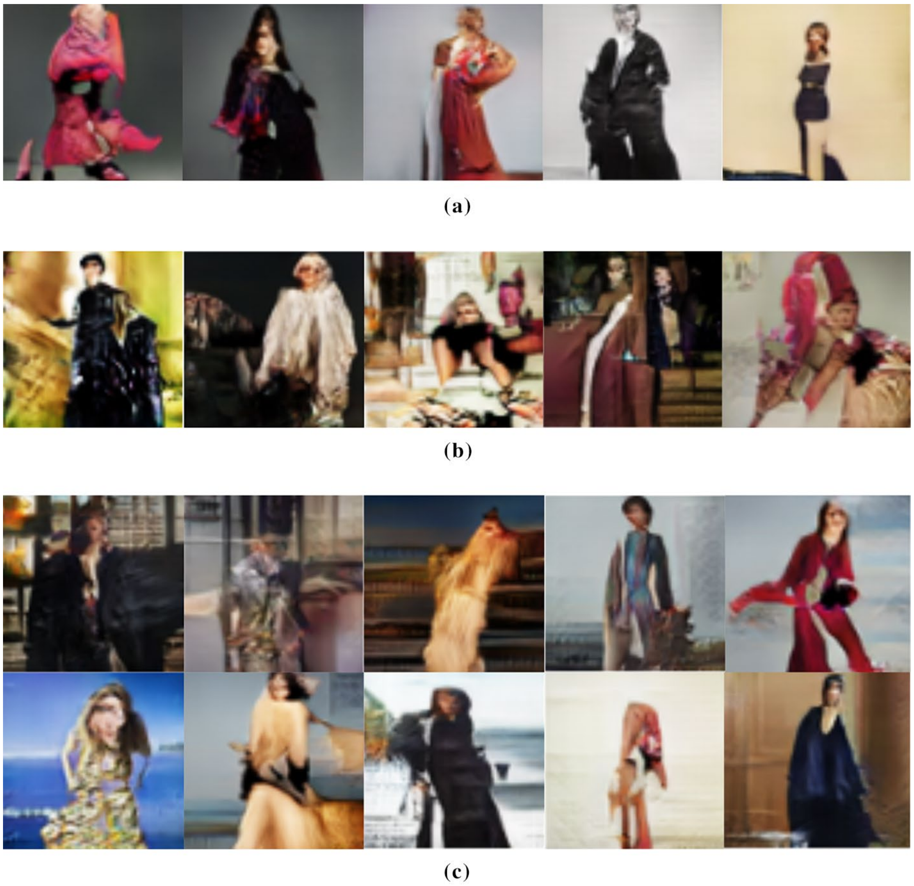
(c)

(d)

Fig. 10 Generated human figures and their apparels. **a** Designs and patterns. **b** Materials and drapes. **c** Fashion items. **d** Postures

Where fashion editorial is shot is not only limited to a photographer's studio. The results from the DCGAN shows consistency with this fact in generating backgrounds. Due to the low image resolution, results where human figure is drawn big and clear enough to be decipherable were mainly presented in this research, but for real editorials in the dataset contains data where human is very distant, the DCGAN has generated such images as well. Since the model has not yet reached a stage to contend that it successfully draws something for sure, especially in relation to the objects placed in the background, editorial results with a highly complex background where the existence of a human figure or the overall existence of anything could be controversial in a small image of 80 × 80 were not presented. Still, the DCGAN has shown to have learn to draw not only simple (Fig. 11a), but also complex backgrounds with or without props (Fig. 11b). Whether interior or exterior could be identifiable in some results (Fig. 11c). For interior backgrounds, corner of the walls was distinguishable through shadows, and wall decorations were drawn in some cases. For exterior backgrounds, separation between sky and ground was observable in some results, water-like surroundings such as an ocean or a lake, field-like surroundings, and natural objects like clouds, trees were examined.

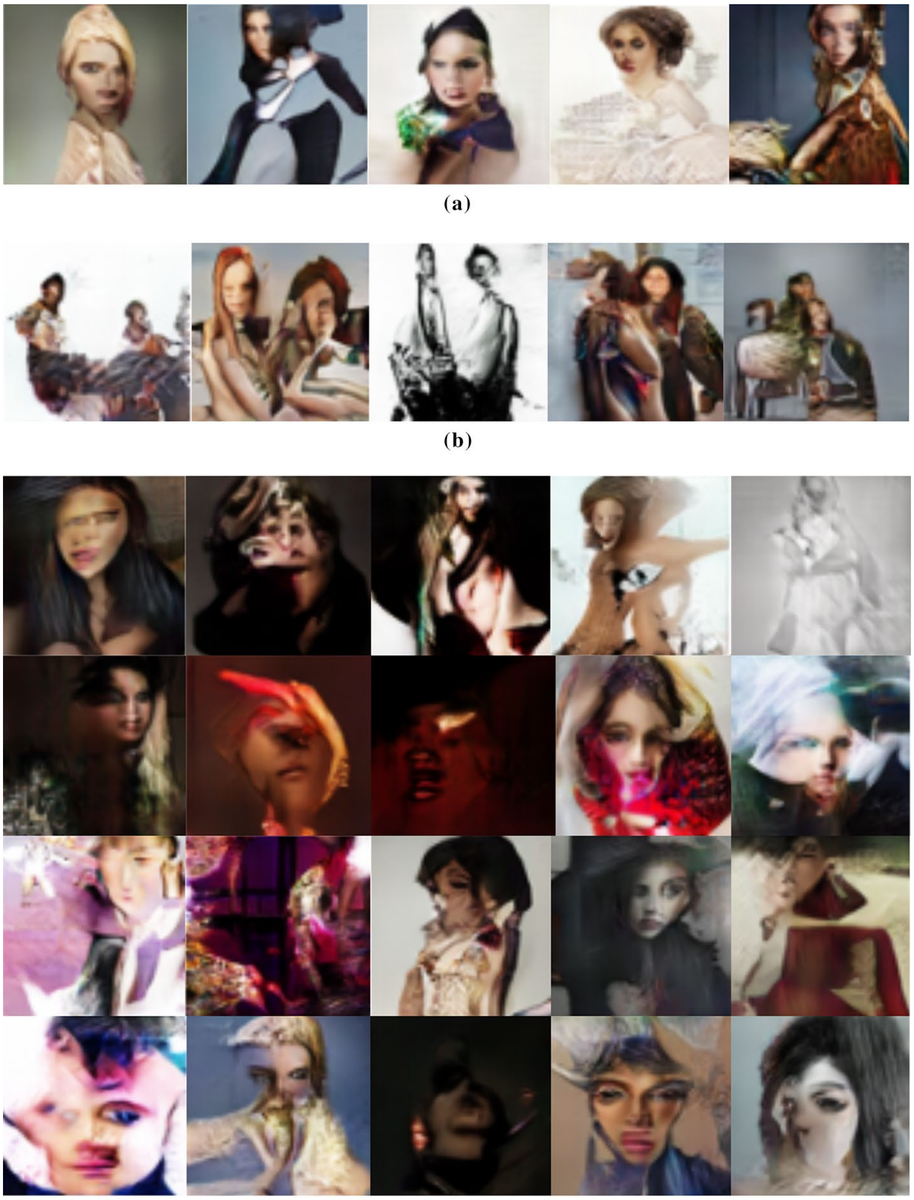
Other miscellaneous editorial results are shown in Fig. 12. In portrait type editorials, rather than staring straight into the camera front-faced, different face and bust angles could be examined (Fig. 12a). There were also results that could be seen as an endeavor to draw multiple human figures in a scene (Fig. 12b). Although ambiguous, what could be considered as different visual effects by either a camera or a photoshop such as low shutter speed which could capture movements of human or object, light and shade contrast, light exposures, grayscaling, blurring or paint-like image effects (Fig. 12c). Nonetheless, the model, DCGAN, trained upon the editorial dataset has learned to draw human figures in diverse context along with different apparel types, styles, designs,


(a)

(b)

(c)

Fig. 11 Generated backgrounds. **a** Solid studio wall. **b** Complex background. **c** Complex indoor or outdoor backgrounds



(a)

(b)

(c)

Fig. 12 Miscellaneous generated editorials. **a** None front-faced portrait-typed editorials. **b** Multiple human figures. **c** Visual effects

materials and so forth. All the exhibited editorial results confirm the validity of the proposed initial editorial dataset.

These results of the DCGAN suggest the proposed editorial dataset to be classified into sub-categories with standards. Dataset could be categorized into those that are revealed through visual inspection as in Figs. 10, 11, 12, or according to the results from further dataset analysis such as PCA (Principle Component Analysis). Once the editorials in the dataset are classified, multiple generative models could be trained to serve different purposes of generating different categories of editorials. The categorization could reduce editorials' scene complexity revealed by the graph in Fig. 6. Reduction in scene complexity would stabilize RGB distribution along the pixels and eventually improve dataset qualities and generative models' performances. Such classification of the dataset is also less costly than data labelling process, so this can be a next step to take for the dataset improvement.

Quantitative evaluations

Although there are no definite and absolute measures to quantitatively evaluate results from GAN, there are several methods that quantitatively assess quality of the generated results. In this study, FID (Frechet Inception Distance) was adopted to compare distribution of three models: GAN, DCGAN with a generator of 128 filters and the DCGAN with a generator of 256 filters (Table 1). Because both GAN and DCGAN with 128 generator filters showed unstable training and their results were visually poor consistently, comparison was made at the 500th-epoch rather than continuing further training. Although the scores are shown to be generally high in value, the inter-comparison has shown that the DCGAN with 256 generator filters has the lowest FID score, meaning it has the most similar distribution with the editorial dataset and generates the best quality results than the others. This verifies that the increase made in the number of generator filters did in fact improve results' qualities. The 1000th-epoch DCGAN model selected for the qualitative study is shown to have the lowest FID score of 200.3927 which indicates the model performs the best in the editorial generation. However, because given that FID scores are high, in-depth data analysis is necessary for the next version of the editorial dataset to further improve the dataset quality in future. As stated earlier, dataset classification or categorization using data distributions could be one of the solutions that can be applied in the next step.

Conclusions

In this study, a fashion editorial dataset of 60,000 images was created and validated using a DCGAN model. The DCGAN model, which can generate creative images in an unsupervised manner, was trained on the dataset without labels. The generator's filter was modified to 256 to enhance detail representation in editorial scenes, and this increase in filters demonstrated improved results which was verified by FID scores. The study acknowledges the dataset's limitations for unsupervised learning and further aims to expand it to improve the generative model's accuracy in depicting objects in scenes. Future studies include dataset categorization, larger editorial image generation, and providing labels through tagging for supervised learning and extending generative models such as diffusion model or VQ-VAE for diverse editorial creation purposes.

Table 1 FID scores of GAN, DCGAN of 128 generator filters, DCGAN of 256 generator filters

	Epoch 500	Epoch 1000
GAN	432.2005	–
DCGAN_g128	223.9395	–
DCGAN_g256	227.7720	200.3927

Abbreviations

AI	Artificial intelligence
DCGAN	Deep Convolutional Generative Adversarial Network
GAN	Generative Adversarial Network
LeakyReLU	Leaky Rectified Linear Unit
VQ-VAE	Vector Quantized Variational Autoencoder
FID	Frechet Inception Distance
PCA	Principle Component Analysis

Acknowledgements
Not applicable.

Authors' contributions
MK specified the research idea, developed and conducted experiments and drafted the manuscript. JK originated the idea of an artificial intelligence that generates fashion editorials. SK supervised and proofread the overall manuscript.

Funding
The authors declare that they have no competing interests.

Availability of data and materials
The editorial generating DCGAN trained for the study is available through nimooj/Editorial-DCGAN repository, https://github.com/nimooj/Editorial-DCGAN.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 31 May 2023 Accepted: 19 December 2023
Published online: 29 January 2024

References

An, H., Lee, K. Y., Choi, Y., & Park, M. (2023). Conceptual framework of hybrid style in fashion image datasets for machine learning. *Fashion and Textiles*. https://doi.org/10.1186/s40691-023-00338-8

Beaumont, R. (2022). LAION-5B: A new era of open large-scale multi-modal datasets. Retrieved May 19, 2023, from https://laion.ai/blog/laion-5b/

Brownlee, J. (2019). How to explore the GAN latent space when generating faces. Retrieved March 23, 2023, from https://machinelearningmastery.com/how-to-interpolate-and-perform-vector-arithmetic-with-faces-using-a-generative-adversarial-network/

Choi, W., Jang, S., Kim, H. Y., Lee, Y., Lee, S. G., Lee, H., & Park, S. J. (2023). Developing an AI-based automated fashion design system: reflecting the work process of fashion designers. *Fashion and Textiles*. https://doi.org/10.1186/s40691-023-00360-w

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems (NIPS 2014)*, 27. https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NIPS 2020)*, Vancouver, Canada, 33, 6840–6851. https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

Jetchev, N., & Bergmann, U. (2017). The Conditional Analogy GAN: Swapping Fashion Articles on People Images. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2287–2292. https://doi.org/10.48550/arXiv.1709.04695

Jones, K. (2017). GANGogh: Creating Art with GANs. Retrieved March 23, 2023, from https://towardsdatascience.com/gangogh-creating-art-with-gans-8d087d8f74a1

Kumar, S., & Gupta, M. D. (2019). *c+GAN: Complementary Fashion Item Recommendation*. Preprint retrieved from https://doi.org/10.48550/arXiv.1906.05596

Lang, Y., He, Y., Dong, J., Yang, F., & Xue, H. (2020, May 4–8). Design-Gan: Cross-Category Fashion Translation Driven By Landmark Attention. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 1968–1972. https://doi.org/10.1109/ICASSP40776.2020.9053880

Lee, H., & Lee, S. G. (2019). Fashion Attribute-to-Image Synthesis using Attention-based Generative Adversarial Network. *IEEE Winter Conference on Applications of Computer Visions (WACV)*, 462–470. https://doi.org/10.1109/WACV.2019.00055

Lin, C. Z., Lindell, D. B., Chan, E. R., & Wetzstein, G. (2022). *3D GAN Inversion for Controllable Portrait Image Animation*. Preprint retrieved from https://doi.org/10.48550/arXiv.2203.13441

Liu, Y., Chen, W., Liu, L., & Lew, M. S. (2019). SwapGAN: A multistage generative approach for person-to-person fashion style transfer. *IEEE Transactions on Multimedia, 21*(9), 2209. https://doi.org/10.1109/TMM.2019.2897897

Lin, J., Song, X., Gan, T., Yao, Y., Liu, W., & Nie, L. (2021). PaintNet: A shape-constrained generative framework for generating clothing from fashion model. *Multimedia Tools and Applications, 80*, 17183–17203. https://doi.org/10.1007/s11042-020-09009-y

Marriott, R. T., Romdhani, S., & Chen, L. (2021, June). A 3D GAN for Improved Large-pose Facial Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13445–13455. Preprint retrieved from https://doi.org/10.48550/arXiv.2012.10545

Pandey, N., & Savakis, A. (2020). Poly-GAN: Multi-conditioned GAN for fashion synthesis. *Neurocomputing, 414*, 356–364. https://doi.org/10.1016/j.neucom.2020.07.092

Pernus, M., Fookes, C., Struc, V., & Dobrisek, S. (2023). *FICE: Text-Conditioned Fashion Image Editing With Guided GAN Inversion*. Preprint retrieved from https://doi.org/10.48550/arXiv.2301.02110

Ping, Q., Wu, B., Ding, W., & Yuan, J. (2019). Fashion-AttGAN: Attribute-Aware Fashion Editing with Multi-Objective GAN. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. https://openaccess.thecvf.com/content_CVPRW_2019/papers/FFSS-USAD/Ping_Fashion-AttGAN_Attribute-Aware_Fashion_Editing_With_Multi-Objective_GAN_CVPRW_2019_paper.pdf

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *4th International Conference on Learning Representations (ICLR) 2016*, San Juna, Puerto Rico. Preprint retrieved from https://doi.org/10.48550/arXiv.1511.06434

Robbiebarrat. (2017). *art-DCGAN*. Retrieved March 23, 2023, from https://github.com/robbiebarrat/art-DCGAN

Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., & Pal, C. (2018). *Fashion-Gen: The Generative Fashion Dataset Challenge*. Preprint retrieved from https://doi.org/10.48550/arXiv.1806.08317

Sohl-Diskstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning (PMLR), 37*, 2256–2265. https://doi.org/10.48550/arXiv.1503.03585

Van Den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). Neural Discrete Representation Learning. *Advances in neural information processing systems (NIPS 2017), 30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/7a98af17e63a0ac09ce2e96d03992fbc-Paper.pdf

Williams, V. (2008). A heady relationship: fashion photography and the museum, 1979 to the present. *Fashion Theory, 12*(2), 197–218. https://doi.org/10.2752/175174108X299998

Wu, J., Zhang, C., Xue, T., Freeman, W. T., & Tenenbaum, J. B. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. *Advances in Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.1610.07584

Xiao, H., Rasul, K., & Vollgraf, R. (2017). *Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms*. Preprint retrieved from https://doi.org/10.48550/arXiv.1708.07747

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Minjoo Kang holds bachelor's degree in both Fashion and Textiles and Computer Science and Engineering, Seoul National University. This enabled specifying the research idea and conducting the research. She is currently a Ph.D. Candidate at the Fashion and Textiles Department, Seoul National University.

Jongsun Kim also is professor of Fashion Design Department, Suwon Women's University.

Sungmin Kim Professor, Research Institute of Human Ecology, Seoul National University, Professor, Research Institute of Human Ecology, Seoul National University.