Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Object Detection Using Convolutional Neural Network To Identify Popular Fashion Product

To cite this article: Andry Alamsyah *et al* 2019 *J. Phys.: Conf. Ser.* **1192** 012040

Object Detection Using Convolutional Neural Network To Identify Popular Fashion Product

Andry Alamsyah[1], Muhammad Apriandito Arya Saputra[2], Riefvan Achmad Masrury[3]
School of Economics and Business, Telkom University, Bandung, Indonesia

E-mail : [1] andrya@telkomuniversity.ac.id,
[2] m.apriandito@gmail.com and [3] riefvan@gmail.com

Abstract. Up to 95 million of photos are uploaded every day to Instagram as the biggest photos-sharing platform in the world. People share photos of their daily activities, hobbies, opinion and outfits in their Instagram post. This phenomenon opened a new visual discovery opportunity where unwritten information can be extracted using computer vision technology. One can extract the insight using Convolutional Neural Network (CNN) by applying object Detection process to recognise object in an image as a part of quantification process and knowledge extraction. In this paper, we re-train the CNN model using top-head fashion accessories dataset, represented by veil, eyeglasses and hat, to recognise the use of those item in each Instagram Image posts gathered as a dataset from 3 representative cities in Indonesia to identify the most popular accessories used in each representative city. The benefit of this model is the ability to identify popular top-head fashion accessories used in a particular area automatically, utilize it as basis information for company in fashion industry to better understand market, thus, increase the accuracy in decision making.

1. Introduction

The use of social media with penetration of 37% of the total Internet users in the world [1] causes a considerable amount of interaction and directly generates huge amount of data. Interaction data from social media contains information that can be extracted using certain methods depending on the type of data [2]. Photos and videos are examples of unstructured data on social media. Instagram is the largest photo-sharing based platform in the world with an average of 95 million photos uploaded every day and more than 40 billion photos uploaded since it was launched in 2010 [2]. This huge amount of Instagram data is certainly being a source of opportunities to implement computer vision technology to extract insights from Instagram post. Insights are expected to be converted to actionable information in order to solve daily problems, especially in business matters.

Unstructured data requires special treatment, since the information was contained in other forms. In photos or videos, the information is contained in the form of visual imagery which is the object in the image. With computer vision technology, computers learn to identify objects in images. *Convolutional Neural Network* (CNN / CovNet) is an object detection model in computer vision study. CNN is a neural networks application that use machine learning algorithms which make it can learn independently from the given dataset and identify specific objects in an image. The results of identification process later be quantified and can be used as a basis information for market research tasks.

Fashion Industry is one of the important industries today. However, market trend that always change make this industry filled with uncertainty [5]. By gathering the market information, company will be able to understand the market condition, reduce the business uncertainty, also assists the company to making better decision and strategy. In this paper, we extract the insight from Instagram post data by applying CNN model with intention to identify the popular top-head fashion accessories in a particular area. We selected 3 cities as the representatives by looking for the most interactive cities in each time zones in Indonesia, which is Jakarta, Bali and Jayapura. We focus on recognize top-head fashion accessories product which in this paper represented by Veil, Eyeglasses, and Hat based on Instagram photos uploaded using specific fashion hashtags for each city. The purpose of this model is to the result of this model is the total number of each item detected in Instagram photos grouped by each city. This result of detection later be quantified and calculated to determine the popular top-head accessories in these 3 cities.

2. Related Work

2.1. Machine learning

Machine Learning is a methodology which concerned with the design and development of algorithms that allow computers automates analytical model building based on empirical data, from sensor data or databases. Machine Learning uses statistics in building mathematical model [3]. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the difficulty lies in the fact that the set of all possible behaviours given all possible inputs is too large to be covered by the set of observed examples [4]. The convolutional neural network model that we use as object-detection models uses machine learning algorithms.

2.2. Convolutional neural network (CNN/CovNet)

Convolutional Neural Network is one of deep neural network class. CNN was inspired by biological process, the connectivity pattern between the neuron resemble the organization of animal visual cortex [5]. CNN are very similar to ordinary Neural Network that made up of neurons that have learnable weights and biases. CNN consist of an input layer, an output layer, and many hidden layers between input and output as shown in figure 1. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. CNN have two major part which is feature extraction and fully connected layers. Feature extraction part was a repetition of three common layers which is convolution layer, activation or Rel-u, and pooling layer.

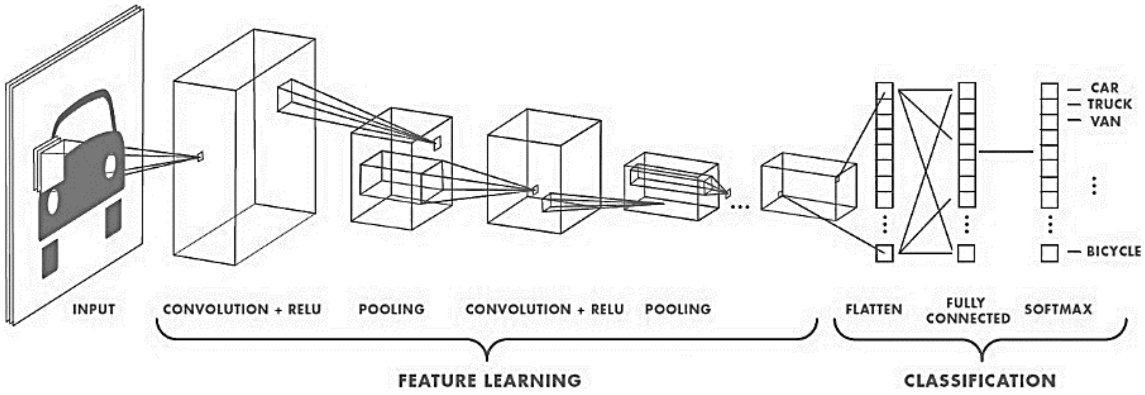

Figure 1. CNN architecture [7].

CNN is able to have tens or hundreds layer which every layer train to detect different features from an image. The Filter applied to every train image with different resolution, and the output from every image that convolved used as input to the next layer. The final layer of the CNN architecture uses a classification layer which is fully connected layer to provide the classification output [7]. CNN can be applied in image and video recognition, recommender systems and natural language processing. [6]

2.2.1. Inception architecture

Inception architecture is the modification of feature extraction in Convolutional Neural Network architecture. The difference lies in the feature extraction section. In inception v2, the feature extraction section uses filter concat and base layer as in the figure 2. The characteristic of Inception architecture is the improved utilization of the computing resources inside the network. The main advantage of Inception Architecture is it is having significant quality gain at a modest increase of computational requirements compared to shallower and less wide networks and have competitive despite of neither utilizing context nor performing bounding box regression [8]. Inception architecture premise to reduce the bottleneck and improve efficiency in term of computational complexity using smart factorization methods. In Inception v2 architecture the 5x5 pixel convolution layer was factorized to 3x3 pixel convolution to improve computational speed shown in Figure 2 [9].

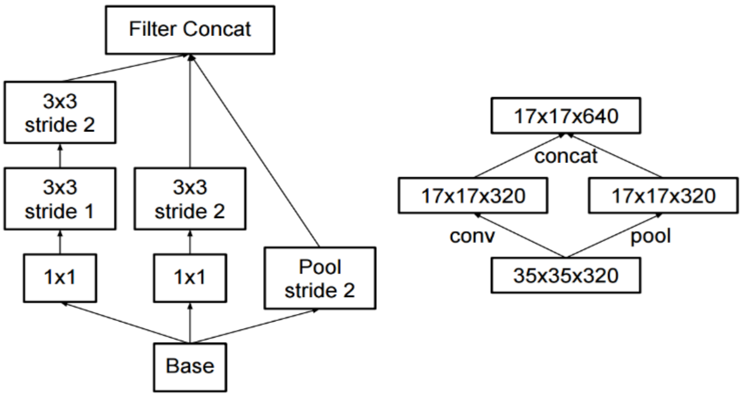

Figure 2. Inception v2 architecture [8].

2.3. Object detection

Object Detection is a computer technology that related to computer vision and image processing to recognize semantic object with a specific class such as Human, Animal, Cars, etc. in an image or video. The object detection determines the location and size of object detected [10].

3. Methodology

3.1. Research framework

The research objective of this paper is to construct a detection model to detect top-head fashion accessories which represented by eyeglasses, veil, and hat then apply it to an Instagram image post dataset that gathered from 3 representative cities in Indonesia. The result of detection in each city will be classified and quantified to present the information. In this paper, we categorize the overall process to 4 main process which is Data Gathering Process, Filtering Process, Modelling Process, detection and Validation process as showed in Figure 3. The cream parallelogram marks for the image dataset and the white box marks for a process.

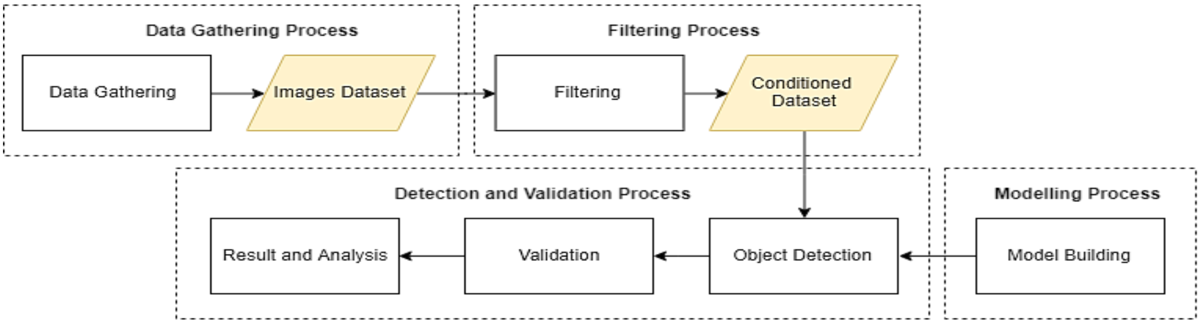

Figure 3. Research framework.

3.2. Data gathering

The gathering process is a process to gather Instagram post image data from the representative cities. We select the representative cities by looking for the most used city hashtag *#namecityhashtag* to find the most interactive cities in 3 time zones in Indonesia as. The number of city hashtag use is in Table 1. We used the time zone as a divider to cover all Indonesia area and to reduce bias of a specific region. We did not use the number of populations, since the large number of populations did not always represent larger number of Instagram interaction.

Table 1. The most used city hashtag in Instagram.

Indonesia Western Time			***Indonesia Central Time***			***Indonesia Eastern Time***		
No.	*City*	*Hashtag Use*	*No.*	*City*	*Hashtag Use*	*N o.*	*City*	*Hashtag Use*
1.	*Jakarta*	45.887.040	1.	*Bali*	41.769.101	1.	Jayapura	1.318.679
2.	*Bandung*	32.764.570	2.	*Makassar*	7.315.042	2.	Ambon	1.309.724
3.	*Surabaya*	21.551.147	3.	*Banjarmasin*	5.120.770	3.	Ternate	696.768
4.	*Medan*	15.495.040	4.	*Samarinda*	4.347.016	4.	Merauke	256.2016
5.	*Solo*	15.200.851	5.	*Manado*	4.347.016	5.	Timika	240.207

The post image data are gathered from Instagram using Instagram Application Programming Interface (API). These image data were collected from 29th June 2018 to 5th August 2018 using specific search keyword in hashtag form with intention to show fashion in cities dated in year 2017 and 2018. The number of total gathered image showed in Table 2.

Table 2. Total gathered Instagram post image.

City	*List of Hashtag*	*Total Images Gathered*
Jakarta	#OOTDJAKARTA #JAKARTAHITS #JAKARTAFASHION	52.931
Bali	#OOTDBALI #BALIHITS #BALIFASHION	32.422
Jayapura	#OOTDJAYAPURA #FASHIONJAYAPURA #JAYAPURAHITS	11.343

Number of data was limited by Instagram through its API by setting a rate limit to only allows 200 calls per hour [11]. The image data can only be gathered from accounts with profile set to public. Deleted or restricted post also cannot be gathered.

3.3. Filtering

The objective of this process is to eliminate images with non-human figure from thousands of images in dataset. This process done by re-trained the inception v2 model to identify whether a face and a body appears in an image and will automatically eliminate images without these objects present. The example of detection area is in Figure 5. The automatic filtering enables a faster and efficient process when dealing with a large dataset. This process aims to get a conditioned dataset suitable for achieving higher accuracy output in identification and annotation process. The number of conditioned datasets used in detection process shoed in Table 3.

Table 3. Conditioned dataset.

City	*Total Image Gathered*	*Total Image Contains Face and Body*
Jakarta	52.931	25.078
Bali	32.422	12.523
Jayapura	11.343	5.319

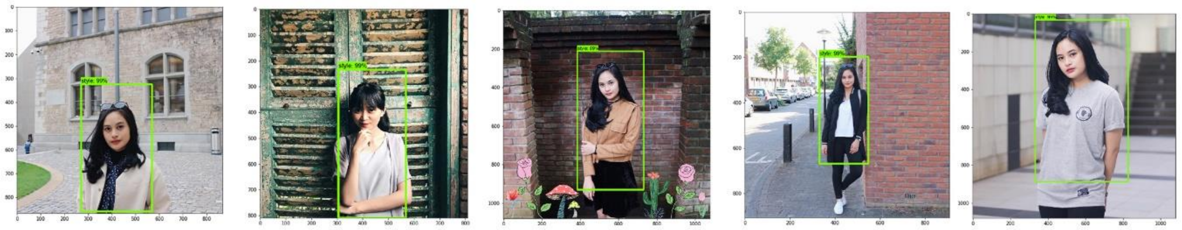

Figure 4. Detection area.

3.4. Model Building

To build the model, we need to have an images dataset that contains all the annotation class that will be used as a learning dataset. In this paper we use 3 classes (Eyeglasses, Veil, and Hat). Every class consists of 500 train Images and 250 test Images. This dataset was chosen randomly from Google Images, and should contain some variability such as *Angle, Lighting, Model, and Distance*. Every row data in Training dataset is labelled accordingly shown in Figure 6. The image labelling was constructed in XML-like structure. Every image will have an object element named "name" as the image label itself and several other elements associated with an object. Labelling process will also add another sub element information regarding the object position in an image inside bndbox element.

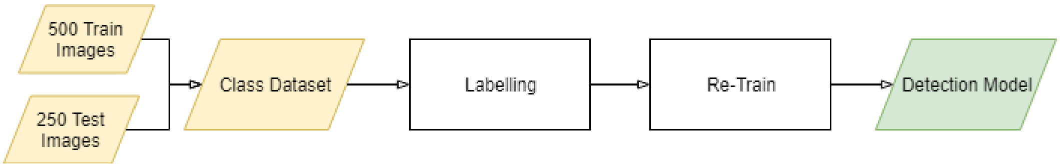

Figure 5. Model building framework.

We used pre-trained model of *Faster RCNN Inception v2* that trained with a COCO dataset. Re-train process or transfer learning process was a process to re-teach the model with new dataset and re-customized the model for identifying top head fashion accessories specifically. The retrain process follows the *TensorFlow* framework. During the re-training process, TensorFlow library functions will give the number of step and the total loss for each step. The number of total loss can be used for performance measurement of the model.

4. Result and analysis

To ensure the model that we make having a high level of accuracy, we took samples of 500 images for each city, then conduct performance test using a confusion matrix showed in table 4. Based on these performance test in table 5, the model accuracy is 0,9474. In this paper we do not aim to find for models with the highest accuracy. We set the accuracy threshold above 85%. we consider a model above those value is considered good enough to recognize top-head fashion accessories.

Table 4. Confusion Matrix.

	True Positive	*True Negative*
Predictive Positive	1407	63
Predictive Negative	95	1437

Table 5. Model performance table.

Accuracy	*Sensitivity*	*Specificity*	*Precision*
0.9474	0.9368	0.9580	0.9571

In the last stage, we applied the model to the entire dataset in order to discover the use of top-head fashion accessories product in 3 cities. Several results of detection shown in figure 9, the model put a square bracket on top of detected objects.

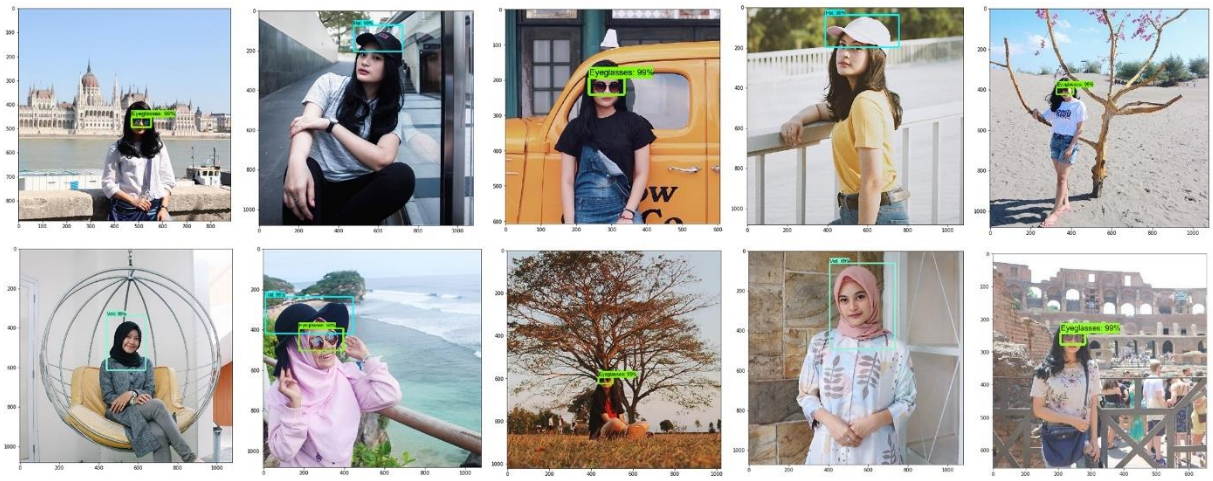

Figure 6. Top-head fashion accessories detection.

We also determine the market penetration value for each product by measuring the number of products identified by the model in a city as a representation of the number of people uses the products. A percentage of people using the product is total number of detected items divided by total images contains human figure for each city. Based on data in Table 6 and Table 7, Veil is much more popular in Jakarta compared to other cities. Hat and Eyeglasses are receiving similar appreciation in Jakarta, Bali, and Jayapura.

Table 6. The number of top-head fashion accessories used in 3 representative cities.

City	*Eyeglasses Only*	*Veil Only*	*Hat Only*	*Multiclass*
Jakarta	2812	5862	4511	3368
Bali	1537	718	1959	749
Jayapura	610	359	954	242

Table 7. Market penetration rate table.

City	*Eyeglasses Penetration*	*Veil Penetration*	*Hat Penetration*
Jakarta	11,2%	23.3%	17.5%
Bali	12,2%	5,7%	15.6%
Jayapura	11,4%	6,7%	17.9%

5. Conclusion

An object-identification models by re-training the Convolutional Neural Networks models can be used to identify the use of top-head fashion accessories in Indonesia's 3 major cities. This quantification result provides basis knowledge in business decision making such as determine market trend in certain area, product penetration, and identify popular product. From the validation test, our model has an accuracy value of 0.9897. The accuracy level is influenced by architecture used, data size and training dataset. The accuracy level can be improved by re-training the models using larger datasets or involving more image attributes. This paper is our first step in information extraction research from social media images. Further research will cover larger data, involving wide array of attributes to detect more fashion items, having better image detection models and incorporating time-series analysis for patterns finding.

References

[1] Kemp S 2017 *Digital in 2017: Global overview*, Retrieved from : https://wearesocial.com/special-reports/digital-in-2017-global-overview, Accessed : September 2018

[2] Alamsyah A, Paryasto M, Putra J and Himmawan R 2016 Network Text Analysis to summarize online conversations for marketing intelligence efforts in telecommunication industry *Proc. International Conference on Information and Communication Technology (Bandung)* pp 1-5 (Indonesia: IEEE)

[3] Alpayadin E 2010 *Introduction to Machine Learning : Second Edition* (Cambridge : Massachuttes Institute of Technology). 978-0-262-01243-0

[4] Matzen K, Bala K, and Snavely N 2017 *StreetStyle: Exploring world-wide clothing styles from millions of photos* (New York) arXiv:1706.01869v1

[5] BoF-McKinsey Global 2018 *The State of Fashion 2018 The Business of Fashion* McKinsey & Company

[6] *Convolutional neural network*, Retrieved from https://en.wikipedia.org/wiki/Convolutional-neural-network , Accessed : September 2018

[7] *Convolutional Neural Network* Retrieved From : https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html , Accessed : September 2018

[8] Szegedy C, Liu w, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2014 *Going deeper with convolutions* (Chapel Hill) arXiv:1409.4842v1

[9] Szegedy C, Vanhoucke V, Loffe S and Shlens J 2015 *Rethinking the Inception Architecture for Computer Vision* (London) arXiv:1512.00567v3

[10] Tebaldi, M. 2010. *Designing a Labeling Application for Image Object Detection . Master of Science in Computer Science and Engineering* (Padua) Universita degli Studi di Padova.

[11] *API graf Instagram* Retrieved from : https://developers.facebook.com/docs/instagram-api/overview/ Accessed : September 2018