**Fashion Conversation Data on Instagram**

Welcome to our data project page. Our fashion dataset is composed of information about 24,752 posts by 13,350 people on Instagram. The data collection was done over a month period in January, 2015. We searched for posts mentioning 48 internationally renowned fashion brand names as hashtag, which are listed below.

**Studied brands**

|  |  |
| --- | --- |
| Brand category | Brand name |
| Mega couture | burberry, cartier, chanel, gucci, hermes, louisvuitton, prada, tiffany |
| Small couture | brioni, brunellocucinelli, cesareattolini, ermenegildozegna, fabianafilippi, kiton, loropiana, nancygonzalez |
| Designer | acnestudios, alexandermcqueen, alexanderwang, coach, dvf, iro, isabelmarant, katespade, maisonmargiela, marcbymarcjacobs, marcjacobs, michaelkors, paulsmith, rebeccaminkoff, sandro, stellamccartney, theory, vince, viviennewestwood |
| High street | abercrombie, americanapparel, americaneagle, calvinklein, forever21, gap, hollister, jcrew, mango, topshop, uniqlo, urbanoutfitters, zara |

Our data contain information about hashtags as well as image features based on deep learning (Convolutional Neural Network or CNN). The list of learned features include selfies, body snaps, marketing shots, non-fashion, faces, etc. Please refer to our paper for the full description of how we built our deep learning model.

**Papers**

* Yu-I Ha, Sejeong Kwon, Meeyoung Cha, Jungseock Joo, Fashion Conversation Data on Instagram, In proc. of the International AAAI Conference on Web and Social Media (ICWSM), May 2017  
  <https://arxiv.org/abs/1704.04137>

**Data Characteristics**

Below describe the data fields, followed by an example. Please note that, in concern for privacy, we anonymized any mentions of user screen names that appear less than 10 times as @00000 (instead of @username). We also do not share the full image content, but give the image URL in case they are public posts.

*1. Basic information about posting user and each post.*

•        User Id: Numeric ID of the posting user

•        Followings: The number other users a given uploader is following (i.e., distinct sources of the uploader)

•        Followers: The number of other users subscribing to the uploader’s account (i.e., fans of the uploader)

•        Media count: The number of total posts (both fashion and non-fashion) contributed by the uploader

•        Brand name: A brand name used in fashion post search process, used as a hashtag in user’s post

•        Brand category: Grouping of a brand

•        Hashtags: The list of hashtags in a post

•        Caption: Text description of a post provided by the user, excluding the hashtag information

•        Image URL: A web link to the image file of the post

•        Likes: The total number of likes per post as well as the list of Instagram audience who liked each photo  (including their user IDs, names, and profile URLs)

•        Comments: The total number of comments as well as the list of Instagram audience who commented on the photo

•        Creation Time: When post was uploaded on Instagram

•        Link: A web URL (if any) contained in each post

*2. The visual content variables and descriptions.*

\* Selfie ~ Outdoor = probability, Number of people and fashion product = count

•        Selfie: Images whose face occupies more than 50% of height

•        Body snap: Images including full or partial body parts and not a selfie

•        Marketing: Editorialized images like runways, awards ceremonies

•        Product-only: Images without face or body parts that contain fashion products

•        Non-fashion: Images not related to fashion such as food, landscape, and animals

•        Face: Images containing frontal or side faces

•        Logo: Images with unmatched brand logo or text

•        Brand logo: Images with the matching logo of the brand in hashtags

•        Smile: Images containing one or more smiling face(s)

•        Outdoor: Images whose background is outdoor

•        People: The number of people detected in the image via body shapes

•        Items: The number of fashion items estimated in the image

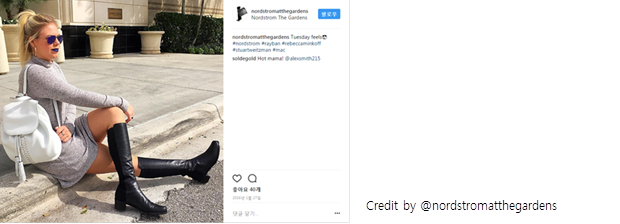
*3. Emotion features*

\* Microsoft emotion API

<https://docs.microsoft.com/ko-kr/azure/cognitive-services/emotion/home>

\* The highest score should be considered as the dominant emotion of the face in the image. For images contained multiple faces, the average value of each emotion was labeled.

**Example data**

****

Field name:

UserId, Followings, Followers, MediaCount, BrandName, BrandCategory, Hashtags, Caption, ImgURL, Likes, Comments, CreationTime, Link, Selfie, BodySnap, Marketing, ProductOnly, NonFashion, Face      Logo, BrandLogo,       Smile,        Outdoor, NumberOfPeople, NumberOfFashionProduct, Anger, Contempt, Disgust, Fear, Happiness, Neutral,        Sadness,        Surprise

Data:

1171190842177271618\_181427732  385, 923, 1123, rebeccaminkoff, Designer, rayban, mac, nordstrom, stuartweitzman, rebeccaminkoff, Tuesday feels\U0001f60e #nordstrom #rayban #rebeccaminkoff #stuartweitzman #mac, https://scontent.cdninstagram.com/t51.2885-15/s640x640/sh0.08/e35/12407295\_933273556750488\_701430114\_n.jpg, 37, 1,        1.45E+09, https://www.instagram.com/p/BBA592gKrdC/, 0.005245, 0.993443, 0.01345, 0.002188, 0.001692, 0.821065, 0.115222, 0.029139, 0.243963, 0.85901, 1.15739, 5.16747, 0.003341, 0.005292, 0.00227, 2.62E-05 0.060212, 0.922894, 0.001611, 0.004352