AI detection of how to manipulate selfies in Instagram: Comparing USA vs. South Korea

Su Jin Yang & Seungwoo Seo

Associate Professor, Department of Living Culture and Consumer Science, Sungshin Women's University
Master's Course, Department of Applied Artificial Intelligence, Korea University Graduate School of Computer & Information Technology

Instagram has been not only famous for its popularity for image-oriented posts, but also notorious for harmful psychological problems like eating disorders (e.g., Lee-Won, et al., 2020; Mills, et al., 2018). Most Instagram users try to build their own self-image through photos of their lifestyle, with dream of becoming influencers that hundreds of people "follow" on social media (Marwick, 2015), as if they were celebrities. However, compared to TV or movie celebrities, people consider influencers to be normal people and feel closer to them (Martensen, Brockenhuus-Schack, & Zahid, 2018). Thus, the influence of Instagram influencers on general users can be even more significant.

Through Instagram, users want to spread their body images out worldwide with expecting to gain the more 'heart' from related and even totally strange users. Selfies means self-taken photos, which are commonly executed and posted as a self-presentation strategy on Instagram. Users updating their own images tend to refer to the other users' postings, which may contribute to formation of self- body image. The critical issue arises here, since most of users in Instagram, especially influencers with vast number of followers, filtering, retouching, and manipulating selfies with following conventional feminine beauty norms of young and slender females (Lowe-Calverley & Grieve, 2020). Thus, viewing Instagram images can make women feel dissatisfied with their bodies, which leads to a worse mood or even depression (Tiggemann & Barbato, 2018). Specifically, a group of people exposed to retouched and manipulated selfies turned out to perceive themselves less attractive than the others exposed to natural selfies (Kleemans, et al.,2018). On the other hand, Instagram users posting natural selfies felt more anxious than the others posting retouched selfies (Mills, et al., 2018). Several research has already suggested that this kind of obsessed behavior toward manipulation of selfies may cause various problems including depression (Mills, et al., 2018) and eating disorders (Lee-Won, et al., 2020), which finally result in lowering quality of life (Abbas, & Dodeen, 2022).

However, to our knowledge, research concerning the Instagram users' selfie related behaviors tend to adopt self-reporting questionnaires asking the degree to utilize manipulating filters or applications (e.g., Fox & Vendemia, 2016; Lee-Won, et al., 2020). Rather, they focused psychological reaction toward retouched selfies within experimental designs (e.g., Mills, et al., 2018; Tiggemann & Barbato, 2018), so specific manipulating behaviors failed to be captured. Thus, the current research attempted to adopt an AI detecting programs to get detailed information on users' manipulating behavior of selfies. This automatic approach detecting manipulating behavior is meaningful for this kind of social norm issues when
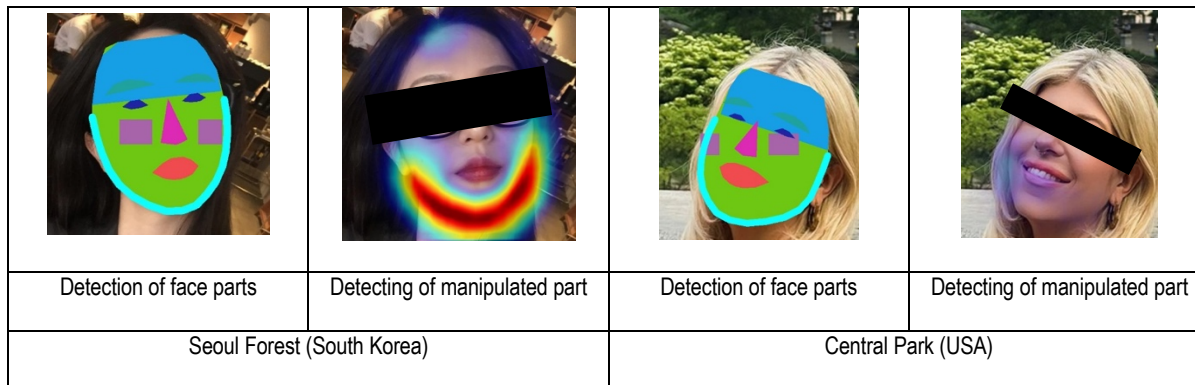
considering that retouching selfies is a kind of lying so socially less desirable. Many researchers in social science argued that social desirability bias generally happen when adopting self-reporting questionnaire toward widely accepted attitudes or behaviors (e.g., smoking, cheating, etc.), since participant tend to answer more socially acceptable way rather than reflect their actual perception and behavior (Grimm, 2010; Van de Mortel, 2008). On the other hand, newly introduced filters and retouching application are most likely high technology enough for users to select the most attractive looks without considering and manipulating parts of face. Thus, there are considerable chances where users may not recognize which part of face retouched and manipulated specifically. Under these circumstances, AI detection seems the better option to adopt to investigate users' manipulating behaviors of selfies.

Image warping filter applications that were originally for digital cameras are now available for smartphones. Before the era of application, Adobe Photoshop is one of the representative tools for expert to utilize for enhancing the quality of their digital pictures. Wang and colleagues (2019) published a paper and share the code of this AI detector of selfie manipulation to the open source of artificial intelligence (AI) with benefits and technical help of Adobe Photoshop. With the AI codes, we can also easily recognize warped face parts and the level of manipulation and even reconstruct the original images. As shown in [Fig 1], the AI codes can visualize the warped area and the level of manipulation represented by the change in color from red to green. These warped areas and level can be detected by using the AI program provided by Wang et al. (2019).

This study seeks to employ this novel methodology to measure the manipulation behavior of Instagram users, specifically which facial features they wish to alter in their own selfies. The social sciences have adopted this AI technology that involves crawling, detecting, and categorizing data. Due to the attractiveness of the open-source culture in computer science, researchers from other fields are able to pursue cutting-edge technologies and adapt them to their own fields of study. With the data that we gathered from the AI detector and Instagram, we investigated how selfie retouching behaviors vary by culture and number of 'likes'. In 2019, South Korea had the highest smartphone adoption rate (95%) (Pew research, 2019). The remaining 5% utilized mobile phones but not smartphones. In contrast, the United States had a smartphone adoption rate of 81%. South Korea had the second-highest SNS adoption rate, at 89.3%, while the United States had a rate of 72.3%.

This research applied the Selenium library to crawl recent Instagram posts on May 24, 2023. Users of Instagram have control over the visibility of their posts, from private to public, as well as their location data. To avoid privacy concerns, we restricted data crawling results to recent posts with complete public visibility and manually entered location information, such as 'Seoul Forest' for South Korea and 'Central Park' for the United States. Due to Instagram's policy of associating each location with a unique ID, we could search for posts containing specific location information during data extraction. Specifically, we only collected the first image posted, location data, and the number of likes. Initially, 843 images for the two locations were examined, and using the face detection AI model by Wang et al. (2019), only photographs

suitable for recognizing facial features were filtered out. Upon completion of the filtering procedure, we had 196 for South Korea and 231 for the United States.

Detection of face parts	Detecting of manipulated part	Detection of face parts	Detecting of manipulated part
Seoul Forest (South Korea)		Central Park (USA)	

[Fig. 1] An example of AI detecting analysis.

Using the open-source codes by Wang et al. (2019), we were able to detect the areas of eight facial regions (e.g., forehead, eyebrow, eye, cheek, nose, mouth, cheek, and the other) from the final datasets of selfies and generate eight intensified pixel values for the designated regions. Generally, Instagram users show strong manipulating behavior on the face regions such as jaw, nose, and cheek. Eyebrow and eye are the locations where they typically retain their natural state. The number of 'likes' that were crawled was recoded as normal (75% below; n=107) and influencers (25% above; n=324). The number of 'likes' ranges from 0 to 18790, with a mean of 141.10 and a standard deviation of 1081.441. With setting factor values of culture (1: South Korea, 0: United States) and influencer (1: top quarter of the number of 'likes', 0: others), we used ANOVA to determine the statistical differences in manipulating intensity on face features. As a result, South Koreans are more likely to alter their selfies in all areas of the face; however, statistically significant differences were found in three areas: forehead ($F = 5.265$, $p = 0.022$), jaw ($F = 3.433$, $p = 0.065$), and others ($F = 4.681$, $p = 0.031$). In addition, influencers demonstrated substantially more manipulative behavior in the mouth ($F = 4.76$, $p = 0.030$), cheek ($F = 6.06$, $p = 0.014$), and jaw ($F = 6.27$, $p = 0.013$). With the application of novel AI-detection technology, this study is anticipated to provide a deeper understanding of Instagram users' image distortion behaviors. Using the open-source algorithms, we could identify the specific features of the face that users may wish to alter. According to cultural differences, South Koreans are likely to retouch the entire face rather than specific features such as the eyes, eyebrow, nose, and mouth. Influencers also have more pronounced manipulative tendencies; however, they tend to modify the lower portion of the face, which can make the face appear smaller.