How to Cite:
Aote, S. S., Potnurwar, A., Vallabhaneni, S., Pogalla, N. R., Kumar, P., Doshi, R. B., & Khurram, S. A. (2022). AI enabled makeup tool. *International Journal of Health Sciences*, *6*(S1), 10357–10364.
https://doi.org/10.53730/ijhs.v6nS1.7488

AI enabled makeup tool

Dr. Shailendra S Aote
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Corresponding author email: *aotess1@rknec.edu*

Archana Potnurwar
Priyadarshini College of Engineering, Nagpur
Email: *archanapotnurwar@gmail.com*

Sreya Vallabhaneni
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Email: *vallabhaneniss@rknec.edu*

Naveen Reddy Pogalla
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Email: *pogallanr@rknec.edu*

Puru Kumar
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Email: *kumarp_2@rknec.edu*

Rohit Bhavesh Doshi
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Email: *doshirb@rknec.edu*

Syed Armaan Khurram
Shri Ramdeobaba College of Engineering Katol Road, Gittikhadan, Nagpur 440010, Maharashtra, India
Email: *khurramss@rknec.edu*

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 9 May 2022

***Abstract*---**The transition of technology over the years increases its application in new domains such as face detection using a pre-trained model. The cosmetic industry is most popular now with the need for automating the experience of a trial virtually. The objective of this project is to create an artificial intelligence-enabled makeup tool which provides the same level of experience. This is done using semantic separation in the profile preview using a pre-trained model called BiSeNet. This model identifies facial features into its classes. The makeup tool changes the color of lips and hair giving a realistic look for a better realisation. This provides a virtual experience to try on different choices of the user without the need to visit a store.

***Keywords*---**Semantic Segmentation, CNN, BiSeNet.

Introduction

In the month of December 2019, we encountered the sudden outbreak of the Coronavirus, later renamed as COVID-19, in the city of Wuhan, China. Leading to a global pandemic like never before. With this, all-time running businesses as well as the new ventures had come to a halt. Focusing on the cosmetic industry that solely worked on trial and choose basis is also trying to reach out to its customers providing the same level of service. To overcome this challenge the same should be available on an online platform with the ease of access. As the customers rely on trying out the products before buying. This poses a serious problem now since the same lipstick used on different people can expose the people to the virus.

The problem we are trying to solve is to provide a makeup tool which will take a front profile image from the customer/user and categorize their lips and hair using the Machine Learning concept of semantic segmentation. Then appropriately change the color of lips and hair in the image to the given user input color as naturally as possible. This will help a customer to understand and realize the changes better without the need of actually leaving their house and physically applying the makeup. If we try to use photoshop just to try the color, it takes a lot of time, effort and skill to achieve a natural look. For that, we decided to use Bilateral Segmentation Network (BiSeNet) which is a convolutional neural network and do this within seconds. BiSeNet is a modern version of the Real-time Semantic Segmentation that uses both Spatial and Context path methods. Spatial Path is a proposed way to save the original size of the included image and encode affluent spatial information (high definition features). The Spatial Path is made up of three layers of convolutions. Context Path is designed to work hand-in-hand with the Spatial method, providing an ample reception platform. In the semantic segregation function, the reception field is critical to performance [1,7].

This also saves the money used in buying expensive software to test the color too as all the libraries used in our project will be open source. Since the makeup tool will be developed in python it can be used by any customer with the understanding of basic fundamentals of computers. Also, it can run on any system with the standard specifications. A survey of women found that six out of ten (60%) wear makeup weekly, while the other 26% apply it at least twice a week and 28% wear it daily[2]. If the makeup tool is accurate, we will try to expand our scope from districts to states, and then to the entire country.

Related Work

Lenskart is the leading e-commerce portal for eyewear in India. Its 3D try on feature uses the futuristic technology by “DITTO”, (virtual try-on and frame-recommendation technology for eyewear retailers), allowing customers to try eyeglasses in virtual reality and see what looks the best on them before buying. The 3D virtual try-on system allows them to see themselves in different eyewear at 180 degree angles [3]. First, the software creates the 3D face model of the customer's head with accurate scale. Then, digitalize the glasses of their choice photo-realistically. Finally, rendering the glasses onto the DITTO (customer’s face model) with proper scale and with matching lighting and shadows so you can see which pairs fit and flatter your unique face. This is a complex and expensive software used by established enterprises. The beauty brand Lakme has come up with the first real application for makeup (Lakme makeup pro). It comes with the latest Facial Recognition Algorithm [4] and has built-in color-correction technology. This app enhances the look based on your skin tone.

We are going to use BiSeNet for semantic segmentation. This convolutional neural network (CNN) classifies the facial features into 19 classes among which are lower lip, upper lip and hair. We decided to specifically use this CNN as it is less complex than other neural network which can also be used for semantic segmentation. And, also the classes can be used to easily access the specific facial features.

Other innovative companies which use facial recognition as their service are Facebook, Google and Amazon. A research group at Facebook created a deep learning facial recognition system called *Deepface* which recognizes human faces in any digital photo. A nine-layer neural network was employed in the program having over 120 million connection weights. Trained on images uploaded by over four million users. The team also stated that the system has an accuracy of 97.35% which means it sometimes recognizes human faces better than human beings. This system is used to auto-suggest people on their images. Google researchers developed a facial recognition system called FaceNet for the services provided by them to their users. The system employed end to end learning with underlying ZF-Net and Inception in its architecture. Their approach high-quality face mapping with an accuracy of 99.63%. This is also used in Google Photos for face clustering, collage or naming detected faces. Two different technologies are used for images and videos. For images cloud vision is used and video-intelligence is used for videos. Amazon uses a cloud-based software as a service (SaaS), called Amazon *rekognition*, to provide analysis. It also provides real-time analysis, scene and text detection and a lot more. The CNN we are using provides an accuracy of 68-75% for semantic segmentation.

There are a lot of approaches based on Fully Convolutional Network (FCN) that have achieved state-of-the-art performance on different levels of the task of semantic segmentation. Most of these methods are designed to encode more spatial information or enlarging the receptive field for faster and high-quality prediction. The CNN encodes all the high-level semantic information with continuously down sampling operations. The spatial information of the image is crucial to predicting the detailed output while in the task of semantic segmentation. Modern existing approaches are solely dedicated to encode affluent spatial information. The U-shape structure of this network (figure 1) can recover a certain extent of spatial information. The original FCN network encodes different layered features by a skip-connected network structure. Some methods employ their specific refinement structure into U-shape network structure for better results. Semantic classification requires contextual information to be used to produce a high quality result. Most common methods include expanding the reception field or integrating information in a different context. Also, various extension measures of the convolution layer are used to extract information on different contexts. Sometimes the structure is used with deconvolutional layers as a refinement.

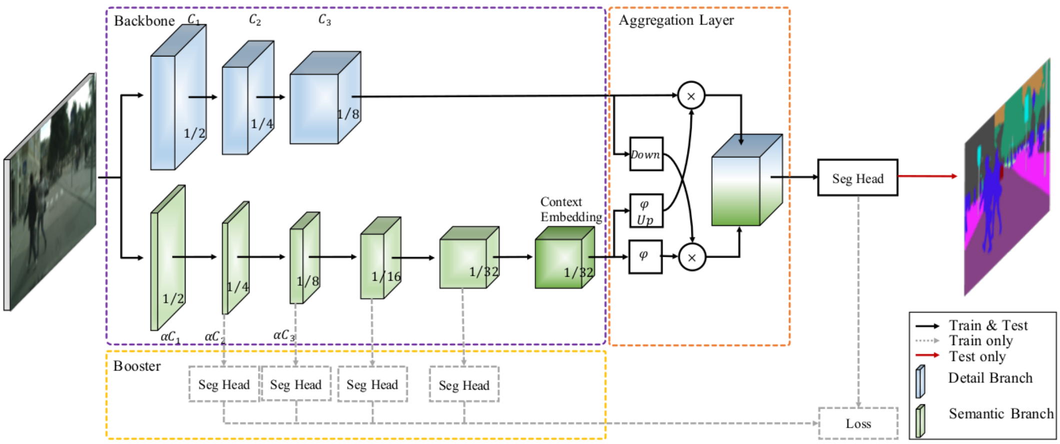

Figure 1: BiSeNet architecture

The U-shape is formed by the backbone and aggregation layer blocks. Further the model's architecture is modified to meet the objective need which is most of the time for real-time semantic segmentation.

Methodology

This part of the paper describes our approach and technologies that are used for recognizing facial features of the user with BiSeNet as a neural network which will give us the data for further processing and generate the expected results.

Semantic Segmentation

The process of identifying and classifying an image at pixel level into a class is called semantic segmentation. Given an image with more than one different object when applied semantic segmentation generates different colors to identify different objects. This approach is trained upon models to carry out the segmentation which can be used for multiple operations like self-driving cars, robotics systems, facial recognition and more.

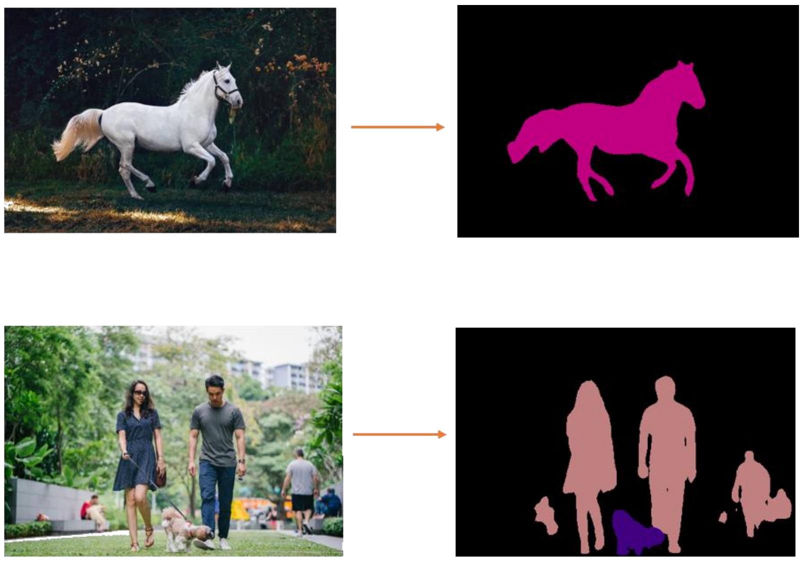

Figure 2: Before and after effect of semantic segmentation

In Figure 2, the first image in which the horse is detected is assigned a color different from the background color (black) which is default if a particular group of pixels don't classify into a class. The second image (figure 2) all the persons are assigned the same color as they all classify into the same class and the dog into another. Since we will be accepting a pre-user profile picture we can ignore the possibility of any such confusion occurring when any two facial features will be the same color.

BiSeNet

The idea of teaching the computer to apply makeup on an image will begin with identification of the part on wfacialhich makeup has to be applied. This work is done using the BiSeNet model which will apply semantic segmentation to the input image which classifies each and every pixel of the image to one of the 19 classes it has been trained on. Among these we will be focussing on upper lip, lower lip and hair classes for our objective point of view.

Figure 3: Transition of a front profile image of a customer before and after applying semantic segmentation using BiSeNet

The above figure 3 shows how this neural network identifies the facial features into its 19 classes. All the features are assigned different colors to be distinguished from the others. For the purpose of our makeup tool we focus on the lips and hair and all the other features are set to their default color. Also, these three classes are assigned a color or shade chosen by the customer for use. This is done by changing the color of the pixels separated between classes. Below mentioned are the 19 classes that a facial feature is classified into in BiSeNet.

labels=[‘background’, ’skin’, ’left_brow’, ’right_brow’, ’left_eye ’,’right_eye’, ’eye_glasses’, ’left_ear’, ’right_ear, ’ear_ring’ ,’nose’, ’mouth’, ’upper_lip’, ’lower_lip’, ’neck’, ’neck_lace’, ‘cloth’, ’hair’, ‘hat’];

In the program code these classes are distinguished in different color codes with the same order as mentioned above. The changes we make are to the upper & lower lip and hair classes to match the choice of the user. And, other classes are left to their default color.

Results & Discussion

Advancing step by step. First, we collected the basic inputs needed for the processing successfully using basic functions provided by the OpenCV [5] and OS library [6]. Then using the functionality provided by BiSeNet [7] we detected lips and hair. The network is used as the go-to approach for semantic segmentation [1]. Using the functionality provided by PIL [8] and pytorch [9] made it possible to make these changes as realistic as possible. Then normalizing this image for the same or better resolution we have used numpy that provides the same functionality. Now, we had the after image with accurate realisation. To give the user a better understanding we decided to output both the before and after image. With the user-friendly GUI all the transitioning is very easy for any user to carry out this process efficiently and at ease. Finally, we ask the user if there are any more changes or exit the process. Also, as the tool is developed in python it can also be used in a system with common definitions that are also advantageous over the use of expensive and sophisticated software.

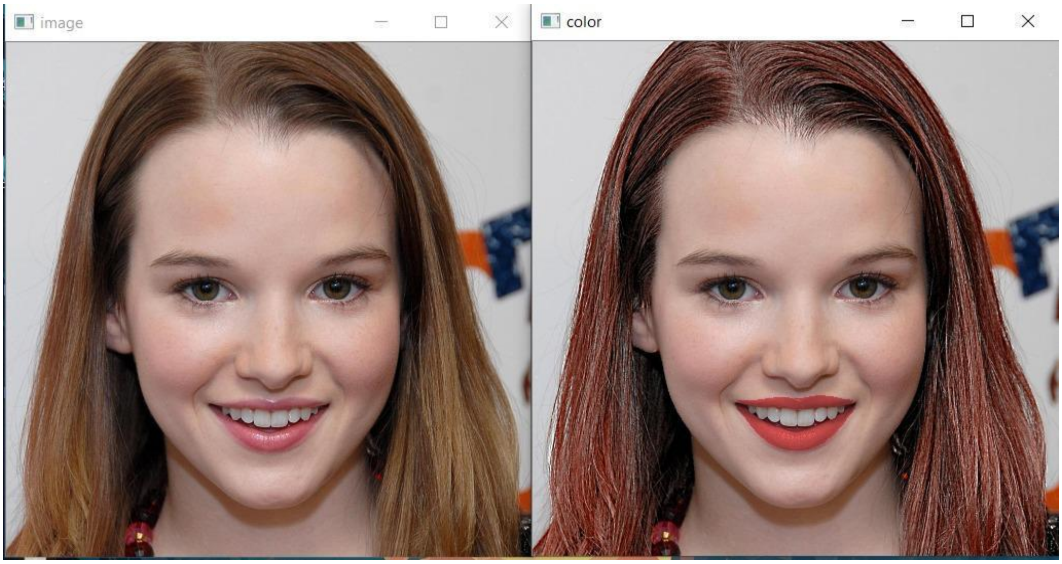

Figure 4: Output of the project

In the above figure 4, the output of this project is demonstrated. It shows the before and after image of the front profile image of the customer with changes to the color of the lips and hair chosen by them. As you can see the changes are as realistic as possible. Also, the differences in the images help the customers to understand better.

Conclusion

The aim to create a makeup tool for a customer to understand the color of lipstick or dye they want to buy without physically visiting a store and applying on themselves was achieved. With the ease and avoiding an expensive software to do the same was also a desired result. The GUI is very easy to use and the processing is all done in the background giving the result to the user in very less time. The after image is very accurate and gives a real like realisation to the customer and gives them a better understanding of their choice.

References

[1] Canuma, P. (2018, December 6). Building custom models using Keras (BiSeNet) Part III. Retrieved from medium.com: https://prince-canuma.medium.com/bisenet-for-real-time-segmentation-part-iii-f2b40ba4e177
[2] Waldersee, V. (2019, March 14). One in six women wear makeup at the gym. Retrieved from yougov: https://yougov.co.uk/topics/lifestyle/articles-reports/2019/03/14/one-six-women-wear-makeup-gym
[3] Lenskart. (2018, January 12). How does Lenskart's 3D try on work? Retrieved from quora: https://www.quora.com/How-does-Lenskarts-3D-try-on-work

[4] Hub, I. B. (2015, March 23). Lakme Makeup Pro App – India's First Real-Time Virtual Makeup App. Retrieved from indianbeautyhub: https://www.indianbeautyhub.com/lakme-makeup-pro-app/
[5] doxygen. (2021, October 16). OpenCV Tutorials. Retrieved from docs.opencv.org: https://docs.opencv.org/master/d9/df8/tutorial_root.html
[6] Pankaj. (2019, 6 11). Python os Library Functions. Retrieved from JournalDev: https://www.journaldev.com/30205/python-os-library-functions#:~:text=%20Python%20os%20Library%20Functions%20%201%20Python,At%20this%20point%20will%20be%20familia...%20More%20
[7] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, & Nong Sang. (2018). BiSeNet: Bilateral Segmentation Network for. ECCV, 17.
[8] Lundh, F. (1995, June 21). Pillow. Retrieved from pillow: https://pillow.readthedocs.io/en/stable/
[9] pytorch. (n.d.). Pytorch. Retrieved from pytorch: https://pytorch.org/