AFRIRAZER: A Deep Learning Model to remove Background and Skin from Traditional African Fashion Images

4 authors:


Wuraola Fisayo Oyewusi
Data Science Nigeria


Gbemileke Onilude
Data Science Nigeria


Olubayo Adekanmbi
City, University of London


Olalekan Akinsande
Data Science Nigeria

AFRIRAZER: A Deep Learning Model to remove Background and Skin from Traditional African Fashion Images

Wuraola Fisayo Oyewusi
Data Science Nigeria
Lagos Nigeria
wuraola@datasciencnigeria.ai

Gbemileke Onilude
Data Science Nigeria
Lagos,Nigeria
gbemileke@datasciencenigeria.ai

Olubayo Adekanmbi
Data Science Nigeria
Lagos Nigeria
olubayo@datasciencenigeria.ai

Olalekan Akinsande
Data Science Nigeria
Lagos,Nigeria
olalekan@datasciencenigeria.ai

Abstract

Africa is known for its rich and diverse culture. Africans are flamboyant and fashion is an expression of our shared heritage as a people.Traditional African fashion is rich, colourful and vibrant,and allows for adaptable customization. In this work, we present AFRIRAZER, a deep learning model for background and skin removal in African fashion images. While the concept of image segmentation using deep learning for fashion is not new, traditional African images are not well represented in the common fashion training images. This work is part of a larger ongoing project on the curation of clean, de-identified and applicable traditional African fashion for Artificial Intelligence related work. AFRIRAZER uses deep learning to remove background and skin from images

1 Background

Image segmentation is critical in extracting and/or understanding the different segments a digital image comprises of, it's goal is to simplify the representation of an image into something more meaningful and easier to analyze. A lot of work has been done on image segmentation, both in the line of medical images and fashion [4,2 5,7]. However,little has been accomplished in terms of low resourced traditional fashion images.There is a need to answer the question on how images can be segmented when the available dataset for training I small and the structures of the images to segment

It is important to look into fashion segmentation because fashion is deeply entrenched in the culture of people in Africa. It is an expression of the culture[8], Traditional communities around the world are identified based on some factors: language, fashion (dress sense),beliefs e.t.c. For instance Nigeria is a country with 250 ethnic groups and each ethnic group has its tradition and unique fashion. Designs differ from one group to the other and they are so distinct that one can identify an ethnic group through fashion outfits. However, these ethnic groups have really low online presence when it

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

comes to text content and images (traditional attires)[9] and are being left out of Machine Learning innovation and research[9]. A simple google search will confirm the low availability of African traditional fashion images and when available,they are mostly on human models with differing image backgrounds which as innocuous as it seems can promote exclusion of use in research.

This paper aims to provide remedies to some of the exclusion issues by building a deep learning model that can extract skin, background, and dress from a low resourced traditional fashion image. In this paper, we present an effective method for image segmentation with low resource traditional fashion images using U-net architecture. The best performing model was trained on a small dataset of 1,945 images and it achieved a mean accuracy of 0.95 over three different training datasets.

2 Methodology

2.1 Model Architecture

In this study, U-net architecture[7] was used to build the image segmentation model, it segments the dress, background, and skin from traditional images. U-net work by learning the feature mapping through the contraction part of the architecture while the expansion part convert feature mapping back to image. The full U-net architecture is shown in figure 1, it depict the contraction,expansion and bottleneck part of the architecture.

The initial filter size of the U-net architecture used was 32 and it was doubled for every repeated application of two 3x3 convolutions layer during the contracting path of the network architecture. The input size of the network was 224 x 224 x 3. Other factors used during the initial development of U-net network architecture remains constant.

The output image from the U-net network architecture is a size of 224 x 224 x 3. The first depth of the output image contains the skin segmentation, the second depth contains the dress segmentation while the last depth the background.Each depth within the image has been normalized between 0 to 1.

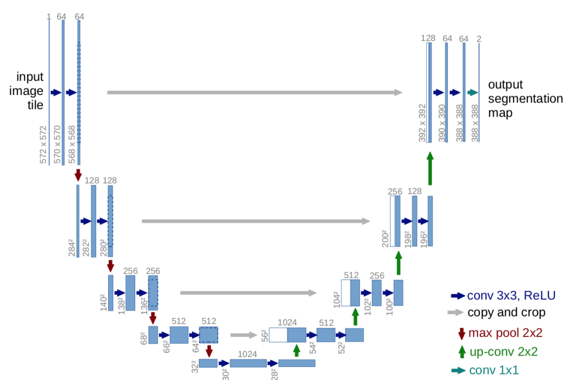


Figure 1: U-net architecture

2.2 Dataset and Preprocessing

For this project about 1900 openly available images of people wearing African fashion items were scrapped from the internet. The copyright of all images used belong to the original owners. The training data was preprocessed by manually removing image backgrounds and skin from the original images. For file management, each image was replicated into 3 folders namely Picture, Body and Dress.The picture folder contains the original images with no alteration. The body folder contains the image with background removed leaving a white background as substitute. The dress folder contains the same image with the background and skin removed, leaving only the dress and white background,The files in each folder were named serially, with same images in Dress, Body, and Pictures folder having the same file name.Figure 2 depicts this process outcome.

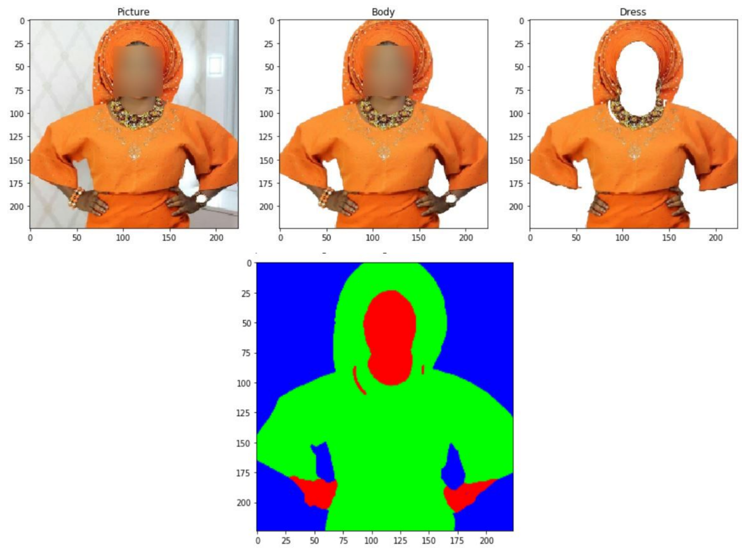


Figure 2: Figures (a) and (b) depict how the dataset was developed

2.3 Model Training

The U-net architecture[7] used in this project is popular for achieving good outcomes even with small datasets, we found this to be true during this project. Three models were trained using different proportions of the available dataset as shown in Table 1 .

3 Result

Table1 compares the performance of the three different models based on accuracy,loss and intersect over union (IoU) values for background,skin and dress

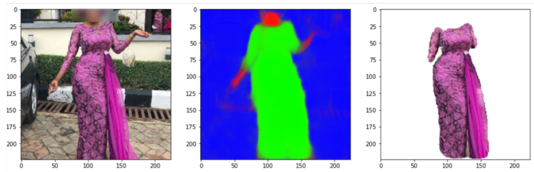
	Model 1	**Model 2**	**Model 3**
Number of training data	512	958	1943
Accuracy	0.93	0.95	0.97
Loss	0.12	0.08	0.06
IoU (Background)	0.83	0.92	0.94
IoU (Skin)	0.40	0.59	0.66
IoU (Dress)	0.68	0.78	0.81

Table 1: Accuracy and IoU Values per Model

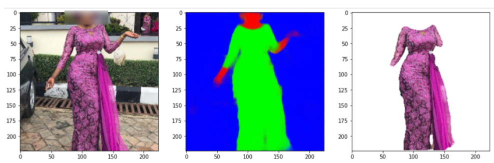
Figure 3 shows a comparison in model performance on the same image. Model 1 with an accuracy of 0.93, trained with the least number of images successfully separates background from skin and dress but did not get a proper outline of the dress in the image. However if the goal is to remove image background, it works well. Model 2 and 3 with respective accuracies of 0.95 and 0.97 were trained with more images and had better performance in dress segmentation as expected. The average size of the model is about 33 Megabyte which is light and easy to deploy. Figure4 shows the performance of the best performing model in different standing and sitting positions.

4 Conclusion

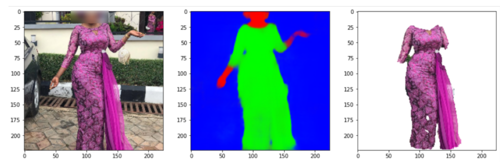
We present AFRIRAZER, a deep learning model that can remove background and skin from images with focus on African images.With as little as 512 training images, the model successfully removed image background and with increased training data,there was a significant performance improvement in removing skin and outlining dresses from images.



(a) Model 1

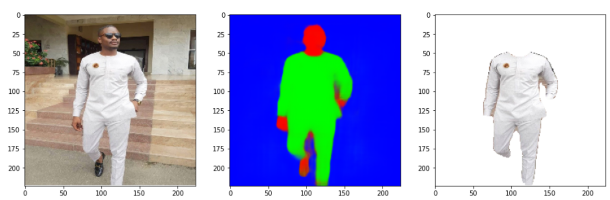


(b) Model 2

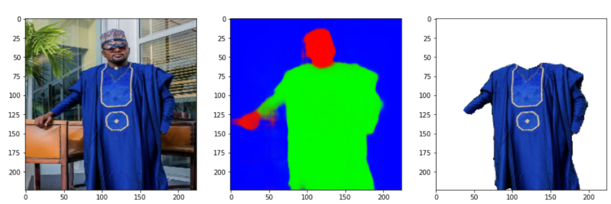


(c) Model3

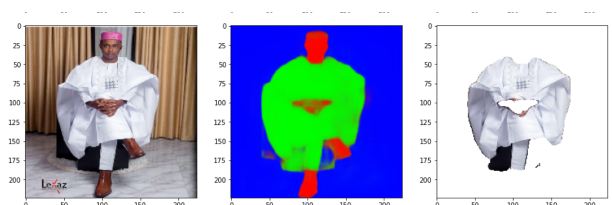
Figure 3: Comparison of the performance of the 3 models on the same image



(a)



(b)



(c)

Figure 4: Figures (a),(b),(c) shows the model in action removing both background and skin from different images with different pose

Broader Impact

This work makes a case for inclusion of African traditional fashion in the application of artificial intelligence. This project aimed to highlight the relevance of deep learning in practical use cases with focus on the fashion industry. All images used belong to the original copyright owners and the authors are not aware of any harmful or discriminatory effect of the work.

References

[1] Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. Large-scale kernel machines, 34(5), 1-41.

[2] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.

[3] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

[4] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (2020, May). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1055-1059). IEEE.

[5] Raffiee, A. H., & Sollami, M. (2020). GarmentGAN: Photo-realistic Adversarial Fashion Transfer. arXiv preprint arXiv:2003.01894.

[6] Baur, C., Denner, S., Wiestler, B., Albarqouni, S.,& Navab, N. (2020). Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study. arXiv preprint arXiv:2004.03271.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[8] Davis, F. (2013). Fashion, culture, and identity. University of Chicago Press.

[9] Orife, I., Kreutzer, J., Sibanda, B., Whitenack, D., Siminyu, K., Martinus, L., ... & Meressa, M. (2020). Masakhane–Machine Translation For Africa. arXiv preprint arXiv:2003.11529.