JOURNAL OF DATA SCIENCE | Vol.2024:06
eISSN:2805-5160

Study on Image Background Removal using Deep Learning

Fakruddin Ali Ahmed[1], Harprith Kaur Rajinder Singh[2]

[1]Dayananda Sagar Academy of Technology and Management College Bangalore, India
[2]Faculty of Data Science and Information Technology, INTI International University, Malaysia

Email: alishaikh9742@gmail.com

Abstract

Removing image backgrounds is a common job in image processing and computer vision. By isolating the main object from the back, background removal in photographs aims to make it easier to examine or edit the image. There are numerous methods for removing the background from an image, including deep learning, color-based segmentation, and human selection. The U-Net architecture, one of the deep learning-based techniques, has demonstrated encouraging results in image segmentation tasks, including image background removal. A convolutional neural network created for biological image segmentation is known as the U-Net architecture. The design consists of an encoder network that stores the context and a decoder network that generates the segmentation map. The U-shape of the U-Net architecture enables it to record both the overall context and the local specifics of the image. For several picture segmentation tasks, including image background removal, U-Net architecture has undergone modification. The suggested method for removing image backgrounds using U-Net entails training a U-Net model on a dataset of pictures with and without background. Then, using the demonstrated methodology, the backdrop is removed from recent photographs. The suggested method differs from current approaches in various, including its high accuracy and capacity to handle complicated backgrounds. Computer vision, object identification, and photo manipulation are just a few of the uses for the suggested method.

Keywords

COCO Dataset, Convolutional Neural Network, U-Net Architecture, Picture Segmentation

Submission: 12 May 2024; **Acceptance:** 29 May 2024

Introduction

The backdrop of a photograph can be removed using computer vision and semantic segmentation. Background removal is a digital image processing technique that can be used to divide an image's components into desirable and unattractive parts (Kc, K., Yin, Z., Li, D., & Wu, Z.,2021).

When using some sort of "marker" and edge detection, background removal is a work that is relatively simple to complete manually or semi-manually (Photoshop, and even PowerPoint, have such tools) (Cheng, C. C., 2021). For an example, see this page. Fully automatic background removal, however, is a difficult task, and although several products attempt it, as far as we know, none of them have successful results. Before further analysis and processing, background reduction is necessary for many applications of image processing and computer vision. For instance, foreground object extraction can use backdrop removal (Lu, Y., & Guo, H., 1999). A multi-layered artificial neural network modeled after the structure and function of the brain is used in deep learning, a branch of machine learning, to comprehend patterns and achieve cutting-edge accuracy (Kang, M. S., & An, Y. K., 2021). Deep learning algorithms have vastly outperformed traditional methods in problems involving the processing of signals, pictures, videos, talks, and texts. An example of a convolutional neural network (CNN) is a widely used image-processing technique in deep learning (Bahri, F., & Ray, N., (2022). A convolutional neural network (CNN), an artificial neural network with numerous convolution layers, may be used for the segmentation of interesting portions from an image (Budianti, N. I., et.al., 2022).

Which context are we going to eliminate? This turned out to be a crucial question because the better the quality of the separation, the more exact the model must be regarding the objects, angle, etc. We had large ideas when we first started working on an overall background remover that could recognize both backgrounds and foregrounds in any kind of image. However, after creating our initial model, we understood that it would be better to focus our efforts on a certain set of images. So, we decided to focus on selfies and human portraits.

Semantic Segmentation

Semantic segmentation, with object detection and classification, is one of the top three computer vision tasks.

In the sense of assigning each pixel to a class, the segment is a categorization process. The segmentation model, in contrast to image classification or detection, really demonstrates some "understanding" of the images by not just stating that "there is a cat in this image" but also by pointing out exactly where and what the cat is, down to the pixel level (Elhabian, S. Y., El-Sayed, K. M., & Ahmed, S. H. , 2008).

With photo background removal, the background is unheeded while the foreground objects are precisely detected and isolated. The technology goes beyond conventional approaches like

eISSN:2805-5160

Chroma entering (green screen) or straightforward thresholding since it uses semantic segmentation, which can handle intricate scenarios with varied backgrounds.

Here is a general description of the procedure:

Dataset Preparation

We began hunting for appropriate datasets once the model had established our broad course. Data segmentation is less frequent than categorization or even detection. Furthermore, manual tagging is not an option. The COCO dataset, which consists of over 80K images with 90 categories, the VOC Pascal dataset, which has 11K images and 20 classes, and the more recent ADE20K datasets were the most often used datasets for segmentation. To train a semantic segmentation model, a sizable dataset is needed. The photos in this dataset each have associated pixel-level annotations that designate the foreground or background of each pixel (Hardas, A., Bade, D., & Wali, V. , 2015).

We debated whether to use a broader comprehensive dataset or just the photographs that are extremely relevant to our task. On the one hand, using a more diverse dataset with more pictures and classifications will provide the model with additional scenarios and problems to handle (Bouwmans, T., 2019). On the other hand, a training session that lasted all night allowed us to review 150K photos. Trimming it a little bit will be helpful because if we use the complete COCO dataset to introduce the model, we'll end up with the model viewing each image twice (on average). It will also result in a model that is crafted to meet our demands.

Model Training

Several deep learning models, including U-Net, Mask R-CNN, and DeepLab, can be used for semantic segmentation. The model is trained using techniques such as supervised learning using the dataset produced in the preceding stage. The model first categorizes each pixel in an image as either foreground or background (Wang, H., Xie, Q., Zhao, Q., & Meng, D., 2020).

Unet

This article demonstrates how to create an image segmentation model using U-Net to predict the mask of an item in a photograph. The model can localize the object in the image using this method. An encoder (for downsampling) and a decoder (for up sampling) with skip connections make up the model's straightforward architecture. It is shaped like the letter U hence the name U-Net.

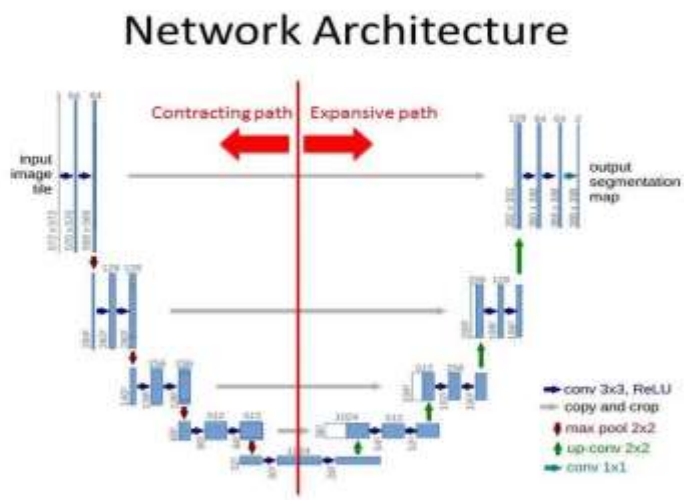

Figure 1. Network Architecture

The concatenation of the encoder feature map with the decoder, shown by the gray arrows, facilitates the backward flow of gradients for better training. Our model was trained using the same schedule as that of the original paper: modest decay, RMSProp optimizer, standard cross-entropy loss, and 1e-3 learning rate. We divided the 11K photos we had into three groups: training, validation, and test. The pictures you see here are all from our test set.

Apply the model for background removal.

The model can be used to infer new images after it has been trained. The trained model receives the input image and forecasts a segmentation mask in which each pixel is categorized as either foreground or background. The segmentation mask can be improved, and the outcome enhanced using post-processing methods like morphological operations or contour detection.

You can use the model to erase the background of new photographs when it has been trained and certified. If you run a picture through the model, it will estimate the likelihood that each pixel is in the foreground or background. You can design a binary mask that distinguishes the foreground from a backdrop determined by the anticipated probability.

Results

We achieved an IoU of 84.6 on our test set, which is below the current state of the art of 85, but our results were still pleasing. However, this quantity is problematic because it varies among datasets and classifications. Certain classes, such as houses and highways, are intrinsically simpler to segment and most models may easily achieve results of 90 IoU.

Humans and trees are two additional more difficult classes, with the outcome for most models hovering around 60 IoU. We assisted our network in narrowing its focus to a particular division and a select group of photographers to assess this challenge. While we still don't feel that our work is "production ready" in the manner we would like it to be, we believe that now is a good moment to pause and talk about our findings because it's a rough estimate.

Figure 2. Test Set of Background Removal

Conclusion

As stated, our goal was to create an exceptional deep-learning product. Deployment is getting effortless and more efficient every time, as you can see from Alon's postings. On the other hand, training a model is challenging and necessitates careful planning, debugging, and results recording, especially when done overnight. Finding a balance between training, improvement, research, and attempting new things is also heavy. Since deep learning is handed down, we constantly have the impression that the ideal model or the precise model we require is just around the corner and will be found by conducting one more Google search or reading one more article. Though, we made progress by merely "squeezing" more and more out of our previous model. And as was already mentioned, we believe there is still a lot more to be gotten out of it.

References

Kc, K., Yin, Z., Li, D., & Wu, Z. (2021). Impacts of background removal on convolutional neural networks for plant disease classification in-situ. *Agriculture, 11*(9), 827.

Cheng, C. C. (2021, February). Single-image Background Removal with Entropy Filtering. In *VISIGRAPP (4: VISAPP)* (pp. 431-438).

Lu, Y., & Guo, H. (1999, September). Background removal in image indexing and retrieval. In *Proceedings 10th International Conference on Image Analysis and Processing* (pp. 933-938). IEEE.

Kang, M. S., & An, Y. K. (2021). Deep learning-based automated background removal for structural exterior image stitching. *Applied Sciences*, *11*(8), 3339.

Bahri, F., & Ray, N. (2022, November). Dynamic background subtraction by generative neural networks. In *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-8). IEEE.

Budianti, N. I., Gibran, M., Risnumawan, A., Anggraeni, M. E., & Ratri, A. (2022, August). Background blurring and removal for 3d modelling of cultural heritage objects. In *2022 International Electronics Symposium (IES)* (pp. 664-670). IEEE.

Elhabian, S. Y., El-Sayed, K. M., & Ahmed, S. H. (2008). Moving object detection in spatial domain using background removal techniques-state-of-art. *Recent patents on computer science*, *1*(1), 32-54.

Hardas, A., Bade, D., & Wali, V. (2015). Moving object detection using background subtraction shadow removal and post processing. *International Journal of computer applications*, *12*(2), 12-20.

Bouwmans, T., Javed, S., Sultana, M., & Jung, S. K. (2019). Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *Neural Networks*, *117*, 8-66.

Wang, H., Xie, Q., Zhao, Q., & Meng, D. (2020). A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3103-3112).