CSCI-4930/5930: Machine Learning
Project Progress Report - Final

Title of the project: Separation of Style and Content: Semantic Segmentation and an Isolated Application of Style Transfer

Team:

I. Ananta Nair
II. Medha Chippa

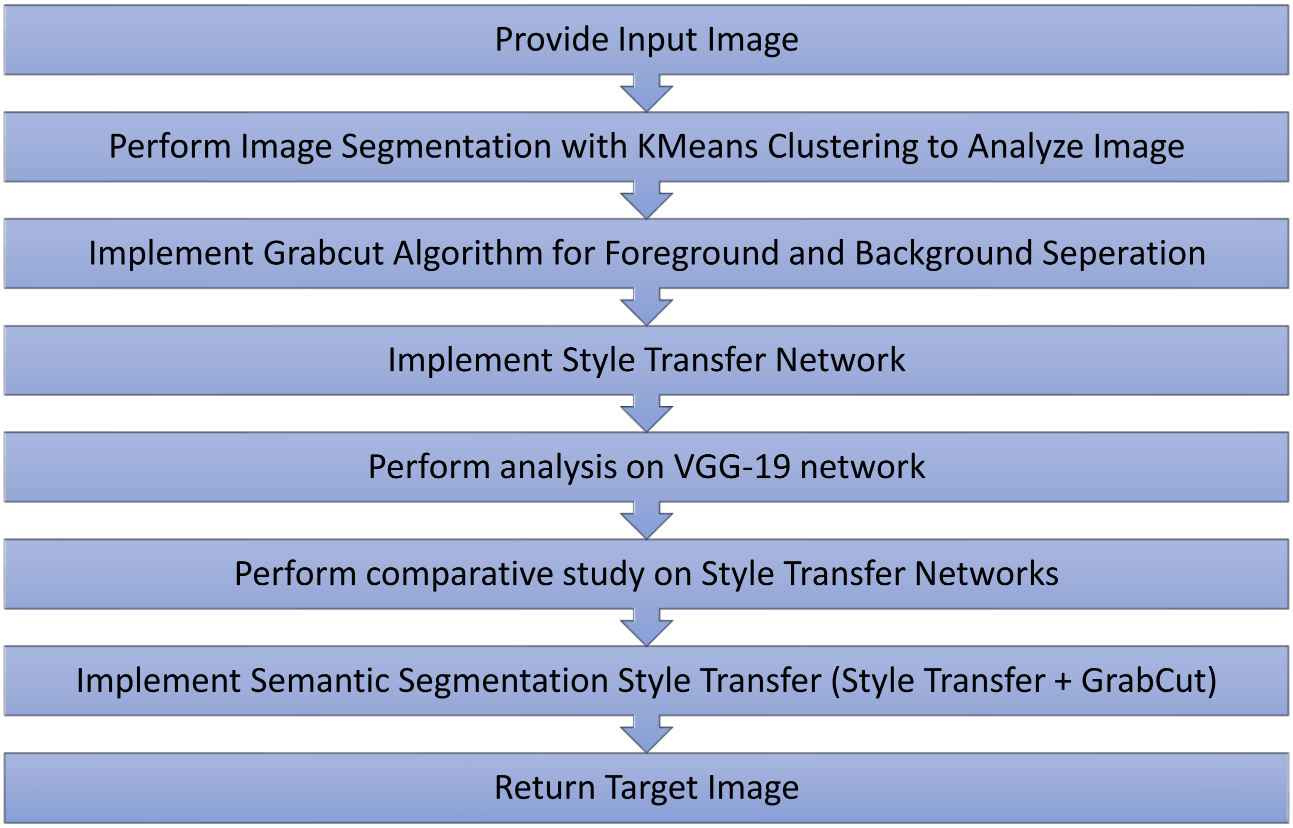
Introduction

Definition of the task(s)/problem:
In this project, we seek to create a neural network that can apply style transfer only to a segmented portion of the image, for example: the foreground or the background. We intended to do this by combining two processes– image segmentation and style transfer.

We firstly aim to perform k-means clustering for image segmentation to preprocess and better understand our images. Following this step, we aim to utilize semantic segmentation techniques (specifically: the grab cut algorithm [1]) to select images (example: portraits) to separate the foreground from the background. We lastly intend to build a deep convolutional neural network that is capable of style transfer. After these three steps are performed individually, we apply the grab cut algorithm to a style transfer target image to produce an image with isolated style transfer; i.e. style transfer is applied only to the foreground or background.

Through the course of this project we also aim to perform an exploratory analysis on style transfer networks. Each member of the team will build their own style transfer network optimized for best performance. A comparative study will then be conducted on both these models to draw conclusions as to what it takes to build a stable, reliable and well performing convolutional deep network. Furthermore, an analysis is also conducted on VGG-19 (the style transfer network used) to better understand its internal workings using feature maps and saliency maps.

Step by Step Diagram of the Project:
Illustrated on the next page.



Purpose of the task(s):
This task has two purposes:
1. Research directed purpose:
To explore both the latest techniques in machine and deep learning and see how they can be optimized and combined. The goal of this research directed approach is also to try to understand what it takes to build a reliable and stable deep learning model that achieves our project goals as well as try to better understand the internal workings of deep networks.

2. Commercial directed purpose:
(as indicated in our project proposal)
(i) Image filters have gained immense popularity through platforms such as Facebook, Instagram, and Snapchat. Though many filters exist and have existed for years, they have a large consumer base and a market that is nowhere near saturation point. As image filter creators can reach great success and growth through popular filters, we would like to expand the concept of style transfer to be applied to portraits and potentially other image scenes.
(ii) Style transfer is an example of where art and artificial intelligence intersect. Our project could serve as a tool that allows photographers and artists to experiment with future mixed media works of art.
(iii) If expanded into the three-dimensional domain, this could be applied to see how custom wallpapers, paintings, or even works of arts would look when applied to specific objects – walls, couches, kitchen counter, etc.

Information on the Dataset:
The goal of this project is not to build a network that runs on large datasets but to explore what it takes

to build a stable and reliable deep network for isolated style transfer. Thus, in line with the work previously done in style transfer we will not be using large established datasets. As indicated in the project proposal and report, we chose images of portraits or objects with a neutral background for our KMeans and GrabCut analysis. These images also were later used for the content aspect of the style transfer. For the style images of the style transfer network, famous artwork with distinctive styles were chosen to create a bold effect. These images, as indicated in the project proposal and report, were collected from wikiart, a large database of paintings. All images used in this project are included in the Github repository under the username annxanair in the Images folder.

Proposed method:

1. Image Segmentation using KMeans Clustering

: A preprocessing step is first conducted whereby image segmentation is performed using k-means clustering. The algorithm segments an image into discrete regions to gain a better understand of the components that exist within an image. This step was helpful in choosing the best images for the project. This analysis is also commonly used in other machine and deep learning tasks such as object segmentation, object recognition, content-based image retrieval etc.

2. GrabCut Algorithm

: For the first processing step to meet our project goals, we used the GrabCut algorithm. This is an image segmentation algorithm that is based on graph cut optimization, gaussian mixture models, and Markov networks and is used to separate the foreground object within an image from the background. This is done by creating a user defined bounding box and grouping pixels into either foreground (source nodes) or background (sink nodes).

3. Style Transfer Network

: For the second processing step to meet our project goals, each group member built a style transfer network. Both pytorch and tensorflow were used and each group member optimized their network to achieve best results. A style transfer network is created using a trained convolutional neural network (for the project, both team members used VGG-19) that takes in two input images (a style and a content image) and uses separate loss functions to create a single target image that combines the two features.

4. VGG-19 Analysis

: After the style transfer models were built, an analysis was done on the VGG-19 network to gain a better understanding as to how deep neural networks work and analyze information. This step was motivated by class discussions on ethical AI as well as Geirhos, Jacobsen, Michaelis et al (2020) [6] that states a large group of problems in deep learning today can be tied to the same issue: Deep networks are not learning the way we think they are learning but using unintended learning strategies or shortcuts. This phase of the project aims to better understand how deep networks are learning by printing out feature and saliency maps.

5. Comparative Analysis on Style Transfer Network

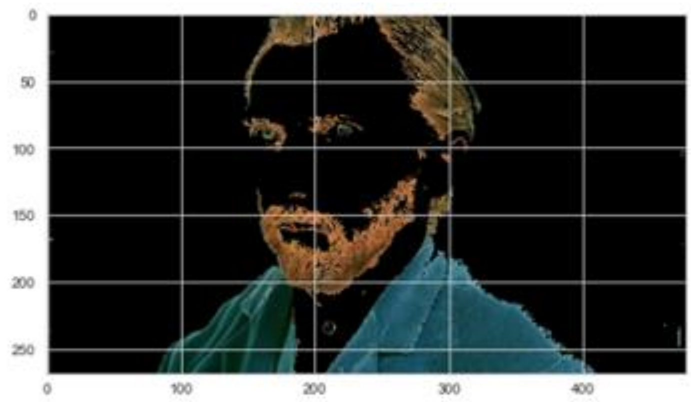
: After both the style transfer networks were created, a comparative analysis between the networks was performed. This analysis was done to give insights into what it takes to build a stable and reliable network and to help us choose the best network for the final step of the project; the semantic segmentation style transfer implementation.

6. Semantic Segmentation Style Transfer Network (Style Transfer + GrabCut)
: The last step of our project combines the best style transfer network with the grabcut algorithm to create a new type of tool. This network first performs style transfer to an image, and, using the target image obtained, grabcut image segmentation is performed to create an image where style transfer is only applied to the foreground or the background.

Results:
The following are the results as presented in the recorded presentation available at:
https://vimeo.com/417305783

K-Means Clustering
Medha: This is a preprocessing step that is applied to segment or partition the image into discrete regions or portions. For the purpose of the project, image segmentation using k-means clustering was conducted to gain a deeper understanding of the divisions and discrete regions that exist within an image. In this implementation, the image is read and preprocessed, the number of clusters are defined, segmentation is performed, and an output image is displayed. The k-means clustering algorithm was carried out by clustering the pixels of the image into distinct, semantically coherent regions. The OpenCV library was used for this initial pre-processing of the images.

The above figure depicts the discrete regions that lie within the image, each of these regions (black and grey) correspond to one of the two clusters that have been used.



To understand the importance of each cluster in the image, one of the two clusters has been blacked out. In the above image, cluster 1 (K=1) has been blacked out. The background and face correspond to cluster 1.

GrabCut Algorithm

Ananta: The GrabCut algorithm is an image segmentation method based on graph cut optimization. It was designed and published in Rother, Kolmogorov & Blake 2004 [1]. In GrabCut, a user defined bounding box is created around the object to be segmented. The algorithm then uses a Gaussian mixture model to estimate a color distribution that separates the foreground object from the background. A Markov network is constructed over the pixel labels and a function that prefers connected regions having the same labels is used to separate the image into two parts using a graph cut based optimization. For this project an open source implementation was used from the OpenCV library. The code is included in my Github repository, under the username annxanair. A result as shown in the presentation are included below. The original image used for this step was obtained from [4].

Style Transfer

Ananta: A deep style transfer network was created using Facebook’s open source machine and deep learning library, Pytorch. The pretrained network I used for this project is VGG-19. The model inputs two images, a content image and a style image, with the purpose of creating a single new image that is representative of the content of the content input but with the style of the style input. The model composes these new target images based on CNN layer activations and their extracted features. The internal workings of this network were described at length in the presentation.

Results from three sample images included in the presentation are below. Each model ran for 5000 epochs. Further details of the model are included on Github and in the comparative study table below. For the first set of images, the content image was obtained from [7], and the style image from [8]. For the second set of images, the content image was obtained from [2], and the style image from [3]. For the third set of images, the content image was obtained from [4], and the style image from [8]. This network is included in my Github repository, under the username annxanair.

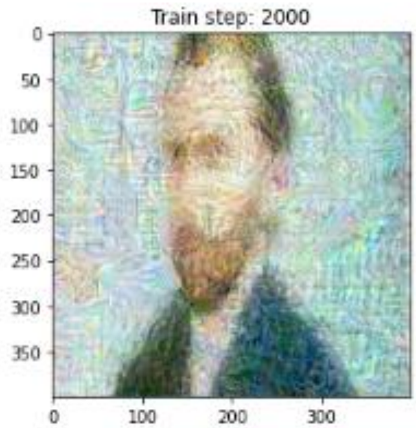
Content Style Target

Content Style Target

Content Style Target

Medha: Google's opensource machine learning and deep learning library, Tensorflow, was used to develop a Neural Style Transfer Network. The Pre-Trained VGG-19 model was used to train the model. From among the 16 layers in the pre-trained VGG 19 model, a custom model was designed by selecting 6 layers that include style layers (block1_conv1','block2_conv1','block3_conv1','block4_conv1','block5_conv1') and content

Layers('block4_conv2'). Style and Content features are extracted from the above layers. Custom weights then given to each of these features to generate the final model. The model takes a content image and style image as input and generates an image as output that is a result of incorporating the style from the style image into the content image.
The code to this is included in the Github repository under the username medha-chippa.



<u>VGG-19 Analysis</u>

Ananta:

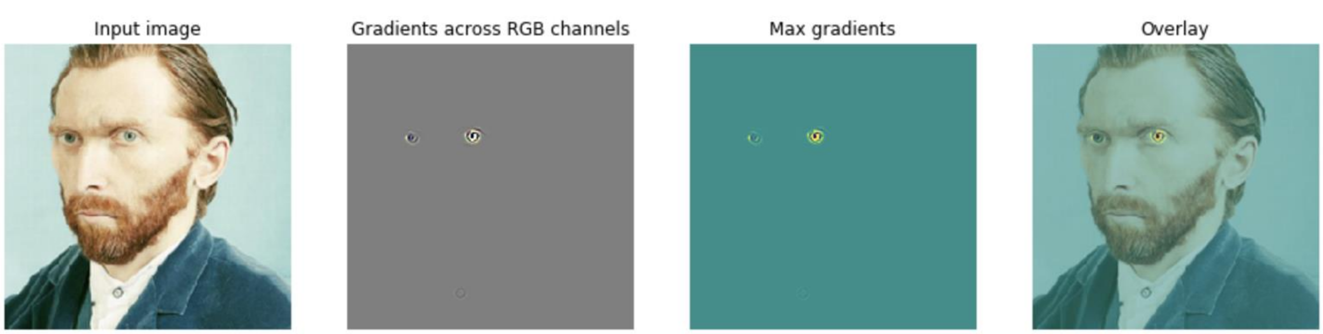
: As indicated above, this analysis was motivated by class discussions on ethical AI as well as Geirhos, Jacobsen, Michaelis et al (2020). Though deep networks have achieved superhuman performance on some tasks, they still face many challenges. The paper believes that many of the problems facing AI are due to unintended learning, or shortcut learning, and a better understanding of deep networks can help overcome these problems. Included below is an analysis of feature maps and saliency maps for VGG-19.

1. Feature Maps
: A feature map, or activation map, is the output activations for a given filter. These maps represent the activation of different parts of the image and the mapping of where a certain kind of feature is found in the image
: In the figure below, the output for the first convolutional layer of the VGG-19 network and shown. The figure includes 64 feature maps.
: This was done using the PytorchVis library. The code is included in my Github repository, under the username annxanair.

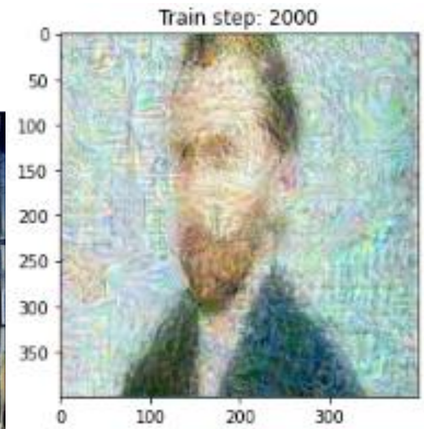
2. Saliency Maps

: The purpose of a saliency map is to represent the ‘saliency’ of an image. Essentially, saliency maps differentiate visual features in images and change the representation of an image into something that is more meaningful and easier to analyze. In the image plotted below, you can see the VGG-19 network learns to classify the image of a face by using the eyes. This image was created after applying guided backpropagation to reduce noise.

: This was done using the Flashtorch library. The code is included in my Github repository, under the username annxanair.



Comparative Study:



The above two images correspond to the target images generated by the two style transfer networks. The image on the left was generated by the network developed using Pytorch. The image on the right was generated by the network developed using Tensorflow.

The architectural constructs of both the models are shown in the table below. Both the models use the VGG-19 network and extract features from the same style and content layer. As seen in the above image, the Pytorch model outperforms the tensorflow model.
The reasons for this are as follows:

- Number of epochs
- Steps per epoch
- Weights for style and content updating
- Weights for the style layers
- Learning Rate

The best model, in this case the Pytorch model, was chosen for the final step of the project, the semantic segmentation style transfer (style transfer + grabcut).

Feature	Model 1	Model 2
Library	Pytorch	Tensorflow
Pre-trained model	VGG-19	VGG-19
Content layers	‘block4_conv2’	‘block4_conv2’
Style layers	block1_conv1 block2_conv1 block3_conv1 block4_conv1 block5_conv1	block1_conv1 block2_conv1 block3_conv1 block4_conv1 block5_conv1
Optimization function	Adam’s Optimizer	Adam’s Optimizer
Learning Rate	0.003	0.02
Style and Content Weights	Style weight= 1e5 Content weight= 1	Style weight= 100 Content weight= 10
Weights for the Style Layers	block1_conv1 – 1.0 block2_conv1 – 0.75 block3_conv1 – 0.2 block4_conv1 – 0.2 block5_conv1 – 0.2	block1_conv1 – 1.0 block2_conv1 – 0.8 block3_conv1 – 0.5 block4_conv1 – 0.3 block5_conv1 – 0.1
Epochs	5000	2000
Steps per epoch (Print Out)	400	100

Pytorch Vs Tensorflow Comparison:

Origins:
Tensorflow was developed by Google Brain and was used by Google for research and production. It was later turned open source.
Pytorch was developed and used at Facebook.

Graph Definitions:
Both Tensorflow and Pytorch view models as a Directed Acyclic Graph (DAG). In Tensorflow, graphs are defined statically before a model is run, whereas Pytorch is imperative and dynamic in nature. Graph nodes can be defined, changed and executed on the go as they are run. It is more tightly integrated with python.

Visualizations:
Tensorflow provides a special tool called Tensorboard to its users for visualization. A few applications of Tensorboard include support for displaying model graphs, visualization of images, distributions, embeddings and histograms, playing audio etc. Pytorch provides a tool called Visdom. It is not feature-complete but is convenient to use.

Support for Debugging:
Tensorboard can be used for the purpose of debugging. Tensorflow also provides a special tool called tfdbg that lets users evaluate their expressions at runtime and browse through all tensors and operations. Since computation graphs in Pytorch are defined at runtime, python debugging tools like pdb, ipdb etc. can be used for the purpose of debugging.

Parallelism:
Pytorch supports declarative data parallelism. Tensorflow provides support for fine tuning every operation to be run on a specific device. Tensorflow can be used to generate more custom and specialized graphs.

Segmentation Style Transfer Network

Ananta: This is the very last step of our project and serves as a stage at which prior work can be combined. As stated above, the GrabCut algorithm is integrated with the best style transfer network with the intention of applying style transfer to either the foreground or background of a segmented image. The code is included in my Github repository, under the username annxanair. A result included in the presentation is below.

Accomplishments:

: The project was able to achieve what was proposed in the project proposal, the final model that is shown above is able to perform isolated style transfer by only applying style transfer to the foreground.

Limitations:

: Even though the project was able to achieve the project proposal goals, style transfer was only applied to the foreground object. This was not expanded to apply style transfer to the background.

Future Work:

: As indicated above I think this project can be expanded to apply style transfer to the background as well, which due to complexity of the project/analysis conducted and time restrictions this could not be done.
: For future work, some suggestions are as below:
- Expanding GrabCut algorithm to create a deep grab cut algorithm and combining this network with a style transfer network
- Using Image segmentation deep networks in combination with Style transfer for isolated style transfer
- Expanding the research on deep networks to different kinds of problems to better understand what deep networks are learning and how they are solving problems.
- Expanding the research on deep networks to different kinds of problems to see what combination of hyperparameters it takes to build stable deep networks.

Summary of contributions:
List of members contributions

Member name	Task	Comment to explain the contribution a little bit better
Ananta Nair	- Project idea development - Literature review - Wrote project proposal - Slides/presentation - Implemented the Grabcut Algorithm - Built the style transfer network in Pytorch - Performed VGG-19 analysis - Implemented the final Sematic Segmentation Style Transfer Network	- Performed literature review on topics and ideas for the project proposal. - Created research goals for project idea development - Performed literature reviews on sematic segmentation and style transfer - Coded up style transfer network for a sample image set. (code included in GitHub) - Performed an exploratory analysis on some hyperparameters of the network and other techniques to achieve best performance - Performed literature review on biases in ML/DL - Performed analysis on VGG-19 network to retrieve feature maps and saliency maps - Read original grab cut algorithm paper [1]

		- Implemented the GrabCut algorithm (code included in GitHub) - Implemented the final semantic segmentation style transfer network (code included in GitHub) - Wrote proposal proposals, report and final report - Created slides/presentation for the above topics
Medha Chippa	- Literature survey - Project proposal - Image Segmentation using K-Means Clustering - Developed a neural style transfer network using Tensorflow - Conducted a comparative study of the two style Transfer Networks - Studied the architectural differences between Tensorflow and Pytorch	- Literature survey on the neural style transfer technique and datasets was conducted - Conducted a literature review on the possible topics and ideas for the project proposal - Contributed to the Project Proposal - Designed research goals for the project idea - Conducted a Literature Survey of Image Segmentation Techniques - Performed Image Segmentation using K-Means Clustering for the purpose of visualizing the discrete portions in the image (code included on GitHub) - Developed a Neural Style Transfer Network using Tensorflow and tried with a sample image set(Code included on GitHub) - Prepared the slides for the presentation - Prepared the final report

Work conducted:

Ananta:

- Performed literature reviews and developed ideas for project.
- Performed literature review on types of image segmentation methods, read the original grabcut algorithm paper [1], and implemented the GrabCut algorithm. This is used as a base for the final semantic segmentation style transfer network.
- Performed literature review on deep neural networks and read [5]. Implemented a style transfer network using Pytorch that was used as a base for the final semantic segmentation style transfer network.
- Performed hyperparameter search and research to create the best optimized network.
- Performed analysis on VGG-19 network to better understand what deep networks learn and learn from. Printed out feature maps and saliency maps.
- Implemented the final semantic segmentation style transfer network, which is the network that combined style transfer techniques with the grabcut algorithm to produce an image that applies style transfer techniques to only an isolated portion of the image.
- Created slides for the presentation on the above material.
- Wrote project proposal, project report 2 and final project report.

Medha:

Conducted a comparative study of the two networks developed using Tensorflow and Pytorch in terms of time and space complexities, model performance etc. and studied the impact of hyperparameters on model performance. Preprocessing of images using k means clustering algorithm.

List of references:

[1] Rother, C., Kolmogorov, V., & Blake, A. (2004). " GrabCut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, *23*(3), 309-314.

[2] https://www.aperturetours.com/blog/2017/best-place-to-photograph-the-notre-dame-de-paris (Last accessed 05-04-2020)

[3] https://www.wikiart.org/en/vincent-van-gogh/the-starry-night-1888-2 (Last accessed 05-04-2020)

[4] https://abcnews.go.com/blogs/headlines/2013/01/artist-recreates-famous-van-gogh-self-portrait-as-modern-day-photograph (Last accessed 05-04-2020)

[5] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

[6] Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut Learning in Deep Neural Networks. *arXiv preprint arXiv:2004.07780*.

[7] https://www.westword.com/news/denver-residents-debate-whether-growth-has-ruined-the-queen-city-of-the-plains-8888771

[8] https://www.wikiart.org/en/vincent-van-gogh/the-starry-night-1889