School of Electronic Engineering

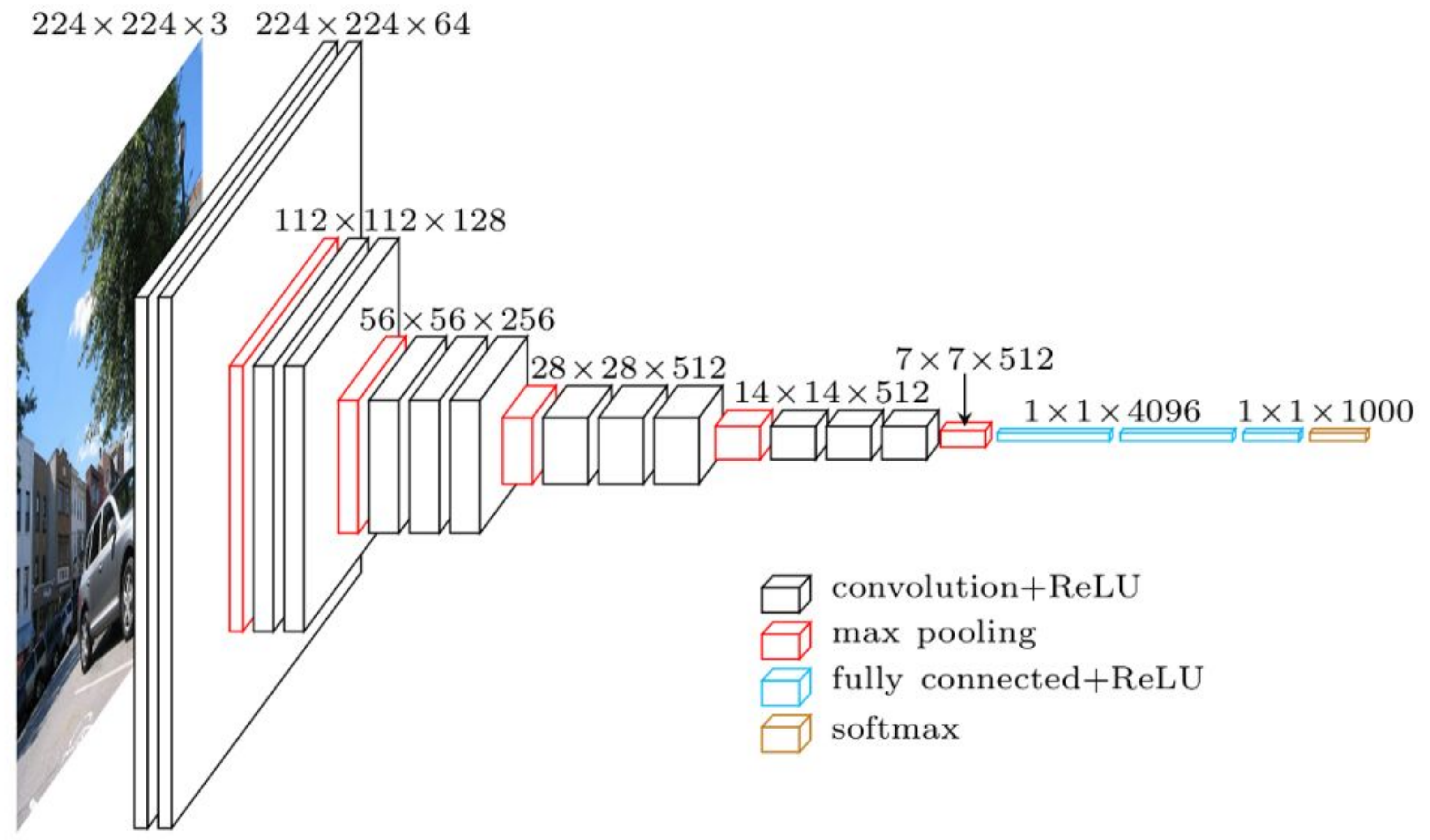
Deep Learning for Semantic Segmentation of Agricultural Imagery

Robert McConnell

Supervised by Prof. Noel O'Connor

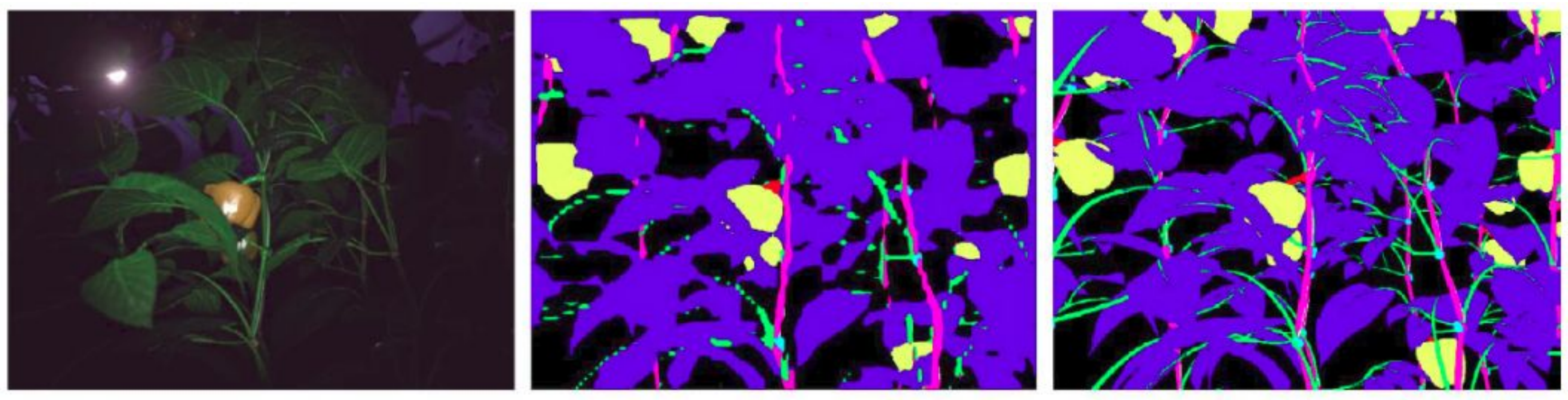
Abstract

With increasing demands on food production, Artificial Intelligence is being used in the Agricultural domain to automate visual inspection tasks and to perform robotic harvesting. Semantic Segmentation is a specific Deep Learning problem which involves labelling regions by assigning every pixel in an image to a meaningful class. The difficulty in performing Semantic Segmentation in an agricultural setting is that it can be difficult for a Convolutional Neural Network to cater for all environmental conditions such as change in seasons and weather along with the large amount of variance between crops. This project aims to address this issue by using Neural Style Transfer techniques to augment the dataset and improve generalisation.



Deep Learning and Semantic Segmentation

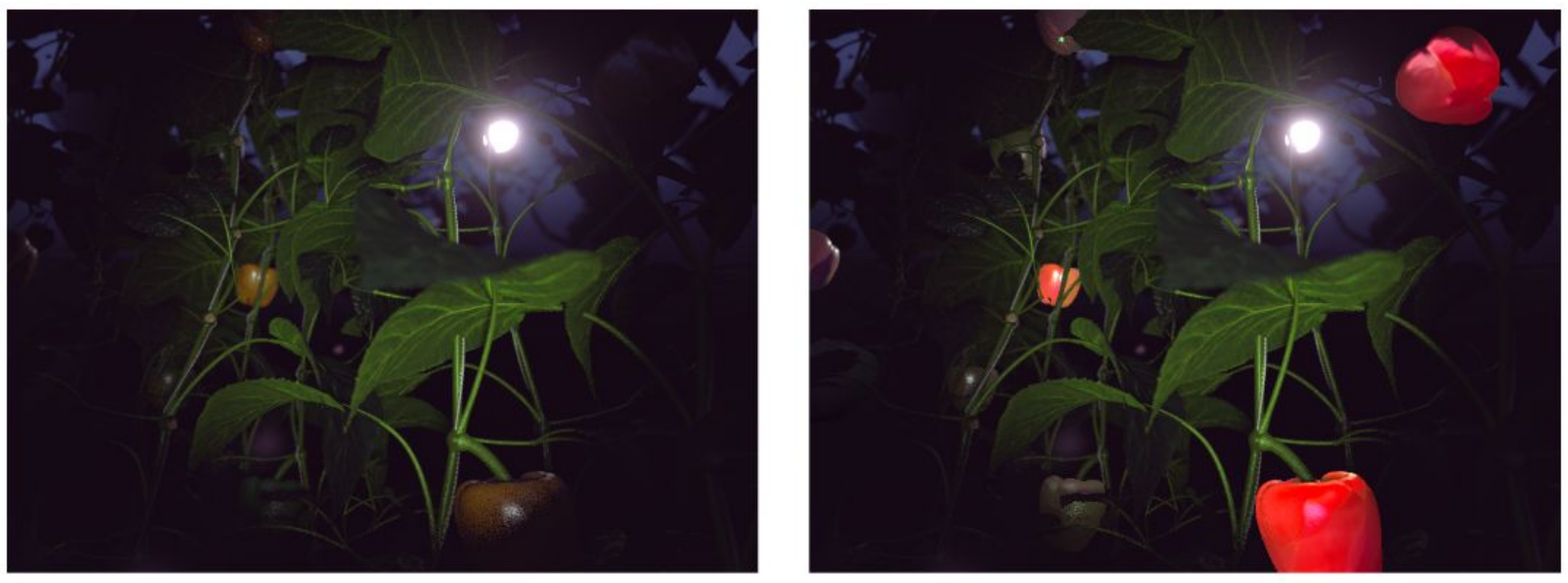
Deep Learning is a subset of machine learning based on learning features in a way similar to humans using large amounts of data and Deep Neural Networks modelled on the human brain to learn features. This project addresses a specific deep learning problem known as Semantic Segmentation which involves labelling regions by assigning every single pixel in an image to a meaningful class. The figure below shows an example from the project. The ground truth is a hand-annotated image which assigns a class label to each pixel in an image and is the ideal predicted output. The aim of the deep learning model is to find the class labels that match the ground truth with a high degree of accuracy.



(a) Input Image (b) Segmentation Result (c) Ground Truth

DeepLab Model Architecture

The baseline experiment was implemented using the DeepLab convolutional neural network model. The underlying architecture of the DeepLab model is VGG-16, shown in the model above[1]. This architecture was originally intended for global object detection but was modified for semantic segmentation by using fully convolutional layers and the use of the á trous algorithm. The Xception network backbone was used as it is a powerful network structure for server-side deployment. The Adaptive Moment Estimation (ADAM) optimisation algorithm was used with β_1=0.9, β_2= 0.999, ε=10^{-8}, a base learning rate of 0.001 for 30,000 iterations and a batch size of 4.



(a) Input Image (b) Stylised Result

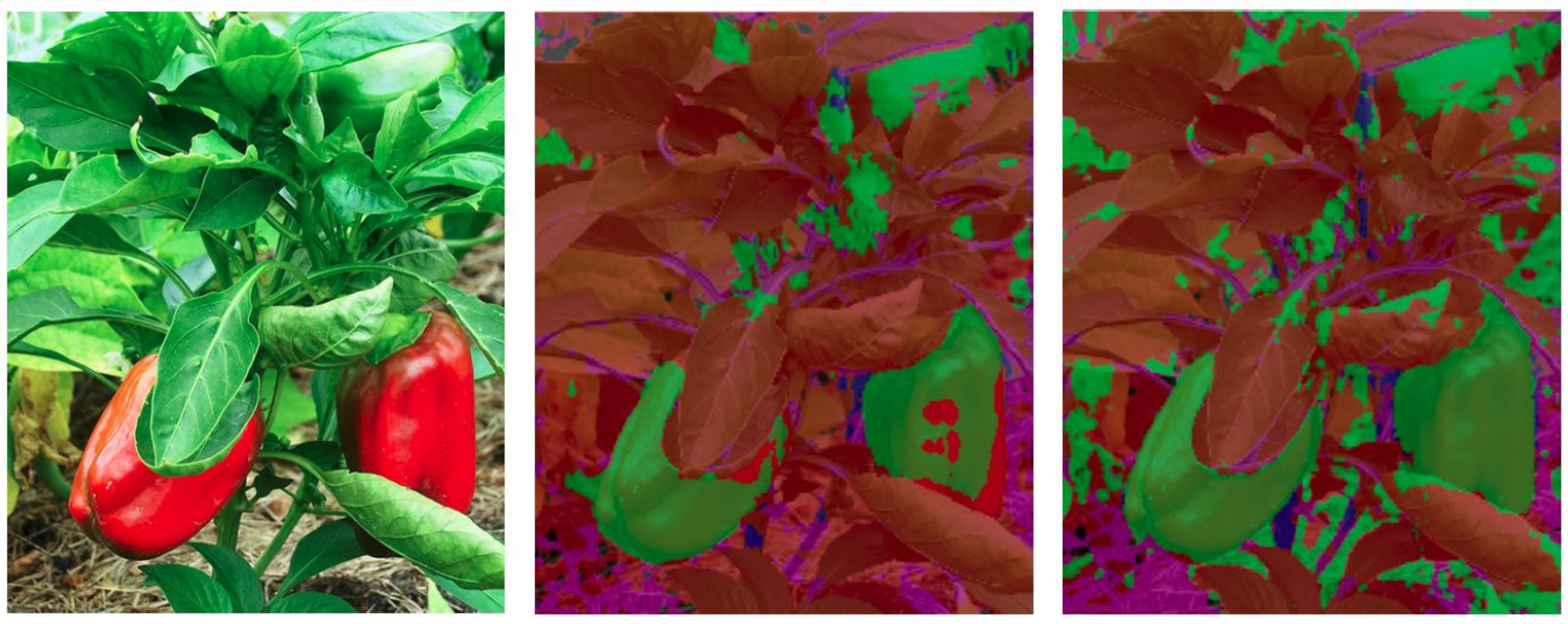
Neural Style Transfer

A key consideration when applying a type of style transfer is that the shapes and locations of the classes must remain unchanged with only their 'styles' altered. The aim was to use the same labels for the altered images as the original ones as having to relabel the images would make the task infeasible. The approach of neural style transfer uses image representations derived from Convolutional Neural Networks to make the high level information explicit and produces images that combine the content of an image with the style of another. To achieve realistic results and to not introduce 'unnecessary distortions' a regularisation term was added so the image colours only undergo affine transformations from the input to the output. The above images show an example of the style transfer where the style of a red bell pepper was mapped onto multiple yellow and green bell peppers.

Results and Performance Metrics

The aim of applying the Neural Style Transfer to the original dataset was to improve the models ability to generalise to new images. This was tested by running both models on some images from ImageNet that contained bell peppers. Based on the small set of test results, the detection of the bell peppers seems to have improved. In the DeepLab baseline tests some parts of the peppers were not fully detected, in the stylised test results the full peppers were detected. This can be seen in the example below. The efficacy of the segmentation was determined using Intersection-Over-Union (IOU) which is a method which quantifies the percentage of overlapped pixels between the target mask and the predicted output. The IOU achieved in the DeepLab baseline experiment was 0.38 but could not be measured on the ImageNet tests due to lack of semantically segmented labels.

$$IOU = \frac{target \cap prediction}{target \cup prediction}$$



(a) Input Image (b) Baseline Segmentation (c) Style Segmentation

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, 2014.

Ollscoil Chaithair Bhaile Atha Cliath
Dublin City University