**Overview**

- **Problem**: in a GUI, click on object in image -> get its outline (useful for image editing).

  - ‘DeepMask’ is great at outputting the outlines of all objects within an image from its pixels[1].
  - ‘DeepCrop’ is a modification of DeepMask which adds a ‘pixel click’ input via a 4th image plane that encodes distance from the click coordinate, to crop clicked object from image.
  - Two versions of DeepCrop are presented, with one slightly outperforming DeepMask.

**Background**

- DeepMask generates an object mask (outline) for images with a single centered object.
- The image is put through a feature extraction network, and then a ‘mask’ branch generates the outline and a ‘score’ branch scores the likeliness the input contains a centered object.
- For segmenting multiple non-centered objects in any image, the model is applied as a convolution and the highest scoring masks are kept.

  - The MS COCO dataset is used [3]. It contains a total of 123,287 images with 886,284 objects, but only 80 categories with outlines.

**DeepCropV1**

- V1 approach is to revise DeepMask by removing the score branch and adding three measures of distance from the ‘click’ pixel as additional inputs to the mask branch.
- The intuition is that providing a click coordinate removes need to assume object centrality and allows directly processing the entire image.
- This was implemented as a direct modification to the open source DeepMask code, in Torch.

**DeepCropV2**

- There are several problems with DeepCropV1:
  - The input image is significantly downsized and so there is a limit to the degree of accuracy the output can have.
  - Having to segment a non-centered object presents additional complexity
  - The feature extraction net is not provided the click location and so cannot use it.

  - A second version of DeepCrop was developed to take this into account. This revised model is the same as DeepMask, but the inputs it is trained on include a fourth image plane that represents distance from the clicked pixel.
  - This ‘DeepCropV2’ works significantly better than DeepCropV1, and also slightly outperforms DeepMask.

**Results**

- The metrics show that the DeepCropV2 model outperforms DeepMask slightly, and the larger ‘-50’ models perform best.
- A reasonable conclusion is that there is not a significant difference due to both models being trained and tested on centered inputs, and a comparison on outlining objects in whole images would yield a more significant difference.

**Discussion**

- The project presented several challenges: understanding DeepMask, learning to program in Torch, and the sheer amount of compute power these models require (3 AWS GPU instances were used for a total of $1000 AWS credit granted by Amazon for this project).
- Overall I consider the project successful, since DeepCropV1 provided interesting results as to the limitations of DeepMask and DeepCropV2 was an actually useful model.
- I expected the outcome, due to the outlined problems with DeepCropV1 and DeepCropV2 being the same as DeepMask but with additional input.

**Future**

- There are three major areas for future work:
  - Evaluating DeepCrop for whole images.
  - Integrating DeepCrop with the segmentation-refinement model SharpMask[2].
  - Further developing a GUI to test the usefulness of DeepCrop.

**References**

