DeepCrop: Directed Object Segmentation with Deep Learning

Andrey Kurenkov andreyk@stanford.edu

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 likelihood 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**.

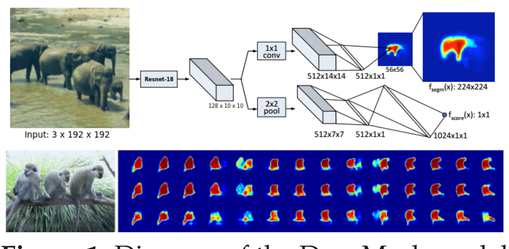

Figure 1: Diagram of the DeepMask model and an example convolution across an image

- 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.

Figure 2: Examples from COCO

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**.

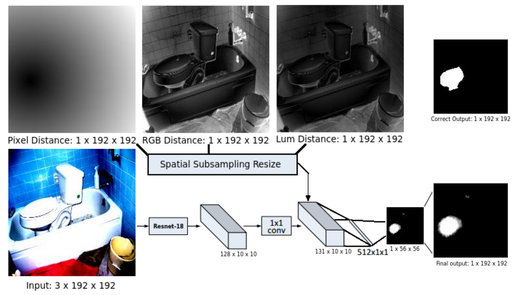

Figure 3: Visualization of the 'DeepCrop1' model

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.

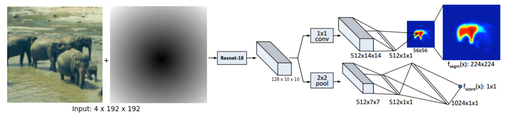

Figure 4: Diagram of the DeepCropV2

- 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

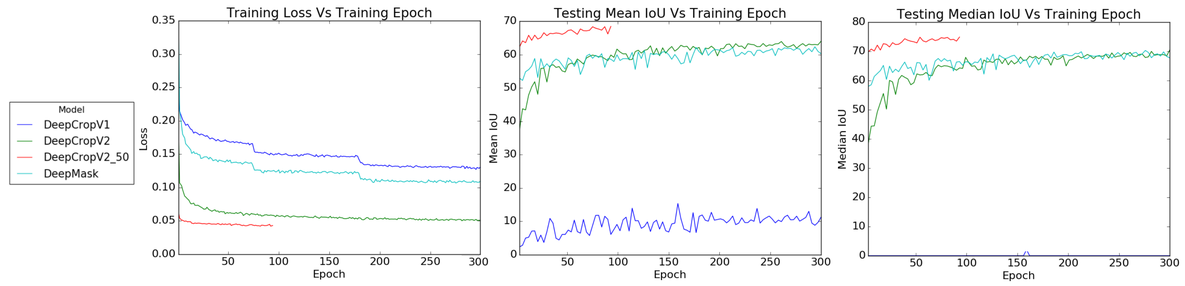

Figure 5: Train graphs. An epoch is 4000 batches, and a batch has 16 examples. DeepMaskV2_50 has a larger feature extraction net and starts with weights copied from a trained DeepMask model.

Metric / Model	Train			Test		
	Mean IoU	Median IoU	IoU@0.7	Mean IoU	Median IoU	IoU@0.7
DeepMask	61.3%	69.0%	48.5%	60.4%	68.16%	47.0%
DeepMask-50	**68.0**%	**75.1**%	**59.9**%	65.8%	73.1%	**55.4**%
DeepCropV1	15.5%	0.0%	8.0%	11.6%	0.0%	4.1%
DeepCropV2	63.8%	70.0%	50.1%	63.5%	69.6%	49.3%
DeepCropV2-50	67.5%	74.6%	58.2%	**66.1**%	**73.2**%	55.3%

Table 1: Comparison of models. 'IoU' is Intersection-over-Union, IoU@0.5 is IoU recall with 50% threshold. Inputs are as in training. Both training and test values were evaluated with 5000 samples from their respective sets.

- 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

[1] Pedro O. Pinheiro et al. Learning to segment object candidates. In *NIPS*, 2015.

[2] Pedro O. Pinheiro et al. Learning to refine object segments. In *ECCV*, 2016.

[3] Tsung-Yi Lin et al. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.