DEPARTMENT OF COMPUTER SCIENCE
AND MATHEMATICS

UNIVERSITY OF APPLIED SCIENCES MUNICH

Master's Thesis in Computer Science

Interactive Segmentation Methods

Alexander Fertig

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS

UNIVERSITY OF APPLIED SCIENCES MUNICH

Master's Thesis in Computer Science

Interactive Segmentation Methods

Author:	Alexander Fertig
Supervisor:	Prof. Dr. David Spieler
Advisor:	Advisor
Submission Date:	Submission date

I confirm that this master's thesis in computer science is my own work and I have documented all sources and material used.

Munich, Submission date   Alexander Fertig

Acknowledgments

Abstract

1. Introductions
   a) DL in Industry
   b) Application of DL and gathering Labels
2. Basics
   a) ML, Dl, CNN
   b) Semantic Segmentation (and IoU)
   c) Interactive Semantic Segmentation (Methods of comparison)
3. Methods
   a) Extreme Points
   b) IOG
4. Benchmark
   a) Motivation and structure of the Benchmark
   b) Applied Methods
   c) Evaluation (or put Evaluation as own chapter)
5. Conclusion

Contents

Acknowledgments **iii**

Abstract **iv**

1 Introduction **1**

1.1 Section 1

1.1.1 Subsection 1

2 Basics **3**

2.1 ML, DL, CNNs 3

2.1.1 Subsection 3

2.2 Semantic Segmentation 3

2.2.1 Subsection 3

2.3 Interactive Semantic Segmentation 5

2.3.1 Subsection 5

3 Methods **7**

3.1 Deep Extreme Cut 7

3.1.1 Method Description 7

3.1.2 Architecture 7

3.1.3 Refinement 7

3.1.4 Results 8

3.2 IOG 8

3.2.1 Method Description 8

3.2.2 Architecture 9

3.2.3 Refinement 10

List of Figures **12**

List of Tables **13**

Bibliography **14**

1 Introduction

1.1 Section

1.1.1 Subsection

See Table 2.3, Figure 1.1, Figure 2.2, Figure 1.3.

Table 1.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

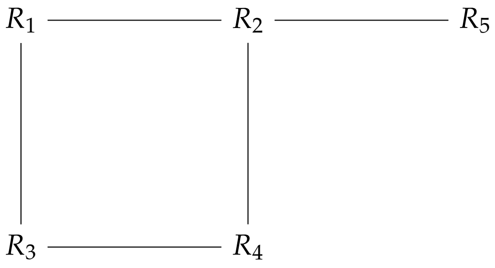

Figure 1.1: An example for a simple drawing.

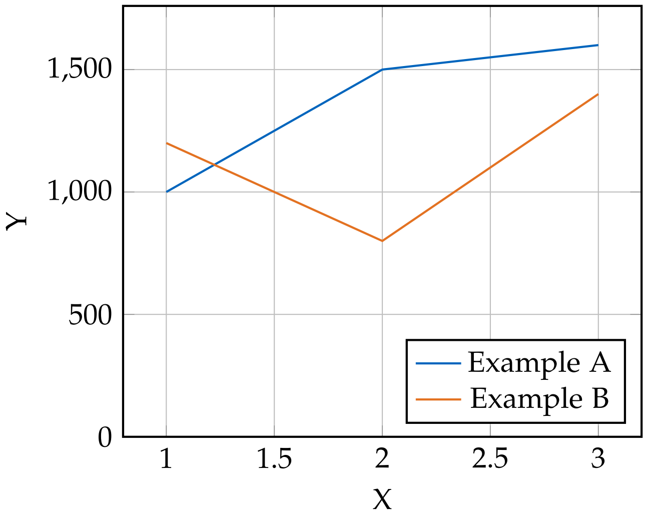

Figure 1.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 1.3: An example for a source code listing.

2 Basics

2.1 ML, DL, CNNs

Citation test [Lam94] [**Zha2020**].

2.1.1 Subsection

See Table 2.2, Figure 2.7, Figure 2.8, Figure 2.9.

Table 2.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

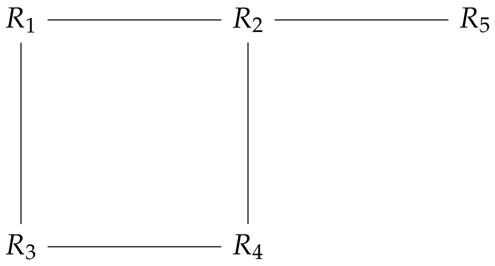

Figure 2.1: An example for a simple drawing.

2.2 Semantic Segmentation

Citation test [Lam94] [**Zha2020**].

2.2.1 Subsection

See Table 2.2, Figure 2.7, Figure 2.8, Figure 2.9.

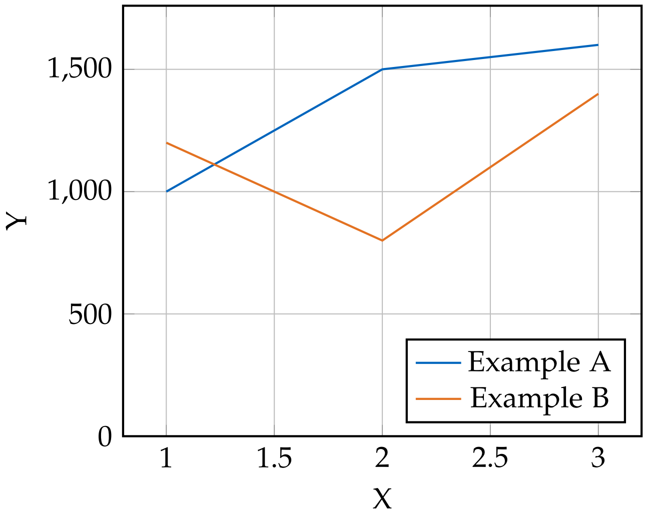

Figure 2.2: An example for a simple plot.

```sql
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 2.3: An example for a source code listing.

Table 2.2: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

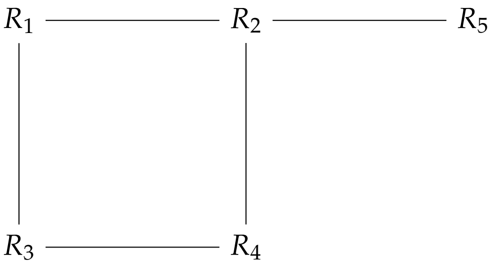

Figure 2.4: An example for a simple drawing.

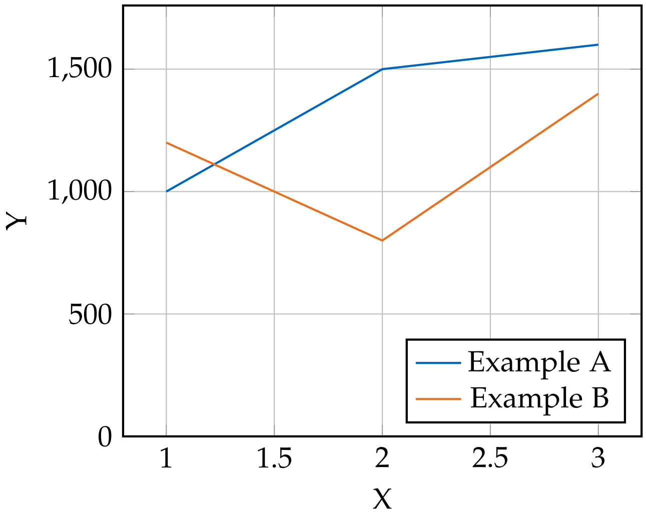

Figure 2.5: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 2.6: An example for a source code listing.

2.3 Interactive Semantic Segmentation

While semantic segmentation performs the task of segmenting an image just with the image itself, Interactive Semantic Segmentation takes advantage of additional information interactively provided by an user. The idea of this concept is to enhance the segmentation result by adding a new sort of information, that is already processed by an user. Because of this, the user input has great value for the network and provides high level guidance for the task of segmentation. Depending on the type of interaction, the receipt of the user input may be more or less elaborately, which leads to a weighing of the advantages and disadvantages. On the one side interactively provided user input may has a high level of information for the segmentation network, but on the other side user interaction may be very expensive, especially when it comes to the sizes of datasets within the context of deep learning. In the following basic concepts of interactive semantic segmentation are introduced by presenting specific methods.

2.3.1 Subsection

See Table 2.2, Figure 2.7, Figure 2.8, Figure 2.9.

Table 2.3: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

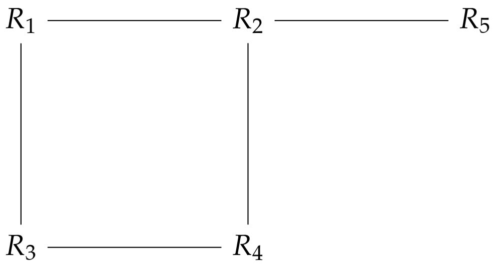

Figure 2.7: An example for a simple drawing.

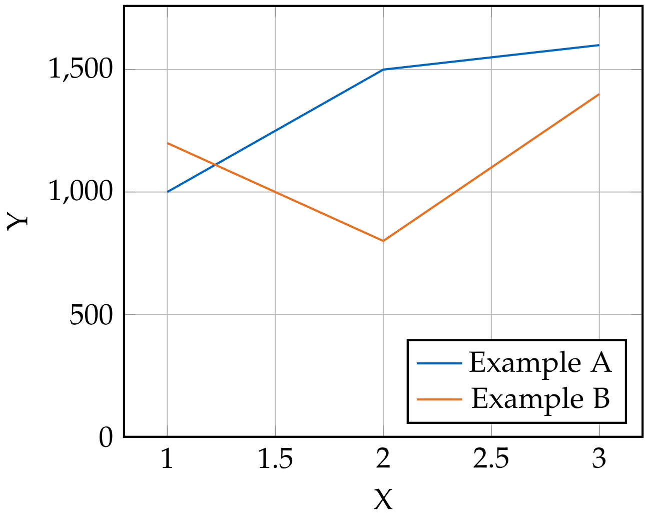

Figure 2.8: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 2.9: An example for a source code listing.

3 Methods

3.1 Deep Extreme Cut

The paper "Deep Extreme Cut: From Extreme Points to Object Segmentation" from Manisis et al. published in 2018 introduces another method to perform interactive object segmentation [Man+18].

3.1.1 Method Description

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

3.1.2 Architecture

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

3.1.3 Refinement

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

3.1.4 Results

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

3.2 IOG

The paper "Interactive Object Segmentation with Inside-Outside-Guidance"[Zha+20] published by Zhang, Liew, Wei, et al. *et al.* in 2020 provides a state-of-the-art method to perform interactive object segmentation.

3.2.1 Method Description

The execution of this method outputs a binary segmentation for a single object of interest within an image. To segment multiple objects in one image, the method has t be applied for each of them sequentially.

IOG is an interactive segmentation method and hence requires user input. The input is given by a three mouse clicks on the object's foreground and on its background. The procedure is shown in Figure 3.1 and described in the following: first, in order to form an *"almost-tight bounding box"*[Zha+20, p. 12235] two exterior clicks are set at the two diagonal locations corners of the object (top-left and bottom-right or bottom-left and top-right). Based on these two points the other two corner points are derived, which leads to four points on the background. Second, to define the object inside the bounding box a single click around the center of the desired object is made, this click is processed as foreground point. The background points *"provide "outside" guidance (indicating the background regions) while the interior click gives an "inside" guidance (indicating the foreground region), thus giving the name Inside-Outside-Guidance"*[Zha+20, p. 12235].

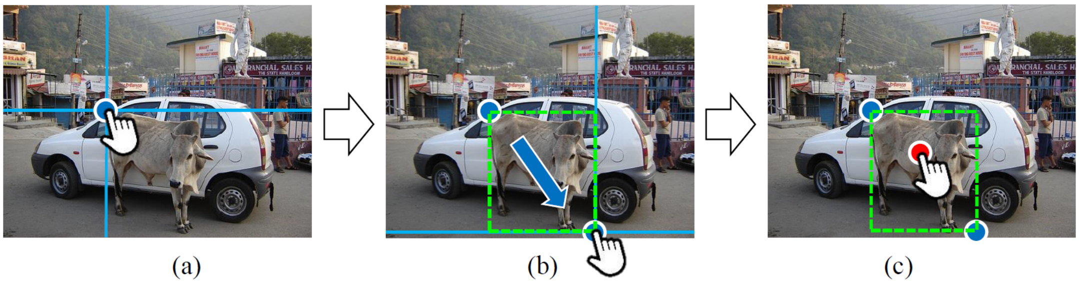

Figure 3.1: Procedure of setting the three IOG clicks [Zha+20]. Set the two background clicks (blue) at the diagonal corner locations of the object. Gather a bounding box based on the background clicks. Set a foreground point (red) at the middle of the object.

These three points are preprocessed before they are input to the actual model. To include context from the surrounding region the bounding box is enlarged by p_{box} pixels. In order to focus on the object of interest the enlarged bounding box is cropped and resized to the size of 512×512 px. For background and foreground points, a separate heatmap is created by centering a 2D Gaussian at each point with

$$Gauss = \frac{\exp -4 * \log 2}{\sigma^2} \tag{3.1}$$

The two heatmaps have the size of 512×512 px and are concatenated with the input RGB image to create a 5-channel input for the model.

3.2.2 Architecture

The architecture of the IOG method is based on a *"coarse-to-fine design"***[Zha20IOGIOG]** (see Figure 3.2), containing two main parts: the CoarseNet and the FineNet.

CoarseNet The CoarseNet contains the heavy encoder part, that mainly consists of a classifier often referred to as backbone. In IOG a ResNet-101 **[He16ResNet]** is used . This ResNet-101 is implemented without the head of fully connected layers. It contains four ResNet blocks and the fourth block outputs 2048 feature maps of the size 32×32 px. After the backbone a PSP-network is applied in order to enrich "the representation with global contextual information"[Zha+20]. The coarse prediction from the PSP-Network [Zha+17] has a spatial dimension of 32×32 px with 512 feature maps. From this onward the layers are enlarged by a four staged upsampling process to obtain the original input size of 512×512 px. During the upsampling process activations from the residual parts of the ResNet are transferred from the ResNet using so lateral connections and concatenated with the upsampled feature maps. A benefit of this architecture is the fusion of information from different network stages.

FineNet The FineNet is based on a "multi-scale fusion structure"[Zha+20]. The activations from all four stages of the upsampling process from the CoarseNet are further processed along different paths. Depending on the spatial dimension, a number of additional convolution and upsampling operations are applied in order to use *"features at deeper layers for better trade-off between accuracy and efficiency"* [Zha+20, p. 12237]. These different paths are concatenated to create the networks final layer. A sigmoid is applied to this final layer, which results in a probability map as final prediction of the IOG network. The author shows in an ablation study, that the FineNet enhances the networks IoU by 0.8%. The ablation study is performed with a ResNet-50 as backbone and PASCAL-1k [Eve+] as dataset.
This architecture especially performs well due to its application of lateral connections from different levels in order to recover local detail. The combination of layers with high localization detail with the layers, that contain high detection details, is helpful to prevent a information loss during the down- and upsampling process.

3.2.3 Refinement

If a segmentation results does not meet the user's expectations a refinement can be performed iteratively. This is done by an additional user click, which can be a fore- or background click on the region with the greatest error. In the refinement iteration of the model, this new point is processed in the same way as the initial user click positions to create a heatmap for fore- and background. These two heatmaps are combined into a two-channel input, which is processed in a so called "lightweight-branch". In this branch five convolution operations are applied and the result is concatenated with the ResNet's output of the first iteration. Hence, the ResNet does not require another execution and leads to a fast refinement process. Further, the normal IOG process is executed from the PSP-module. Zhang states that the usage of the lightweight-branch performs better than adding the refinement click into the normal 5-channel input.

In their experiments Zhang compares the IOG method to other state-of-the-art methods on different benchmarks, as shown in Figure **??**. They also evaluate the generalization abilities of IOG on unseen classes. Zhang claims that IOG outperforms all other methods.

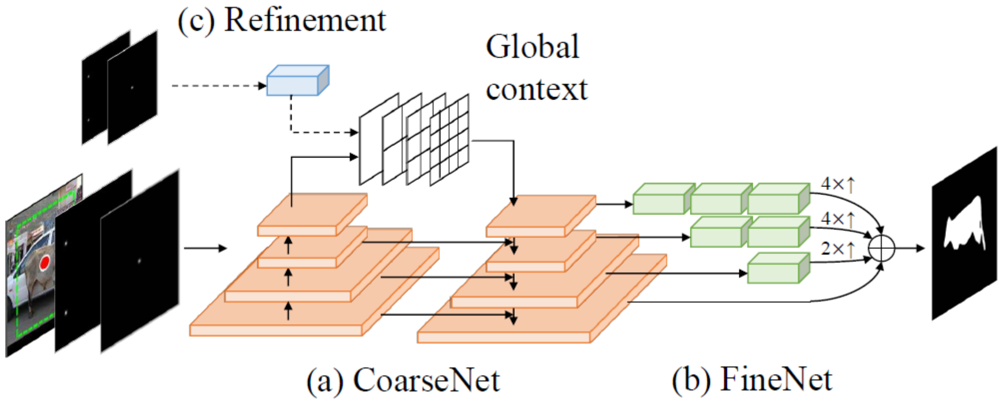

Figure 4. **Network Architecture.** (a)-(b) Our segmentation network adopts a coarse-to-fine structure similar to [14], augmented with a pyramid scene parsing (PSP) module [68] for aggregating global contextual information. (c) We also append a lightweight branch before the PSP module to accept the additional clicks input for interactive refinement.

Figure 3.2: IOG architecture (not final).

List of Figures

1.1 Example drawing 1
1.2 Example plot 2
1.3 Example listing 2

2.1 Example drawing 3
2.2 Example plot 4
2.3 Example listing 4
2.4 Example drawing 4
2.5 Example plot 5
2.6 Example listing 5
2.7 Example drawing 6
2.8 Example plot 6
2.9 Example listing 6

3.1 Procedure of setting the three IOG clicks [Zha+20]. Set the two background clicks (blue) at the diagonal corner locations of the object. Gather a bounding box based on the background clicks. Set a foreground point (red) at the middle of the object. 9
3.2 IOG architecture (not final). 11

List of Tables

1.1 Example table 1

2.1 Example table 3
2.2 Example table 4
2.3 Example table 6

Bibliography

[Eve+] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. *The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results*. http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[Lam94] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.

[Man+18] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool. "Deep Extreme Cut: From Extreme Points to Object Segmentation." In: *Computer Vision and Pattern Recognition (CVPR)*. 2018.

[Zha+17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. "Pyramid Scene Parsing Network." In: *Computer Vision and Pattern Recognition (CVPR)*. 2017.

[Zha+20] S. Zhang, J. H. Liew, Y. Wei, S. Wei, and Y. Zhao. "Interactive Object Segmentation With Inside-Outside Guidance." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 12234–12244.