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) IOG
   b) Extreme Points

4. Benchmark
   a) Motivation and structure of the Benchmark
   b) Applied Methods
   c) Evaluation (or put Evaluation as own chapter)

5. Conclusion

Contents

1 Introduction

1.1 Section

Citation test [Lam94] [Zha+20].

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.

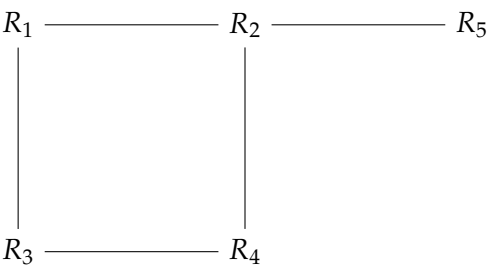| A	B	C	D
1	2	1	2
2	3	2	3

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] [Zha+20].

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.

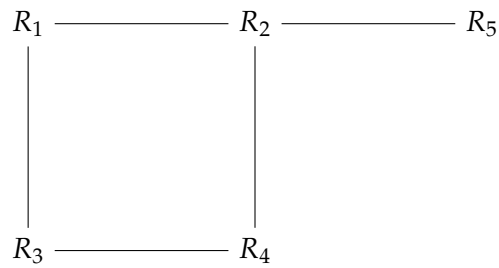| A	B	C	D
1	2	1	2
2	3	2	3

Figure 2.1: An example for a simple drawing.

2.2 Semantic Segmentation

Citation test [Lam94] [Zha+20].

2.2.1 Subsection

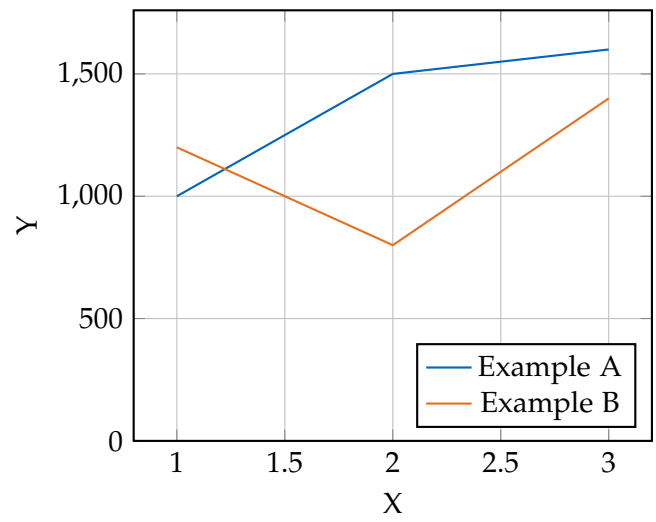See Table 2.2, Figure 2.7, Figure 2.8, Figure 2.9.

Figure 2.2: An example for a simple plot.

```
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.

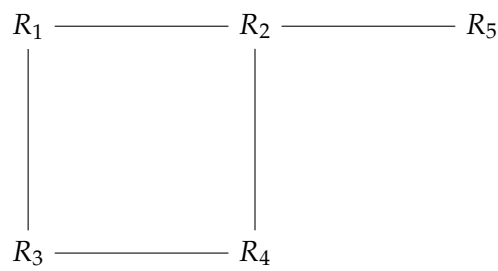| A	B	C	D
1	2	1	2
2	3	2	3

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

Citation test [Lam94] [Zha+20].

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.

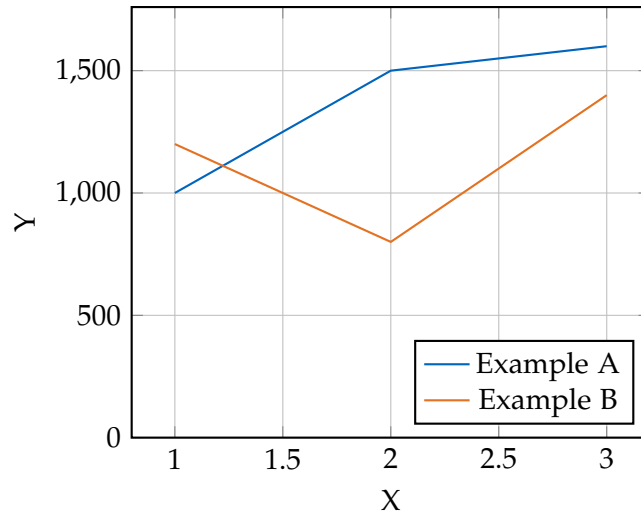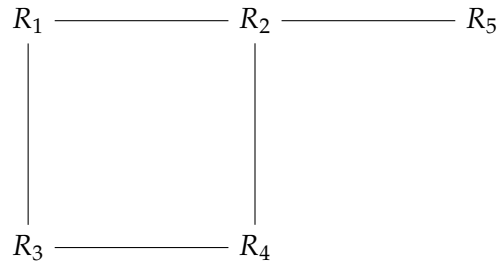| A	B	C	D
1	2	1	2
2	3	2	3

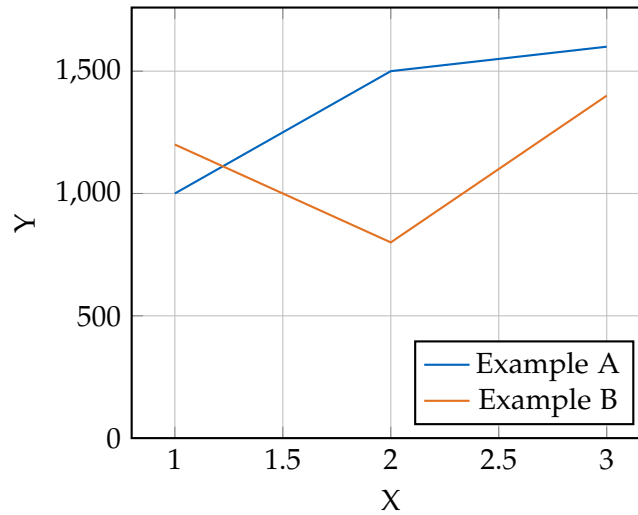Figure 2.7: An example for a simple drawing.

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 IOG

IOG is a state-of-the-art method to perform interactive object segmentation and was published by Zhang et. al in 2020, details can be read in [Zha+20].

3.1.1 Method Description

The execution of this method delivers a binary segmentation result of one object in an image. To segment multiple objects in one image, the methods requires multiple executions.

As interactive methods IOG requires a user input in the form of three mouse-clicks on fore- and background of the input image, the procedure is the following: First, in order to form an "almost-tight bounding box"[Zha+20] two exterior clicks are set at the corner locations 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 will be 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].

These three points are preprocessed before applying to the actual model. To include context from the surrounding region the bounding box is enlarged by several pixels. In order to focus on the target of interest the enlarged bounding box is cropped and resized to the size of 512x512 pixels. For background and foreground points, a separate heatmap is created by centering a 2D Gaussian around each point. The two heatmaps have the size of 512x512 pixels as well and are concatenated with the input RGB image to create a 5 channel input for the model.

3.1.2 Architecture

The architecture of the IOG method is based on a "coarse-to-fine design" [Zha+20] (see Figure 3.2), containing two main parts: the CoarseNet and the FineNet.
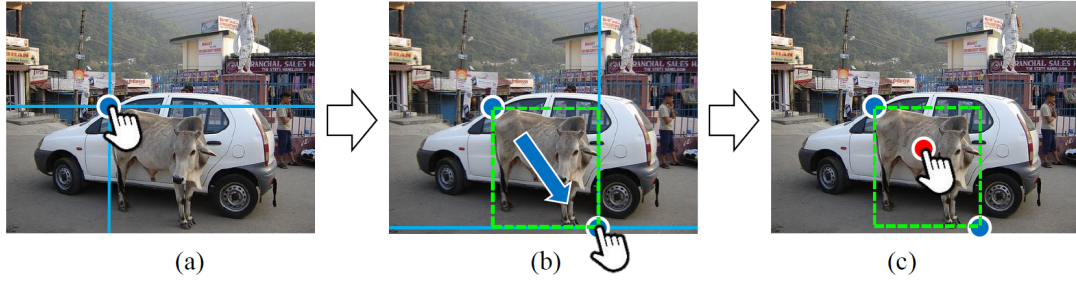
Figure 3.1: Inside-Outside guidance.

CoarseNet   The CoarseNet contains the heavy encoder part, also refereed as backbone, for what a ResNet-101 is used. This ResNet-101 is implemented with out a head of fully connected layers, but ends with the casual ResNet blocks. 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-Net has a of shape of 32x32 pixels with 512 feature maps. From this on the layers are enlarged by a four staged upsampling process to obtain the original input size of 512x512 pixels. 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 layers. This is advantageous due to the combinations of localization detail from the several ResNet layers with the deeply processed activations from the PSP-net.

FineNet   The FineNet is based on a multi-scale fusion structure. The activations from all four stages of the upsampling process from the CoarseNet are further processed in different paths. In most of them additional convolution and upsampling operations are applied in order to pressure "features at deeper layers for better trade-off between accuracy and efficiency" [Zha+20]. These different paths are concatenated to create the final prediction in the form of a probability map. The FineNet allows to recover a more precise segmentation prediction, due to the combination of activations from multiple layers.

This architecture especially stands out due to its application of lateral connection from different levels in order to recover local detail. Especially, the combination of layers with high localization detail with the layers, that contain high detection details, is precious in order to prevent the loss of information during the process of down- and upsampling.
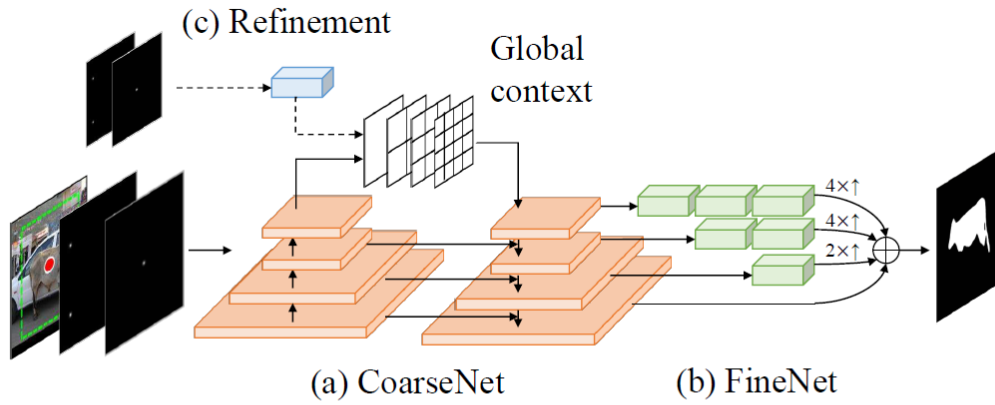
Figure 4. **Network Architecture.** (a)-(b) Our segmentation network adopts a coarse-to-fine structure similar to [14], augmented with a pyramid scene parsing (PSP) module [68] for aggregating global contextual information. (c) We also append a lightweight branch before the PSP module to accept the additional clicks input for interactive refinement.

Figure 3.2: IOG architecture.

3.1.3 Refinement

For the case a segmentation results does not meet the users expectations a refinement can be performed several times. This is realized by obtaining an additional click from the user, this can be a fore- or background click on the region with the greatest error. In the second refinement iteration of the model, this new point is processed as the original points before to create two heatmaps. They are combine into a two-channel input, which is processed in a so called "lightweight-branch" and then concatenated with the last activation of the ResNet before the PSP-module. Zhang states that this special architecture performed best in an ablation study. The further process proceeds normally and results in a probability map as result.

3.1.4 Results

In their experiments Zhang compares the IOG method to other state-of-the-art methods on different benchmarks, as shown in Figure 3.3. They also evaluate the generalization abilities of IOG on unseen classes. Zhang claims that IOG outperforms all other methods.

3.2 Extreme Points

Citation test [Lam94] [Zha+20].

3.2.1 Subsection

See Table 3.1, Figure 3.4, Figure 3.5, Figure 3.6.

Table 3.1: An example for a simple table.

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

Methods	Number of Clicks		IoU(%) @ 4 clicks	
	PASCAL@85%	GrabCut@90%	PASCAL	GrabCut
Graph cut [5]	> 20	> 20	41.1	59.3
Random walker [23]	16.1	15	55.1	56.9
Geodesic matting [2]	> 20	> 20	45.9	55.6
iFCN [66]	8.7	7.5	75.2	84.0
RIS-Net [38]	5.7	6	80.7	85.0
DEXTR [46]	4	4	91.5	94.4
Li et al. [37]	-	4.79	-	-
ITIS [45]	3.4	5.7	-	-
FCTSFN [28]	4.58	3.76	-	-
IOG-ResNet101 (**ours**)	3	3	93.2*	96.3*
IOG-ResNet101 (**ours**)	4	4	94.4	96.9

Table 1. Comparison with the state-of-the-art methods on PAS-CAL and GrabCut in terms of the number of clicks to reach a certain IoU and in terms of quality at 4 clicks. *denotes the IoU of our IOG given only 3 clicks.

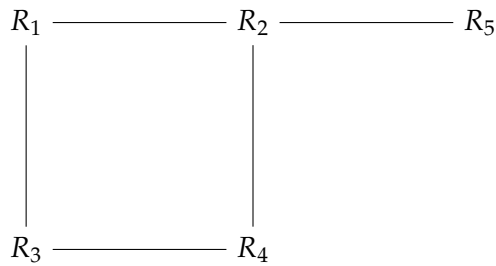Figure 3.3: Comparison with state-of-the-art methods.

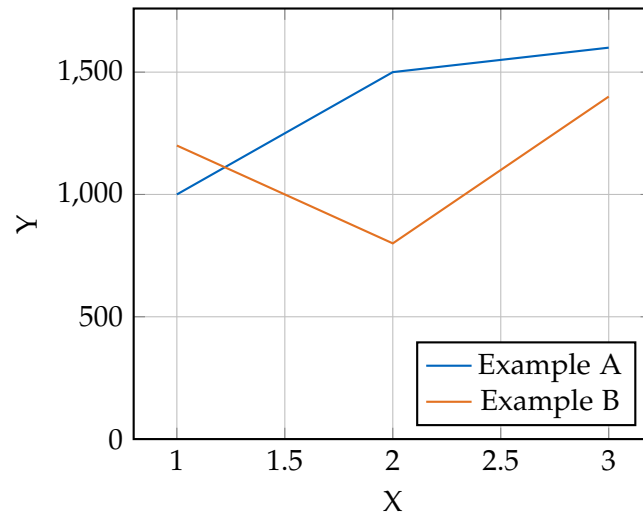Figure 3.4: An example for a simple drawing.

Figure 3.5: An example for a simple plot.

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

Figure 3.6: An example for a source code listing.

List of Figures

List of Tables

Bibliography

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

[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.