



**POLITECNICO**  
**MILANO 1863**

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE

# A Parametric Approach for Instance-level Cell Generation

TESI DI LAUREA MAGISTRALE IN  
BIOMEDICAL ENGINEERING - INGEGNERIA BIOMEDICA

Author: **Francesco Benedetto**

Student ID: 225776

Advisor: Prof. Giacomo Boracchi

Co-advisors: Roberto Basla, Luca Magri

Academic Year: 2024-25



# Abstract

Deep Learning models for instance segmentation represent the current State-of-the-Art for automating Cell Segmentation and Tracking tasks in histological imaging. These models, however, require large annotated datasets for training, which are expensive and time-consuming to obtain. Annotations rely on domain experts manually tracing cell boundaries in tissue images and videos, making data availability a major bottleneck that limits research in critical domains including cancer research, drug development, and tissue analysis. We tackle this data scarcity problem by proposing a novel generative framework for Instance-level Cell generation. Unlike existing Synthetic Data Generation methods that produce entire histological images in a single step or operate directly in pixel space, we propose a parametric approach based on *Elliptical Fourier Descriptors* (EFDs). This parametric representation enables efficient generation of single cell masks in a low-dimensional parametric domain rather than in high-dimensional spatial space, ensuring computational efficiency and scalability—fundamental requirements for reproducible and practical large-scale data augmentation. We test our method on the Fluo-N2DL-HeLa dataset, evaluating both static (single images) and dynamic generation (video sequences). For static generation, our parametric approach achieves Deep Learning performance in terms of morphological realism while maintaining computational efficiency through lightweight training. For dynamic generation, our framework outperforms existing baseline in both morphological realism and temporal dynamics of shape evolution. Our synthesis scheme provides an efficient solution to produce biologically plausible synthetic cell masks that can be integrated into complete synthesis pipelines, decreasing the experts’ effort in annotating ground truth for training Deep Learning models in the biomedical domains.

**Keywords:** Synthetic Data Generation, Cell Segmentation, *Elliptical Fourier Descriptors*, parametric representation, temporal modeling, Deep Learning



# Abstract in lingua italiana

I modelli di Deep Learning per la segmentazione di istanze rappresentano lo Stato dell'Arte attuale per l'automazione dei compiti di Segmentazione e Tracciamento Cellulare nell'imaging istologico. Tuttavia, questi modelli richiedono grandi dataset annotati per l'addestramento, che sono costosi e richiedono molto tempo per essere ottenuti. Le annotazioni si basano su esperti del settore che tracciano manualmente i confini delle cellule in immagini e video di tessuti, rendendo la disponibilità di dati un importante collo di bottiglia che limita la ricerca in domini critici tra cui la ricerca sul cancro, lo sviluppo di farmaci e l'analisi dei tessuti. Affrontiamo questo problema di scarsità di dati proponendo un nuovo framework generativo per la generazione di cellule a livello di istanza. A differenza dei metodi di Generazione di Dati Sintetici esistenti che producono intere immagini istologiche in un singolo passaggio o operano direttamente nello spazio dei pixel, proponiamo un approccio parametrico basato su *Elliptical Fourier Descriptors* (EFD). Questa rappresentazione parametrica consente una generazione efficiente di maschere di singole cellule in un dominio parametrico a bassa dimensionalità piuttosto che in uno spazio spaziale ad alta dimensionalità, garantendo efficienza computazionale e scalabilità—requisiti fondamentali per un'augmentation dei dati su larga scala riproducibile e pratica. Testiamo il nostro metodo sul dataset Fluo-N2DL-HeLa, valutando sia la generazione statica (immagini singole) che quella dinamica (sequenze video). Per la generazione statica, il nostro approccio parametrico raggiunge prestazioni di Deep Learning in termini di realismo morfologico mantenendo l'efficienza computazionale attraverso un addestramento leggero. Per la generazione dinamica, il nostro framework supera la baseline esistente sia in termini di realismo morfologico che di dinamica temporale dell'evoluzione della forma. Il nostro schema di sintesi fornisce una soluzione efficiente per produrre maschere di cellule sintetiche biologicamente plausibili che possono essere integrate in pipeline di sintesi complete, diminuendo lo sforzo degli esperti nell'annotare ground truth per l'addestramento di modelli di Deep Learning nei domini biomedici.

**Parole chiave:** Generazione di Dati Sintetici, Segmentazione Cellulare, *Elliptical Fourier Descriptors*, rappresentazione parametrica, modellazione temporale, Deep Learning



# Contents

<b>Abstract</b>	<b>i</b>
<b>Abstract in lingua italiana</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>Introduction</b>	<b>1</b>
<b>1 Chapter one</b>	<b>3</b>
1.1 Sections and subsections . . . . .	3
1.2 Equations . . . . .	3
1.3 Figures, Tables and Algorithms . . . . .	4
1.3.1 Figures . . . . .	4
1.3.2 Tables . . . . .	5
1.3.3 Algorithms . . . . .	6
1.4 Theorems, propositions and lists . . . . .	7
1.4.1 Theorems . . . . .	7
1.4.2 Propositions . . . . .	7
1.4.3 Lists . . . . .	7
1.5 Use of copyrighted material . . . . .	7
1.6 Plagiarism . . . . .	7
1.7 Bibliography and citations . . . . .	8
<b>2 Conclusions and future developments</b>	<b>9</b>
<b>Bibliography</b>	<b>11</b>

A	Appendix A	13
B	Appendix B	15
	List of Figures	17
	List of Tables	19
	List of Symbols	21
	Acknowledgements	23



# Introduction

Deep Learning (DL) models for Instance Segmentation represent the current State-of-the-Art for automating Cell Segmentation and dynamic Cell Tracking tasks in biomedical imaging. These models have demonstrated high potential in identifying and outlining individual cells within tissue samples images and videos. However, the success of these data-driven architectures critically depends on the availability of large annotated datasets for training, which represents a major limitation in the biomedical domain.

In fact, especially in histological applications, creating high quality annotations requires domain experts—typically pathologists or trained biologists—to manually create binary segmentation masks by tracing individual cell boundaries. This process requires not only specialized expertise but also significant time investment. Moreover, this task becomes even more complex in dynamic scenarios, where experts must track cells across multiple temporal frames while maintaining consistent labeling. This process is consequently expensive and time-consuming, thus limiting the availability of training data and creating a significant bottleneck in the development of DL-based segmentation models.

Such data scarcity problem is therefore particularly critical as it limits the full use of DL capabilities in quantitative biomedical research. This problem manifests in several key areas, including cancer research, drug development, and tissue analysis, where accurate Cell Segmentation and Tracking are essential for both scientific and clinical applications.

Among the various strategies employed to address this bottleneck, Synthetic Data Generation has emerged as a promising but challenging approach. Here, the idea is to automatically generate artificial training samples that can augment limited real datasets, thus reducing the labeling effort. However, the value of this approach is fundamentally dependent on the realism of the generated data, that depends primarily on synthesizing high quality individual cell shapes. Such instance-level cell geometries must accurately capture the highly variable and complex morphologies exhibited across different cellular populations, and at the same time their generation should be computationally efficient for large scale applications. Additionally, in dynamic settings, the synthetic cells should also model realistic temporal dynamics of cell shape evolution, reflecting biological processes

such as the mitotic cycle.

Although these challenges have motivated the development of various Synthetic Data Generation frameworks in recent years, available works in the literature face several critical limitations. First, most existing strategies focus primarily on synthesizing entire histological images or videos in an end-to-end manner with black-box DL models, treating generation as a single step. While this approach can produce visually realistic texturized outputs, it provides limited control and interpretability over individual cellular entities and their annotations. Second, traditional approaches that typically rely on pixel-based representations, lack computational efficiency as they process and generate data at the full image resolution, requiring important computational resources and prohibitive training times. This computational cost limits their scalability and practical deployment. Third, existing methods often struggle to accurately model the temporal dynamics of individual cell shape changes in videos, leading to synthetic data that may not faithfully represent biological processes over time.

To overcome these limitations, we introduce a novel generative framework that produces binary masks of individual cell instances using a parametric representation based on *Elliptical Fourier Descriptors* (EFDs) instead of operating in the traditional pixel domain. By encoding cell shapes as compact set of descriptors, our approach dramatically reduces the dimensionality of the generation problem, moving from high-dimensional pixel space to a low-dimensional parametric space. Beyond being computationally efficient, this methodological choice is also geometrically interpretable, as each harmonic component of the EFDs representation has a clear geometric meaning and can easily model temporal dynamics, as the compact parametric space facilitates learning temporal trajectories for video generation. These properties make our approach particularly suitable for generating cell phantoms in both static settings (individual images) and dynamic settings (temporal sequences), addressing the needs of both Cell Segmentation and Cell Tracking applications. Our experiments demonstrate that this parametric method efficiently generates morphologically variable and biologically plausible synthetic data that can effectively augment limited real annotations, achieving higher fidelity and better morphological realism than similar generative methods while maintaining superior computational efficiency and explainability. This work therefore provides a significant step towards addressing the data scarcity challenges that currently limit the application of Deep Learning in biomedical imaging.

# 1 | Chapter one

In this chapter additional useful information are reported.

## 1.1. Sections and subsections

Chapters are typically subdivided into sections and subsections, and, optionally, sub-subsections, paragraphs and subparagraphs. All can have a title, but only sections and subsections are numbered. A new section is created by the command

```
\section{Title of the section}
```

The numbering can be turned off by using `\section*{}`.

A new subsection is created by the command

```
\subsection{Title of the subsection}
```

and, similarly, the numbering can be turned off by adding an asterisk as follows

```
\subsection*{}
```

## 1.2. Equations

This section gives some examples of writing mathematical equations in your thesis.

Maxwell's equations read:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{D} = \rho, \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}, \\ \nabla \cdot \mathbf{B} = 0, \\ \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}. \end{array} \right. \quad \begin{array}{l} (1.1a) \\ (1.1b) \\ (1.1c) \\ (1.1d) \end{array}$$

Equation (1.1) is automatically labeled by `cleveref`, as well as Equation (1.1a) and Equation (1.1c). Thanks to the `cleveref` package, there is no need to use `\eqref`.

Remember that Equations have to be numbered only if they are referenced in the text.

Equations (1.2), (1.3), (1.4), and (1.5) show again Maxwell's equations without brace:

$$\nabla \cdot \mathbf{D} = \rho, \quad (1.2)$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}, \quad (1.3)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.4)$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}. \quad (1.5)$$

Equation (1.6) is the same as before, but with just one label:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{D} = \rho, \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}, \\ \nabla \cdot \mathbf{B} = 0, \\ \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}. \end{array} \right. \quad (1.6)$$

### 1.3. Figures, Tables and Algorithms

Figures, Tables and Algorithms have to contain a Caption that describe their content, and have to be properly referred in the text.

#### 1.3.1. Figures

For including pictures in your text you can use `TikZ` for high-quality hand-made figures, or just include them as usual with the command

```
\includegraphics[options]{filename.xxx}
```

Here xxx is the correct format, e.g. `.png`, `.jpg`, `.eps`, ....



Figure 1.1: Caption of the Figure to appear in the List of Figures.

Thanks to the `\subfloat` command, a single figure, such as Figure 1.1, can contain multiple sub-figures with their own caption and label, e.g. Figure 1.2a and Figure 1.2b.

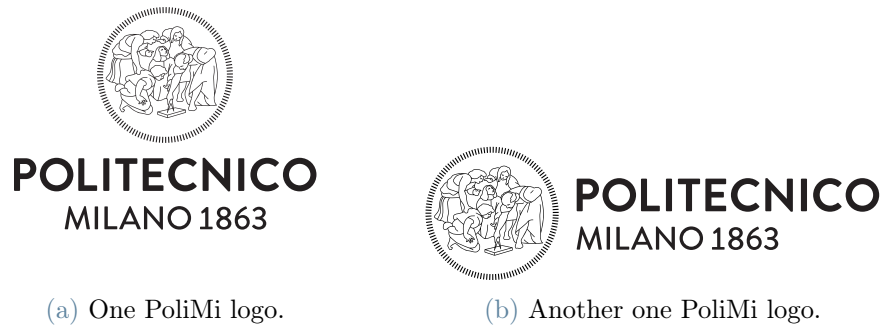


Figure 1.2: This is a very long caption you don't want to appear in the List of Figures.

### 1.3.2. Tables

Within the environments `table` and `tabular` you can create very fancy tables as the one shown in Table 1.1.

Title of Table (optional)			
	column 1	column 2	column 3
row 1	1	2	3
row 2	$\alpha$	$\beta$	$\gamma$
row 3	alpha	beta	gamma

Table 1.1: Caption of the Table to appear in the List of Tables.

You can also consider to highlight selected columns or rows in order to make tables more

readable. Moreover, with the use of `table*` and the option `bp` it is possible to align them at the bottom of the page. One example is presented in Table 1.2.

	column1	column2	column3	column4	column5	column6
row1	1	2	3	4	5	6
row2	a	b	c	d	e	f
row3	$\alpha$	$\beta$	$\gamma$	$\delta$	$\phi$	$\omega$
row4	alpha	beta	gamma	delta	phi	omega

Table 1.2: Highlighting the columns

	column1	column2	column3	column4	column5	column6
row1	1	2	3	4	5	6
row2	a	b	c	d	e	f
row3	$\alpha$	$\beta$	$\gamma$	$\delta$	$\phi$	$\omega$
row4	alpha	beta	gamma	delta	phi	omega

Table 1.3: Highlighting the rows

### 1.3.3. Algorithms

Pseudo-algorithms can be written in  $\text{\LaTeX}$  with the `algorithm` and `algorithmic` packages. An example is shown in Algorithm 1.1.

---

Algorithm 1.1 Name of the Algorithm

---

```

1: Initial instructions
2: for for – condition do
3:   Some instructions
4:   if if – condition then
5:     Some other instructions
6:   end if
7: end for
8: while while – condition do
9:   Some further instructions
10: end while
11: Final instructions

```

---

## 1.4. Theorems, propositions and lists

### 1.4.1. Theorems

Theorems have to be formatted as:

**Theorem 1.1.** *Write here your theorem.*

*Proof.* If useful you can report here the proof.

### 1.4.2. Propositions

Propositions have to be formatted as:

**Proposition 1.1.** *Write here your proposition.*

### 1.4.3. Lists

How to insert itemized lists:

- first item;
- second item.

How to insert numbered lists:

1. first item;
2. second item.

## 1.5. Use of copyrighted material

Each student is responsible for obtaining copyright permissions, if necessary, to include published material in the thesis. This applies typically to third-party material published by someone else.

## 1.6. Plagiarism

You have to be sure to respect the rules on Copyright and avoid an involuntary plagiarism. It is allowed to take other persons' ideas only if the author and his original work are clearly mentioned. As stated in the Code of Ethics and Conduct, Politecnico di Milano *promotes the integrity of research, condemns manipulation and the infringement of*

*intellectual property*, and gives opportunity to all those who carry out research activities to have an adequate training on ethical conduct and integrity while doing research. To be sure to respect the copyright rules, read the guides on Copyright legislation and citation styles available at:

<https://www.biblio.polimi.it/en/tools/courses-and-tutorials>

You can also attend the courses which are periodically organized on "Bibliographic citations and bibliography management".

## 1.7. Bibliography and citations

Your thesis must contain a suitable Bibliography which lists all the sources consulted on developing the work. The list of references is placed at the end of the manuscript after the chapter containing the conclusions. We suggest to use the BibTeX package and save the bibliographic references in the file `Thesis_bibliography.bib`. This is indeed a database containing all the information about the references. To cite in your manuscript, use the `\cite{}` command as follows:

*Here is how you cite bibliography entries: [2], or multiple ones at once: [3, 4].*

The bibliography and list of references are generated automatically by running BibTeX [1].



## 2 | Conclusions and future developments

A final chapter containing the main conclusions of your research/study and possible future developments of your work have to be inserted in this chapter.



# Bibliography

- [1] CTAN. BiBTeX documentation, 2017. URL <https://ctan.org/topic/bibtex-doc>.
- [2] D. E. Knuth. Computer programming as an art. *Commun. ACM*, pages 667–673, 1974.
- [3] D. E. Knuth. Two notes on notation. *Amer. Math. Monthly*, 99:403–422, 1992.
- [4] L. Lamport. *LaTeX: A Document Preparation System*. Pearson Education India, 1994.



# A | Appendix A

If you need to include an appendix to support the research in your thesis, you can place it at the end of the manuscript. An appendix contains supplementary material (figures, tables, data, codes, mathematical proofs, surveys, . . . ) which supplement the main results contained in the previous chapters.



## B | Appendix B

It may be necessary to include another appendix to better organize the presentation of supplementary material.





## List of Figures

1.1	Caption of the Figure to appear in the List of Figures. . . . .	5
1.2	Shorter caption . . . . .	5



# List of Tables

1.1	Caption of the Table to appear in the List of Tables. . . . .	5
1.2	Highlighting the columns . . . . .	6
1.3	Highlighting the rows . . . . .	6



## List of Symbols

Variable	Description	SI unit
$\boldsymbol{u}$	solid displacement	m
$\boldsymbol{u}_f$	fluid displacement	m



# Acknowledgements

Here you might want to acknowledge someone.

