



---

## Calculating Optimal Feasible Decompositions of Simple Polygons for Identifying Biomarkers

---

### Leonie Selbach

Department of Computer Science and Center for Protein Diagnostics  
Ruhr University Bochum, Germany  
[leonie.selbach@rub.de](mailto:leonie.selbach@rub.de)

### Tobias Kowalski

Department of Biophysics and Center for Protein Diagnostics  
Ruhr University Bochum, Germany  
[tobias.kowalski@rub.de](mailto:tobias.kowalski@rub.de)

### Klaus Gerwert

Department of Biophysics and Center for Protein Diagnostics  
Ruhr University Bochum, Germany  
[klaus.gerwert@rub.de](mailto:klaus.gerwert@rub.de)

### Maike Buchin

Department of Computer Science, Ruhr University Bochum  
Germany  
[maike.buchin@rub.de](mailto:maike.buchin@rub.de)

### Axel Mosig

Department of Biophysics and Center for Protein Diagnostics  
Ruhr University Bochum, Germany  
[axel.mosig@rub.de](mailto:axel.mosig@rub.de)

### ABSTRACT

*In identifying biomarkers and molecular profiling of diseases, laser capture microdissection serves as an exceptionally efficient method for isolating disease-specific areas from intricate, varied tissue specimens. These areas must be broken down into manageable pieces that meet specific size and shape requirements for successful extraction. We approach the challenge of constrained shape decomposition by calculating optimal feasible decompositions of simple polygons. Our framework is based on a skeleton-oriented method and provides an*

*algorithmic structure that supports the integration of different feasibility criteria and optimization objectives. Driven by our application focus, we explore various constraints and analyze the resulting fragmentations. Additionally, we implement our technique on lung tissue samples and demonstrate its benefits over a heuristic decomposition strategy.*

**Keywords:** Polycons, Biomarkers, Optimization algorithms

**Received:** 25 August 2024, Revised 27 November 2024, Accepted 4 December 2024

**Copyright:** with Authors

## 1. Introduction

Laser capture microdissection (LCM) [14] is a highly effective approach to extract specific cell populations from complex, heterogeneous tissue samples and has been used extensively in the context of biomarker discovery [15] as well as the molecular characterization of diseases [17]. Since LCM separates homogeneous and disease-specific regions from their heterogeneous and unspecific surrounding tissue regions, the characterizations obtained from genomic, transcriptomic or proteomic characterizations of samples processed with LCM provide more accurate molecular markers of diseases [9, 20]. With LCM being used more and more commonly in clinical studies, there is a need to automate all procedures involved in sample processing.

### 1.1 Problem Statement

In our present contribution, we address one central problem of processing samples with LCM. Namely that the regions of interest (ROIs) consist of complex shapes of varying size which in general cannot be extracted in one piece from the tissue sample. Rather, the ROI needs to be fragmented into small subregions that satisfy certain constraints in size and morphology: Fragments must not exceed certain limits of minimal or maximal size and should be of approximately round shape. If a fragment does not meet these constraints, it cannot be properly extracted from the surrounding tissue. This negatively affects the sample quality and thus compromises the advantages of LCM-based sample preparation.

By interpreting each connected component of the ROI as a simple polygon, we can model this problem of constrained shape decomposition as the computation of optimal feasible decompositions of polygons. The constraints can be modeled as certain feasibility criteria and optimization goals. Our decomposition method utilizes a skeleton of the shape and follows a dynamic approach. Specifically, we restrict our cuts to certain line segments based on the skeleton. This not only results in simple cuts but also in a flexible framework that allows to integrate various criteria. With respect to our application, we consider different criteria regarding for example the area, convexity or fatness of a polygon or the length of inserted cuts. However, other constraints as well as combinations of multiple criteria are possible.

### 1.2 Application

Our contribution is motivated by an application introduced in [15] in which the ROI to be dissected from the tissue sample is identified using label-free hyperspectral infrared microscopy. In this approach, an infrared microscopic image of the sample yields infrared pixel spectra at a spatial resolution of about 5  $\mu m$ . A previously trained random forest classifier assigns each pixel spectrum to one tissue component such as *healthy* or *diseased*, with the *diseased* class being further subdivided into *inflamed tissue* as well as several subtypes of

thoracic tumors. The general sample preparation task in the context of LCM is to dissect all tumor regions (or all regions identified as one specific tumor subtype) from a sample. While our current contribution deals with the specific context of label-free infrared microscopy, our shape decomposition approach equally applies more broadly to LCM in the context of other microscopic modalities, most notably H&E stained images [9] for which recent digital pathology approaches facilitate reliable computational identification of disease specific regions [26, 28].

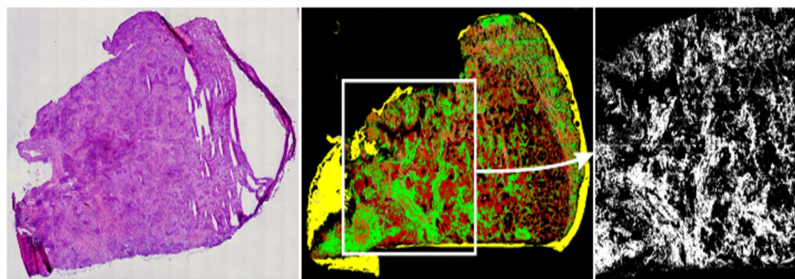
## 2. Related Work

Polygon decomposition is an important tool in computational geometry, as many algorithms work more efficiently on certain polygon classes, for example convex polygons [16]. Moreover, polygon decomposition is frequently used in applications such as pattern recognition or image processing [16]. Object recognition, biomedical image analysis and shape decomposition are typical areas of application that utilize skeletons [23]. Skeletons are oftentimes used to analyze the morphology of a given shape and work especially well on elongated structures, such as vessels [11], pollen tubes [27] or neuron images [19, 24]. There are several shape decomposition methods based on the skeleton or some other medial representation of a shape. However, most of these methods are designed for object recognition and thus focus on decomposing a shape into “natural” or “meaningful” parts [22, 18, 21]. In some approaches even decompositions with overlapping parts are allowed [10, 25]. None of the established decomposition methods facilitate a straightforward introduction of adjustable size and shape constraints as needed for our application.

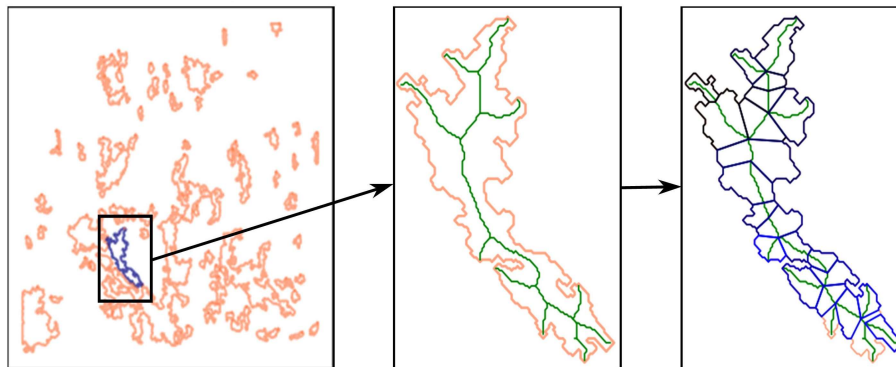
We utilize the skeleton for two main reasons: it is well-established to represent shape morphology and has proved useful for shape decomposition previously. As cancerous tissue regions often present themselves as highly complex and ramified shapes, we apply the skeleton to obtain a morphological representation, based on which we compute a decomposition that includes the morphological features.

## 3. Methods

As an input, we receive a binary mask of a microscopic slide with the region of interest (ROI) as the foreground (Fig. 1a). After a preprocessing step, the foreground is reduced to connected components without holes. We can interpret each of these components as a polygon by taking each boundary pixel as a polygon vertex. The goal is to produce a decomposition of each component in such a way that the fragments fulfill certain constraints in size and morphology. We compute this fragmentation using a skeleton-based approach for polygon decomposition (Fig. 1b).



(a) Selection of one region of interest in a histopathological tissue sample (H&E-stained image of a subsequent sample on the left) in which different regions have been identified using the method in [15]



(b) After a preprocessing step, each connected component is a simple polygon without holes and is decomposed using the method presented in this paper

Figure 1. Decomposition of ROI polygons in a histopathological tissue sample

### 3.1 Skeleton of a Shape

Our approach is based on the medial axis or skeleton of the polygon. The medial axis of a shape is the set of points that have more than one closest point on the shape's boundary. The medial axis was introduced for the description of biological shapes [3, 4] but is now widely used in other applications such as object recognition, medical image analysis and shape decomposition [23]. An important property is that the medial axis represents the object and its geometrical and topological characteristics while having a lower dimension [8, 29].

Formally, the medial axis of a shape  $D$  is defined as the set of centers of maximal disks in  $D$ . A closed disk  $B \subset D$  is maximal in  $D$  if every other disk that contains  $B$  is not contained in  $D$ . A point  $s$  is called *skeleton point* if it is the center of a maximal disk  $B(s)$  (see Fig. 2). For a skeleton point  $s$ , we call the points where  $B(s)$  touches the boundary the *contact points* – every skeleton point has at least two contact points. A skeleton  $S$  is given as a graph consisting of connected arcs  $Sk$ , which are called *skeleton branches* and meet at branching points. Given a simple polygon without holes the skeleton is an acyclic graph.

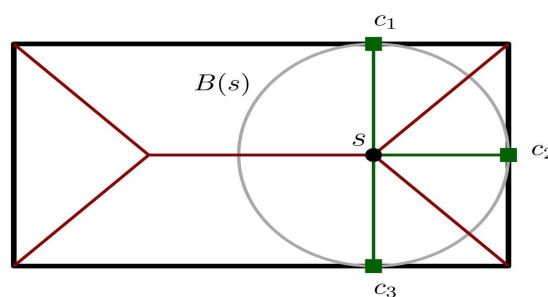


Figure 2. Medial axis of a simple shape. The skeleton point  $s$  is a branching point with three contact points  $c_1$ ,  $c_2$ ,  $c_3$

There are various methods for the computation of the medial axis in practice [23]. In general, the medial axis is very sensitive to noise in an object's boundary. This is a problem that often occurs in digital images and leads to spurious skeleton branches. Therefore, many approaches apply some kind of pruning method to remove those branches. In our application, we have a discrete input and a discrete output is expected. Because of that, we use a skeletonization algorithm that computes a discrete and pruned skeleton, which consists of a finite number of

skeleton pixels as our skeleton points [2]. Furthermore, the computed skeleton has the property that every branching point has a degree of exactly three.

### 3.2 Skeleton-based Polygon Decomposition

We consider the following problem: Given a simple polygon  $P$ , compute an optimal feasible decomposition of  $P$ . A decomposition is *feasible* if every subpolygon is feasible, in the sense that it fulfills certain conditions on for instance its size and shape. We present an algorithmic framework that allows the integration of various criteria for both feasibility and optimization, which are discussed later. As for now, we only consider criteria that are locally evaluable.

In our skeleton-based approach, we allow only cuts that are line segments between a skeleton point and its corresponding contact points. Thus, the complexity of our algorithm mainly depends on the number of skeleton points rather than the number of boundary points of the polygon. Every subpolygon in our decomposition is generated by two or more skeleton points. We present two decomposition algorithms: One in which we restrict the subpolygons to be generated by exactly two skeleton points and a general method. In the first case, each subpolygon belonging to a skeleton branch can be decomposed on its own and in the second case the whole polygon is decomposed at once.

#### Decomposition based on Linear Skeletons

First, let us consider the restriction that the subpolygons are generated by exactly two skeleton points. In this case, the corresponding skeleton points have to be on one skeleton branch  $Sk$ . In our work, a branching point belongs to exactly three branches and has exactly three contact points. When a branching point generates a subpolygon with another skeleton point on an adjacent branch, we choose those two line segments that correspond to this branch. Due to the Domain Decomposition Lemma (see Fig. 3, proof in [8]) and the following corollary, we can decompose each skeleton branch on its own.

**Theorem 1** (Domain Decomposition Lemma). Given a domain  $D$  with skeleton  $S(D)$ , let  $p \in S(D)$  be some skeleton point and let  $B(p)$  be the corresponding maximal disk. Suppose  $A_1, A_2, \dots, A_k$  are the connected components of  $D \setminus B(p)$ . Define  $D_i = A_i \cup B(p)$  for all  $i$ . Then:

$$S(D) = \bigcup_{i=1}^k S(D_i).$$

Moreover, we have

$$S(D_i) \cap S(D_j) = p \quad \forall i = j.$$

**Corollary 2** Let  $p \in S(D)$  and  $A_1, A_2, \dots, A_k$  be as above. For each skeleton point  $q \neq p$  exists an  $i$  such that all contact points of  $q$  are contained in  $A_i$ .

Let  $S_k$  be a skeleton branch with a linear skeleton of size  $n_k$  and let  $P_k$  be the polygon belonging to this branch. By  $P_k(i, j)$ , we denote a subpolygon that is generated by two skeleton points  $i$  and  $j$  on  $S_k$  (see Figure 4). Thus, we have  $P_k(1, n_k) = P_k$ . First, we consider the decision problem. Note that there exists a feasible decomposition of a polygon  $P_k(i, n_k)$  if either

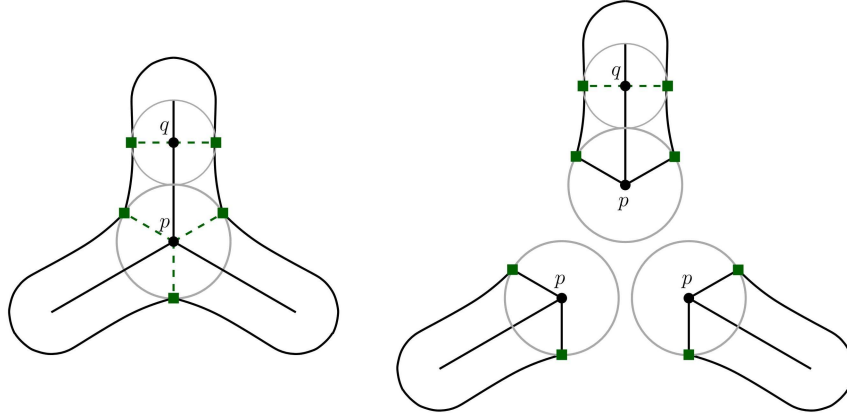
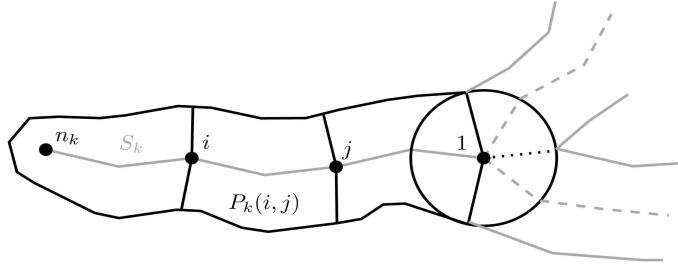


Figure 3. Domain Decomposition Lemma

- $P_k(i, n_k)$  is feasible or
- There exists  $j > i$  such that  $P_k(i, j)$  is feasible and  $P_k(j, n_k)$  has a feasible decomposition.

Thus, we can solve the problem by using dynamic programming and use backtracking to compute the corresponding decomposition. We can include different optimization criteria by choosing an optimal point  $j$ .

**Lemma 3** Given a subpolygon  $P_k$  with a linear skeleton  $S_k$  consisting of  $n_k$  points, one can compute a feasible decomposition of  $P_k$  based on  $S_k$  in time  $O(n_k^2 F)$ , with  $F$  being a factor depending on the feasibility criteria.


 Figure 4. Polygon  $P_k$  belonging to a skeleton branch  $S_k$ 

**Proof** For every skeleton point  $i$ , for  $i = n_k - 1$  down to 1, we compute  $X(i)$  such that  $X(i)$  equals True if there exists a feasible decomposition of  $P_k(i, n_k)$ . To compute  $X(i)$ , we consider  $O(n_k)$  other values  $X(i')$  for  $i < j \leq n_k$  and check in time  $O(F)$  if the polygon  $P_k(i, j)$  is feasible. The correctness follows inductively.

The factor  $F$  is determined by the runtime it takes to decide whether a subpolygon is feasible. This factor might depend on for instance the number of points in the skeleton or in the boundary of the polygon. We discuss examples in the following section. After computing decompositions for each subpolygon corresponding to a skeleton branch, we can combine those to obtain a decomposition of the entire polygon. This leads to the following result.

**Theorem 4** Given a simple polygon  $P$  with skeleton  $S$  consisting of  $n$  points, one can compute a feasible decomposition of  $P$  based on the skeleton branches of  $S$  in time  $\mathcal{O}(n^2F)$ , with  $F$  being a factor depending on the feasibility criteria.

Note that there might not exist a feasible decomposition of the entire polygon or for certain subpolygons. By using this method, we are able to obtain partial decompositions. Thus, this approach can be favorable in practice.

### General Decomposition

In the general setting, subpolygons are allowed to be generated by more than two skeleton points. In this paper, we will briefly explain the idea of our method (see [5] for a more detailed description and the corresponding formulas). Recall that our skeleton is an acyclic graph consisting of a finite number of vertices, i.e. skeleton points. The skeleton computed for our application (method of Bai et al. [2]) has the property that the maximal degree of a skeleton point is three. We select one branching point as a root and consider a rooted skeleton tree. Since branching points belong to three different branches, these nodes are duplicated in the skeleton tree such that each node corresponds to the cut edges on the respective branch (see Fig. 5). Our method and its runtime are based on two main observations.

**Observation 5** The maximal number of skeleton points that can generate a subpolygon is equal to the number of endpoints in the skeleton, i.e. the number of leaves in the skeleton tree.

**Observation 6** Every subpolygon can be represented as the union of subpolygons generated by just two skeleton points.

Let  $i$  be a node in the skeleton tree and  $T_i$  the subtree rooted in  $i$ . By  $P(i)$ , we denote the subpolygon ending in the skeleton point  $i$ . This polygon corresponds to the subtree  $T_i$  in the given tree representation (see Fig. 6). For each node  $i$  (bottom-up), we compute if there exists a feasible decomposition of the polygon  $P(i)$ . Such a decomposition exists if either

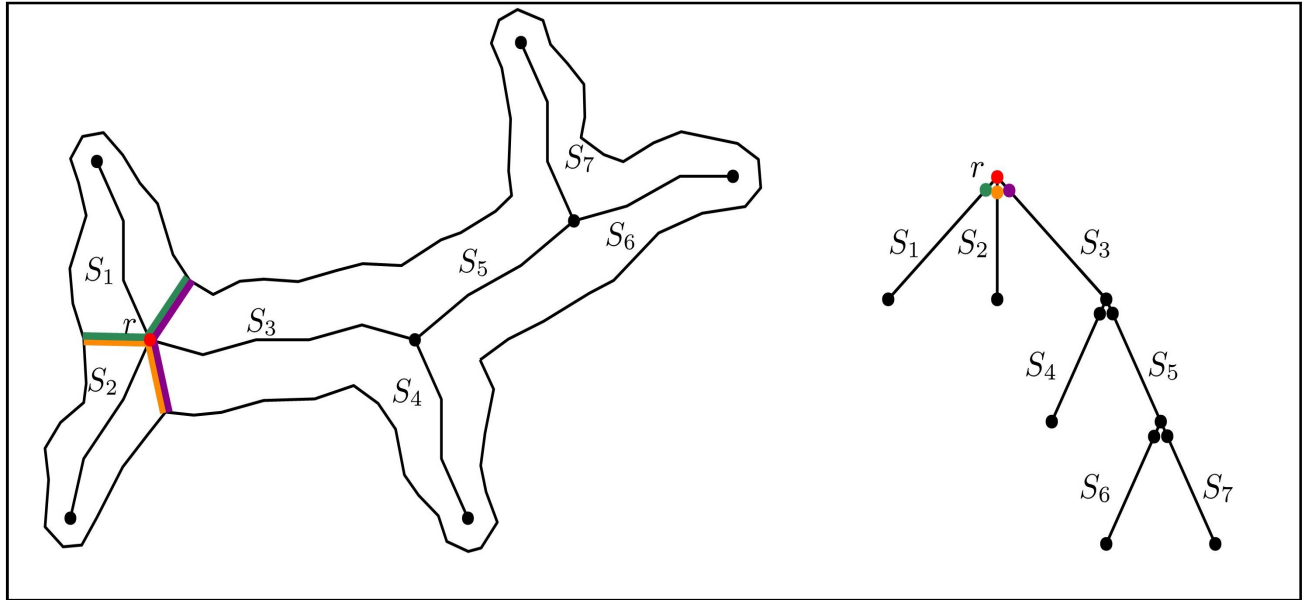


Figure 5. Representing the skeleton graph as a tree rooted at the point  $r$



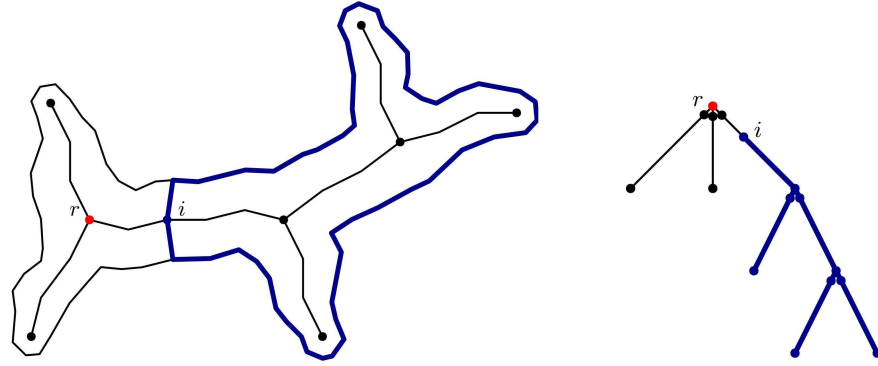


Figure 6. Subpolygon  $P(i)$  ending in the skeleton point  $i$  and the corresponding subtree  $T_i$

- $P(i)$  is feasible or
- There exists a feasible polygon  $P'$  ending in  $i$  and feasible decompositions of the connected components of  $P(i) \setminus P'$ .

Thus, we have to consider all different combinations of skeleton points that together with  $i$  can form such a polygon  $P'$ . In a top-down manner, we consider the different combinations of nodes  $[i_1, i_2, \dots, i_l]$  such that  $i_j \in T_i$  and  $T_{i_j} \cap T_{i_{j'}} = \emptyset$  for all  $j = j'$ . The polygon  $P_0$  corresponds to the subtree rooted in  $i$  with  $i_1, i_2, \dots, i_l$  as the leaves, depicted in blue in Fig. 7. Note that we can compute  $P'$  as a union of subpolygons iteratively. We check if  $P'$  is feasible and we have feasible decompositions for each  $P(i_j)$ , meaning every subtree  $T_{i_j}$  (gray in Fig. 7). Because of Observation 5, we know that  $l \leq k$ , for  $k$  being the number of leaves in the skeleton tree. We have a feasible decomposition of the whole polygon if there exists one of the polygon  $P(r)$ . This computation dominates the runtime with the maximum number of combinations to consider being in  $\mathcal{O}(n^k)$ . Note that this approach does not depend on the initial choice of the root node.

**Theorem 7** Given a simple polygon  $P$  with skeleton  $S$  consisting of  $n$  points with degree at most three, one can compute a feasible decomposition of  $P$  based on  $S$  in  $\mathcal{O}(n^k F)$  time, with  $k$  being the number of leaves in the skeleton tree and  $F$  as above.

### 3.3 Feasibility Constraints and Optimization

Our decomposition method is highly versatile framework that can be adjusted for different feasibility constraints and optimization goals. In the following, we present a few examples of criteria we considered with regard to our application. As stated before, the ROI in our tissue sample needs to satisfy constraints in size and morphology for the LCM. In LCM, a laser separates a tissue fragment from its surrounding sample and a subsequent laser pulse catapults the fragment into a collecting device. On the one hand, the fragment has to have a certain size to ensure that enough material is supplied to be analyzed. On the other hand, the size cannot be too large otherwise the transferring process will fail. The transferring process also fails if the fragment has an irregular shape, since the laser pulse is concentrated on only one boundary point of the fragment. Specifically, elongated shapes or objects with narrow regions (bottlenecks) are problematic, as the tissue can tear and is only transferred partly or not at all. For simplicity, we consider only polygons with a linear skeleton in the following, but all mentioned constraints can be applied to the general decomposition method as well.



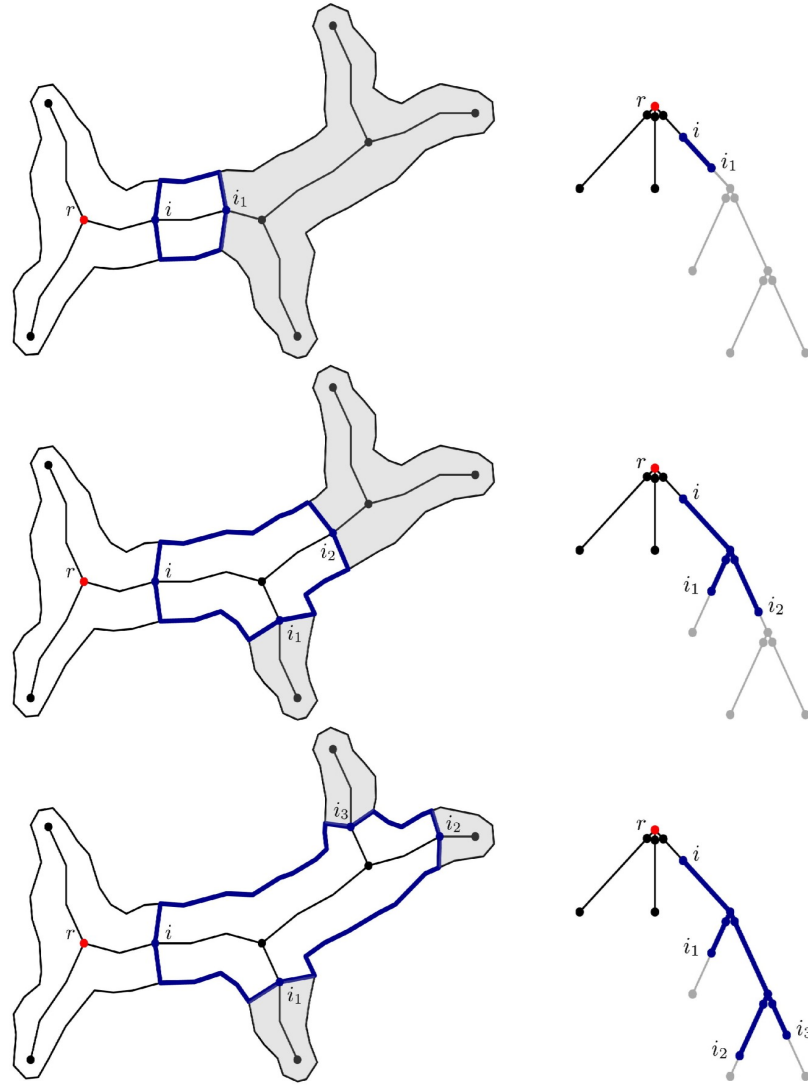


Figure 7. Different possibilities of skeletons points to consider when computing the decomposition of  $P(i)$

### Feasibility Constraints

First, we consider the size constraint by constraining the *area* of the subpolygons. Given two bounds  $l$  and  $u$ , a polygon  $P$  is feasible if  $l \leq A(P) \leq u$ , for  $A(P)$  being the area of the polygon. Instead of the area one could also constrain the number of boundary points. Regarding a shape constraint for the polygons, we considered approximate convexity and fatness. For the convexity, a polygon  $P$  would be feasible if every inner angle lies between two given bounds. However, this does not prevent elongated shapes. By constraining the fatness of a polygon, we can achieve less elongated shapes. Fatness is defined by the aspect ratio  $AR(P)$  of a polygon, which is the ratio of its width to its diameter. For a simple polygon, the diameter is defined as the diameter of the smallest enclosing circle and the width as the diameter of the largest inscribed circle. The aspect ratio lies in  $[0, 1]$  and the higher its value the more circular and less elongated the shape is. We have the following constraint: A polygon is feasible if it is  $\epsilon$ -fat, meaning that  $AR(P) \geq \epsilon$  for some parameter  $\epsilon$ .

## Optimization goals

To solve the decision problem, we assign a value  $X(i)$  to every skeleton point  $i$ . This value equals True if there exists a feasible decomposition of the polygon  $P(i, n)$  and False otherwise. We initialize  $X(1) = \text{True}$  and compute  $X(i)$  as described before, that is  $X(i) = \text{True}$  if there exists  $j \geq i$  such that  $P(i, j)$  is feasible and  $X(j) = \text{True}$ . We can define  $X(i)$  in different ways to achieve various optimization goals. For a point  $i$ , let  $I$  be the set of points  $j$  such that  $P(i, j)$  are feasible. One possible optimization goal would be to find the minimal decomposition (MinNum). We define  $X(i)$  as the number of subpolygons in an optimal feasible decomposition of  $P(i, n)$ , set  $X(1) = 0$  and compute  $X(i) = \min_{j \in I} X(j) + 1$ . If one prefers the cuts to be at bottlenecks of the polygon, one could also consider *minimizing the length of the cut edges* (MinCut). Since every skeleton point  $i$  is the center of a maximal disk, we can obtain the cut length by the corresponding radius  $r(i)$ . We can define  $X(i)$  either as the length of the longest cut or as the sum of cut lengths in the optimal decomposition of  $P(i, n)$  and compute  $X(i) = \max\{\min_{j \in I} X(j), r(i)\}$  resp.  $X(i) = \min_{j \in I} X(j) + r(i)$ . Another optimization goal we considered is maximizing the fatness (MaxFat). We define  $X(i)$  as the smallest aspect ratio of a subpolygon in the decomposition and compute  $X(i) = \max_{j \in I} \{\min\{X(j), AR(P(i, j))\}\}$ . The runtime of the algorithm stays the same for different optimization goals. Obviously, one can combine different feasibility criteria and optimization goals.

The versatility of our algorithm allows the implementation of many different criteria. Note that for certain combinations other (faster) methods might exist. One example is finding the minimal (MinNum) decomposition in which the area of the subpolygons is bounded. For polygons with linear skeletons this can be modeled as finding the minimal segmentation of a weighted trajectory (in  $\mathcal{O}(n \log n)$  time [1]). For general polygons, this problem can be modeled as computing the minimal  $(l, u)$ -partition of a weighted cactus graph (in  $\mathcal{O}(n^6)$  time [6, 7]).

## 4. Results

We evaluated our algorithm by computing decompositions of tumor regions in infrared microscopic images of lung tissue samples that were identified using the method in [15]. With the tumor regions as our ROI, we decomposed each connected component by applying our algorithm on each branch of the corresponding skeletons separately. This approach follows the practical consideration that a polygon as a whole may not possess a feasible decomposition, while some individual branches do. We assessed the decomposition outcome in two respects. First, we present the differences in the decompositions when using different feasibility criteria and optimization goals. Then, we compare our decomposition approach based on the maximal fatness (MaxFat) to a heuristic recursive bisection method (BiSect) that was used to decompose tissue samples for LCM in previous work [15].

### 4.1 Comparison of feasibility constraints and optimization goals

Our algorithm facilitates the use of a wide range of feasibility criteria and optimization goals. The number of subpolygons as well as the positions of cuts depend on these constraints. Having a larger upper bound on, for example, the maximal area of a subpolygon results in fewer subpolygons, as illustrated in Fig. 8a and 8b. These decompositions both minimized the number of subpolygons, but the solutions are not necessarily unique and one optimal decomposition is chosen arbitrarily. This can be seen at the bottom-most skeleton branch in both polygons as the decomposition in Fig. 8a would be feasible with the constraints in 8b as well. In Fig. 8c and 8d, a fatness criterion has been added through a lower bound on the aspect ratio of subpolygons. This criterion avoids the tendency towards elongated subpolygons that can be observed in Fig. 8a. If the bounds in some constraints are too tight, a feasible decomposition might not exist. We illustrate this case in Fig. 8d, where the algorithm did not decompose the polygon parts depicted in gray. This is not favorable for our application, as

it reduces the amount of extracted tissue material.

Regarding size as the feasibility criterion, we applied the different optimization goals described in Section 3.3, which we denoted by MinNum, MinCut and MaxFat. While choosing a different optimization criterion will not affect the area of the polygon that is successfully decomposed, the amount and positions of cut edges may change significantly. These changes may have an influence on the amount of successfully extracted fragments with LCM in practice later on. If we look at the decompositions of the top left skeleton branch, we can see that in Fig. 8a and 8e we have the same number of fragments, but with MinCut a cut with a lower length is chosen. However, maximizing the fatness of each subpolygon usually results in a higher number of subpolygons, but, as can be seen in Fig. 8f, each subpolygon is less elongated and somewhat rounder in shape. We expect these to be the desired shapes for our application. Thus, we used this optimization goal in the comparison of decomposition methods.

## 4.2 Comparison to BiSect

In this section, we evaluate our method with respect to its benefits for LCM. As mentioned before, the success of tissue extraction with LCM highly depends on the size and shape of the given tissue region. For the comparison, we applied the algorithm that maximizes the fatness (the aspect ratios) of the computed subpolygons while using a size constraint (lower and upper bound on the subpolygon's area). We denote this approach by *MaxFat* and compare it to a heuristic method we call *BiSect*. This method decomposes a polygon by recursively bisecting it if its area exceeds the upper size bound.



(a) area in  $[50, 300]$ , MinNum



(b) area in  $[50, 500]$ , MinNum



(c) area in  $[50, 300]$ , fatness 0.4, MinNum



(d) area in  $[50, 300]$ , fatness 0.5, MinNum



(e) area in  $[50, 300]$ , MinCut



(f) area in  $[50, 300]$ , MaxFat

Figure 8. Decomposition with different feasibility criteria and optimization goals

recursively bisecting it if its area exceeds the upper size bound. If the area of a (sub)polygon is below the given lower bound, it is discarded. For technical reasons with LCM, this method is designed to leave a strip of tissue behind with every bisection (see Fig. 9).

We computed the decompositions of 10 different lung tissue samples with both methods and the same area bounds, namely, a minimal and maximal area of 100 px and 2800 px respectively. These 10 tissue samples contained 460 connected components of tumor regions as our regions of interest (ROI). Each ROI covered in average an area of 4100 px. Hence, for many ROIs the decomposition size with BiSect is fairly low. In fact, the average number of subpolygons per ROI was 2.3 for BiSect, but 9.3 for MaxFat.

Assessing the quality of results is of key importance for comparing results between the two methods. While an ideal assessment would compare the actual physical yield of tissue material, such experimental validation was not available in our present study, so that we rely on purely computational measures. Specifically, we employed two main measures: On the one hand, the amount of tissue loss in both decompositions, and on the other hand, the fatness of the resulting subpolygons, which as a single parameter captures reasonably how favourable a specific shape is for LCM. While MaxFat in fact optimizes towards obtaining fat fragments, it is still important to assess how far this translates into practice and how it compares to the fatness obtained by BiSect.

### Area loss

First, we compare the amount of tissue loss (in percent) for each individual polygon/ROI. Both methods inherently involve tissue loss. In BiSect, each bisection results in some tissue loss, therefore, the amount of lost tissue rises with the number of subpolygons. In MaxFat, however, we lose tissue only if no feasible decomposition exists for a given skeleton branch. This mainly occurs when the corresponding (sub)polygon is either too slim or too wide. The first case is depicted in Fig. 9a: The polygons belonging to the bottom two skeleton branches (depicted in gray) are too small and thus no feasible decomposition was computed, resulting in this area being lost. This can be attributed to shortcomings of the underlying skeleton pruning method [2]. Improving the pruning of the skeleton may avoid such short branches. The second case of too wide (or fat) shapes is exemplified in Fig. 9d: In our approach, cut edges are introduced as line segments between a skeleton point and its closest boundary points. In the case of a wide shape being decomposed using a small upper bound for the size constraint, this leads to either thin-slicing or no feasible solution at all. This illustrates that our approach is tailored towards complex, ramified shapes rather than fat objects whose interior can be decomposed effectively through a simple grid pattern. It is also noteworthy that Polygon in Fig. 9d covers an area of around 43000 px and therefore presents an huge outlier in our sample.

The mean area loss with MaxFat is lower than the mean area loss with BiSect (see Fig. 10), yet with a greater variability in values and some high-loss outliers, which can be assigned to large and fat objects. While such objects do not occur frequently in our samples, the resulting area loss is obviously very high. This contributes to the higher standard deviation that we observe when considering the individual ROIs. Since the resulting fragments for each component in one tissue sample are collectively gathered, it is reasonable to validate on the level of samples and determine the entire yield for all ROIs in each sample. As can be seen in Table 1, the decomposition with MaxFat yields overall more tissue (with respect to area) to be collected for most tissue samples.

### Fatness

It is important to note that the success of tissue collection using LCM depends not only on the size but also the

shape of the fragments. BiSect applies merely a size constraint, whereas MaxFat considers both factors. While BiSect only decomposes polygons that exceed the given upper size bound, MaxFat searches for an decomposition in which the subpolygons have the highest possible aspect ratios.

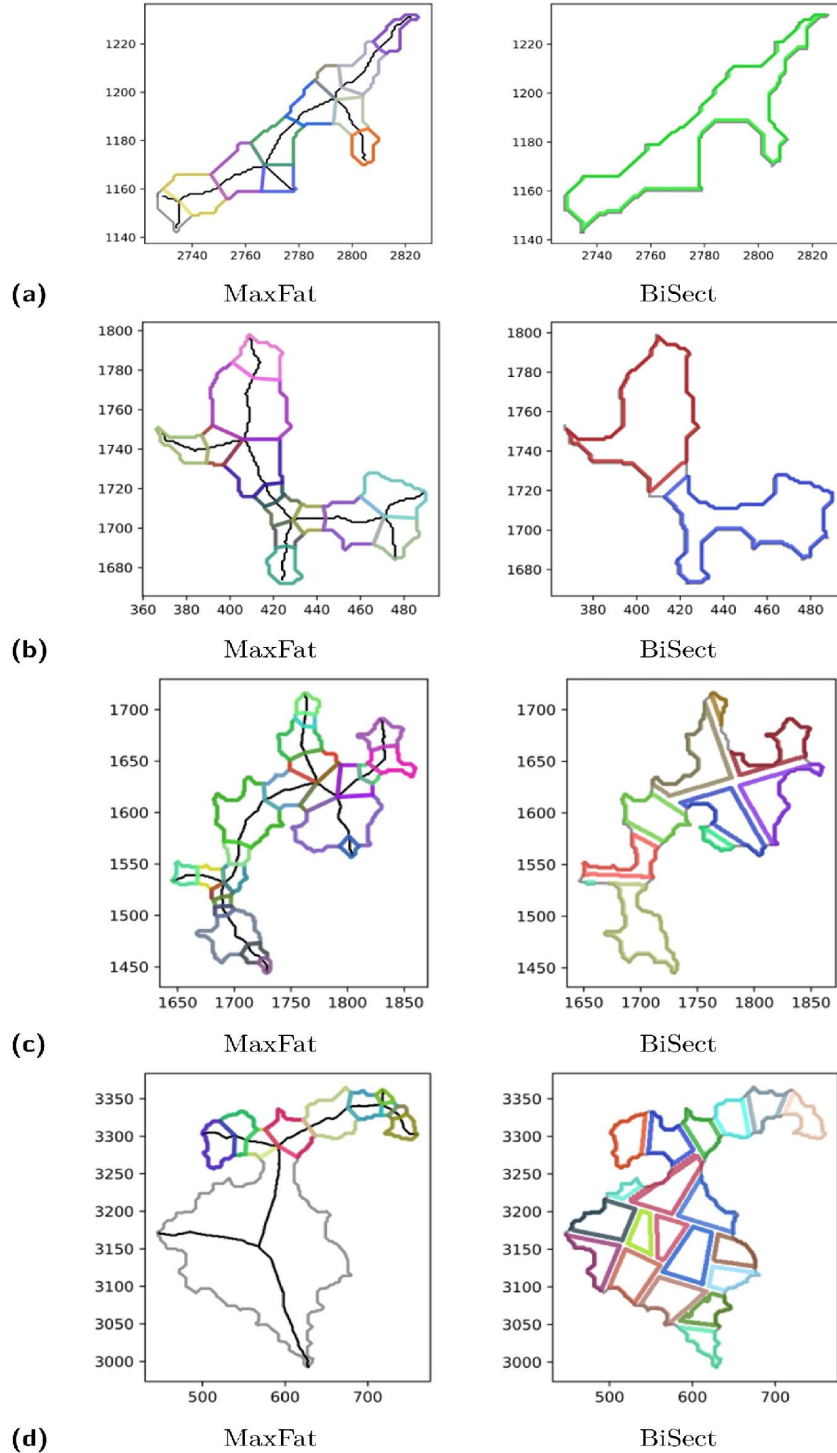


Figure 9. Four example shapes decomposed by MaxFat (left) and BiSect (right)

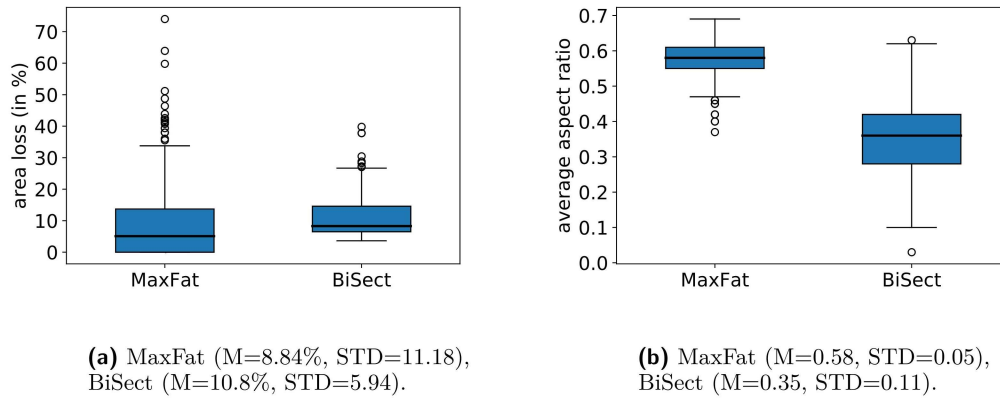
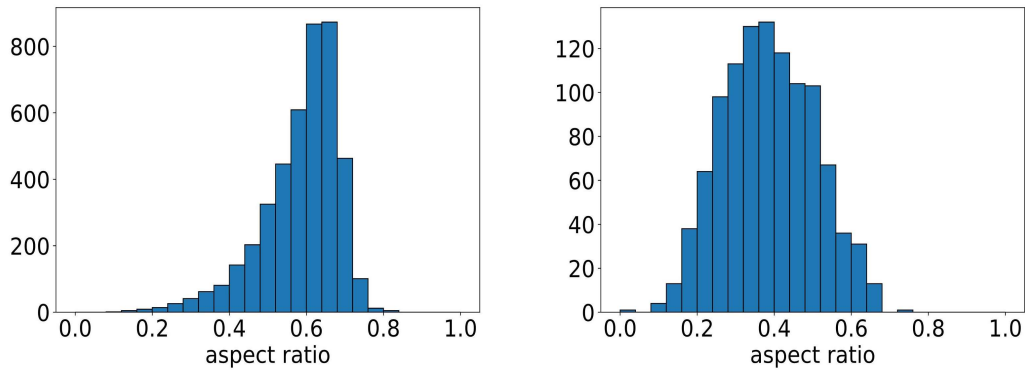


Figure 10. Comparison of the distribution of the percentage of area loss and the average fatness in the decompositions with MaxFat and BiSect for each ROI ( n=460)

Sample	1	2	3	4	5	6	7	8	9	10
MaxFat	91.42	78.57	94.78	76.15	96.58	88.29	96.60	86.44	89.55	94.78
BiSect	89.38	81.96	86.56	76.59	86.53	80.98	87.00	85.41	81.52	86.56

Table 1. Comparison of the yielded area (in %) of the different decomposition methods



(a) MaxFat (n=4285, M=0.58, STD=0.1)

(b) BiSect (n=1066, M=0.38, STD=0.12)

Figure 11. Distributions of the aspect ratios for each subpolygon in the decompositions with MaxFat and BiSect for all samples combined

This obviously results in a higher number of subpolygons over all samples: MaxFat resulted in 4285 and BiSect in 1066 fragments. In some cases, the size of a ROI did not make it necessary to place any cut when decomposing with BiSect, but the shape of the ROI oftentimes is too complex or irregular for LCM to be successful. Additionally, the cut placement with BiSect is far from being optimal.



As can be seen in the examples of Fig. 9b and 9c, the shapes resulting from BiSect are highly irregular – except for most “internal” subpolygons in large, round ROI. In most cases, the resulting subpolygons are far from being convex (in an approximate sense), but elongated and show narrow bottlenecks. All these shape properties pose a risk for the tissue to tear in the extraction process. With the MaxFat approach, however, we overall receive less elongated and rounder shapes. When comparing the average fatness – meaning the average of aspect ratios of the subpolygons – in the decompositions for each ROI (Fig. 10), we clearly achieved higher values with MaxFat with a smaller variability. In fact, for over 75 % of ROI our method achieved an average fatness higher than 0.5, whereas with BiSect nearly 75 % of ROI have an average fatness lower than 0.4. When comparing the distribution of aspect ratios for the individual subpolygons, it is revealed that BiSect shows the pattern of a normal distribution, whereas the distribution for MaxFat is clearly left-skewed (see Fig. 11). This shows that without applying additional shape constraints the decomposition does not result in fragments of the desired shape whereas our method consistently obtains such fragments.

## 5. Conclusion

In this paper, we presented a skeleton-based decomposition method for simple polygons as a novel approach to decompose disease-specific regions of tissue samples while aiming to optimize the amount of tissue obtained by laser capture microdissection (LCM). Compared to naive heuristic approaches that are currently used, our approach provably optimizes target functions under side constraints that are tailored towards relevance for tissue yield in LCM, as this requires fragments of suitable size and morphology to minimize tissue loss. As we demonstrated, these theoretical properties translate into practice when comparing our approach to a recursive bisection approach that considers fragment size as the only criterion for decomposition.

Our approach is designed towards complex morphological structures that are commonly found in cancerous tissue and are usually the most challenging to extract using LCM without major loss of tissue mass. Not surprisingly, it does not perform well when fragmenting relatively fat shapes into small fragments. In this case, it is expected to yield thin slices not ideal for LCM or find no feasible decomposition at all. Yet, such fat shapes does not occur frequently and can easily be decomposed using simple approaches, e.g. bisection-based approaches, without major tissue loss. Thus, we expect to achieve the best outcome for the practical application if we combine both approaches. Our method can be used for the majority of the complex tissue regions and for simple fat morphologies, which can be easily distinguished and separated, other approaches can be applied.

The implementation of our approach relies on a skeletonization of the underlying polygons. Specifically, we utilizes the approach by Bai et al. [2], which implements a heuristic pruning approach. It is likely that recent improvements for skeletonization and pruning [12, 13] will further improve results, as in particular the pruning step of these recent methods promises to avoid short and other spurious branches which negatively affect the amount of yielded tissue in our approach.

Finally, our validation is merely based on quantitative morphological indicators about the resulting fragments. For future work, it will be important to validate the improvements experimentally, e.g. by comparing the actual yield of protein or DNA from different approaches. Overall, our work contributes to further optimization and automation of LCM and thus promises to contribute to the further maturing of the technology and enhancing its suitability for systematic use in larger scale clinical studies.



## References

- [1] Alewijnse, Sander P. A., Buchin, Kevin., Buchin, Maike., Kölzsch, Andrea., Kruckenberg, Helmut Westenberg, Michel A. (2014). A framework for trajectory segmentation by stable criteria. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '14* (pp. 351–360). Association for Computing Machinery.
- [2] Bai, Xiang., Latecki, Longin Jan Liu, Wen-Yu. (2007). Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3).
- [3] Blum, Harry. (1967). A transformation for extracting new descriptors of shape. In *Models for Perception of Speech and Visual Forms* (pp. 362–380).
- [4] Blum, Harry. (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology*, 38(2), 205–287.
- [5] Buchin, Maike., Mosig, Axel Selbach, Leonie. (2019). Skeleton-based decomposition of simple polygons. In *Abstracts of 35th European Workshop on Computational Geometry*. Retrieved from <http://www.eurocg2019.uu.nl/papers/3.pdf>
- [6] Buchin, Maike Selbach, Leonie. (2020). Decomposition and partition algorithms for tissue dissection. In *Computational Geometry: Young Researchers Forum* (pp. 24–27). Retrieved from <http://www.computational-geometry.org/YRF/cgyrf2020.pdf>
- [7] Buchin, Maike Selbach, Leonie. (2020). A polynomial-time partitioning algorithm for weighted cactus graphs. *arXiv Preprint*. arXiv:2001.00204
- [8] Choi, Hyeong In., Choi, Sung Woo Moon, Hwan Pyo. (1997). Mathematical theory of medial axis transform. *Pacific Journal of Mathematics*, 181(1), 57–88.
- [9] Datta, Soma., Malhotra, Lavina., Dickerson, Ryan., Chaffee, Scott., Sen, Chandan K Roy, Sashwati. (2015). Laser capture microdissection: Big data from small samples. *Histology and Histopathology*, 30(11), 1255.
- [10] Sanniti di Baja, Gabriella Thiel, Edouard. (1994). (3,4)-weighted skeleton decomposition for pattern representation and description. *Pattern Recognition*, 27(8), 1039–1049.
- [11] Drechsler, Klaus Oyarzun Laura, Cristina. (2010). Hierarchical decomposition of vessel skeletons for graph creation and feature extraction. In *Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on* (pp. 456–461). IEEE.
- [12] Durix, Bastien., Chambon, Sylvie., Leonard, Kathryn., Mari, Jean-Luc Morin, Géraldine. (2019). The propagated skeleton: A robust detail-preserving approach. In *International Conference on Discrete Geometry for Computer Imagery* (pp. 343–354). Springer.

- [13] Durix, Bastien., Morin, Géraldine., Chambon, Sylvie., Mari, Jean-Luc Leonard, Kathryn. (2019). One-step compact skeletonization. In *Eurographics (Short Papers)* (pp. 21–24).
- [14] Emmert-Buck, Michael R., Bonner, Robert F., Smith, Paul D., Chuaqui, Rodrigo F., Zhuang, Zhengping., Goldstein, Seth R., Weiss, Rhonda A Liotta, Lance A. (1996). Laser capture microdissection. *Science*, 274(5289), 998–1001.
- [15] Großerueschkamp, Frederik., Bracht, Thilo., Diehl, Hanna C., Kuepper, Claus., Ahrens, Maike., Kallenbach-Thieltges, Angela., Mosig, Axel., Eisenacher, Martin., Marcus, Katrin., Behrens, Thomas., et al. (2017). Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection, and proteomics. *Scientific Reports*, 7(1), 1–12.
- [16] Keil, J. Mark. (2000). Polygon decomposition. In *Handbook of Computational Geometry* (Vol. 2, pp. 491–518).
- [17] Kondo, Yutaka., Kanai, Yae., Sakamoto, Michiie., Mizokami, Masashi., Ueda, Ryuzo Hirohashi, Setsuo. (2000). Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis—A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. *Hepatology*, 32(5), 970–979.
- [18] Leonard, Kathryn., Morin, Geraldine., Hahmann, Stefanie Carlier, Axel. (2016). A 2D shape structure for decomposition and part similarity. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (pp. 3216–3221). IEEE.
- [19] Narro, Martha L., Yang, Fan., Kraft, Robert., Wenk, Carola., Efrat, Alon Restifo, Linda L. (2007). NeuronMetrics: Software for semi-automated processing of cultured neuron images. *Brain Research*, 1138, 57–75.
- [20] Ong, Chin-Ann J., Tan, Qiu Xuan., Lim, Hui Jun., Shannon, Nicholas B., Lim, Weng Khong., Hendrikson, Josephine., Ng, Wai Har., Tan, Joey W. S., Koh, Kelvin K. N., Wasudevan, Seettha D., et al. (2020). An optimized protocol harnessing laser capture microdissection for transcriptomic analysis on matched primary and meta-static colorectal tumors. *Scientific Reports*, 10(1), 1–12.
- [21] Papanelopoulos, Nikos., Avrithis, Yannis Kollias, Stefanos. (2019). Revisiting the medial axis for planar shape decomposition. *Computer Vision and Image Understanding*, 179, 66–78.
- [22] Reniers, Dennie Telea, Alexandru. (2007). Skeleton-based hierarchical shape segmentation. In *Shape Modeling and Applications, 2007. SMI'07. IEEE International Conference on* (pp. 179–188). IEEE.
- [23] Saha, Punam K., Borgefors, Gunilla Sanniti di Baja, Gabriella. (2016). A survey on skeletonization algorithms and their applications. *Pattern Recognition Letters*, 76, 3–12.
- [24] Schmuck, Martin R., Temme, Thomas., Dach, Katharina., de Boer, Denise., Barenys, Marta., Bendt, Fa

rina., Mosig, Axel Fritsche, Ellen. (2017). Omnisphero: A high-content image analysis (HCA) approach for phenotypic developmental neurotoxicity (DNT) screenings of organoid neurosphere cultures in vitro. *Archives of Toxicology*, 91(4).

[25] Tanase, Mirela Veltkamp, Remco C. (2003). Polygon decomposition based on the straight line skeleton. In *Geometry, Morphology, and Computational Imaging* (pp. 247–268). Springer.

[26] Vu, Quoc Dang., Graham, Simon., Kurc, Tahsin., To, Minh Nguyen Nhat., Shaban, Muhammad., Qaiser, Talha., Koohbanani, Navid Alemi., Khurram, Syed Ali., Kalpathy-Cramer, Jayashree., Zhao, Tianhao., et al. (2019). Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in Bioengineering and Biotechnology*, 7.

[27] Wang, Chaofeng., Gui, Cai-Ping., Liu, Hai-Kuan., Zhang, Dong Mosig, Axel. (2013). An image skeletonization-based tool for pollen tube morphology analysis and phenotyping. *Journal of Integrative Plant Biology*, 55(2), 131–141.

[28] Wang, Shidan., Rong, Ruichen., Yang, Donghan M., Fujimoto, Junya., Yan, Shirley., Cai, Ling., Yang, Lin., Luo, Danni., Behrens, Carmen., Parra, Edwin R., et al. (2020). Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer Research*.

[29] Wolter, Franz-Erich. (1993). Cut locus and medial axis in global shape interrogation and representation.