

# Model-based Correction of Segmentation Errors in Digitised Histological Images

Randell, David; Galton, Antony; Fouad, Shereen; Mehanna, Hesham; Landini, Gabriel

DOI:

[10.1007/978-3-319-60964-5\\_63](https://doi.org/10.1007/978-3-319-60964-5_63)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Randell, D, Galton, A, Fouad, S, Mehanna, H & Landini, G 2017, Model-based Correction of Segmentation Errors in Digitised Histological Images. in Medical image understanding and analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings. Communications in computer and information science, vol. 723, Springer, pp. 718-730, Medical Image Understanding and Analysis (MIUA) 2017, United Kingdom, 11/07/17. [https://doi.org/10.1007/978-3-319-60964-5\\_63](https://doi.org/10.1007/978-3-319-60964-5_63)

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Model-Based Correction of Segmentation Errors in Digitised Histological Images

David A. Randell<sup>1</sup>(✉), Antony Galton<sup>2</sup>, Shereen Fouad<sup>1</sup>, Hisham Mehanna<sup>3</sup>,  
and Gabriel Landini<sup>1</sup>

<sup>1</sup> School of Dentistry, College of Medical and Dental Sciences,  
University of Birmingham, Birmingham, UK

{d.a.randell,s.a.fouad,g.landini}@bham.ac.uk

<sup>2</sup> Department of Computer Science, University of Exeter, Exeter, UK  
apgalton@ex.ac.uk

<sup>3</sup> Institute of Cancer and Genomic Sciences, University of Birmingham,  
Birmingham, UK  
h.mehanna@bham.ac.uk

**Abstract.** This paper describes an application of topological, model-based methods for the algorithmic correction of segmentation errors in digitised histological images. The topological analysis is provided by the spatial logic Discrete Mereotopology and integrates qualitative spatial reasoning and constraint satisfaction methods with classical image processing methods. A set of eight topological relations defined on binary segmented regions are factored out and reworked as nodes of a set of directed graphs. The graphs encode and constrain a set of set-theoretic and topological segmentation operations on regions, so that the interpreted images and any proposed changes made to the regions can be made to conform to a valid histological model. Worked examples are given using images of H&E stained H400 cell line cultures.

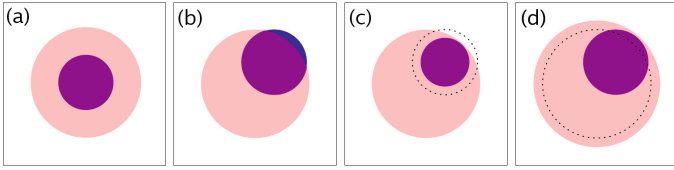
**Keywords:** Mereotopology · Graph theory · Histological image processing

## 1 Introduction

This paper describes an application of model-based methods for the algorithmic correction of segmentation errors in digitised histological images. This is a real-world application where qualitative spatial reasoning (QSR) and constraint-satisfaction programming methods have been integrated with classical image processing methods to develop context-based histological imaging algorithms.<sup>1</sup> The context here arises from: (i) making an ontological stand whereby regions rather than pixels in digitised images are deemed to be the main carriers of

---

<sup>1</sup> See our project page at <http://www.mecourse.com/landinig/software/intellimic.html>.



**Fig. 1.** Idealised image of an H&E-stained nucleated cell, with pink-stained cytoplasm and blue-stained nucleus; overlap between cytoplasm and nucleus is shown in magenta. In (a) the nucleus forms a proper part of the cytoplasm, as expected; (b) shows an anomalous image in which the nucleus partially overlaps the cytoplasm; (c) and (d) show possible morphological corrections of (b), with dotted lines marking the original extent of nucleus and cytoplasm respectively. (Colour figure online)

histological content, and (ii) highlighting the importance of and explicitly representing topological (and in particular relational) information encoded in digitised histological images. The topological analysis and constraints are provided by the spatial logic Discrete Mereotopology (DM) [1, 2], which we use to augment the operations of classical Mathematical Morphology (MM) by explicitly encoding a set of binary relations such as contact, overlap and the part-whole relation on pairs of regions. These mereotopological relations are used both to model the domain and to guide algorithms for correcting segmentation results so that they conform to the requirements for a valid histological model.

In three-dimensional reality a cell consists of a nucleus surrounded by a body of cytoplasm; the two are in contact but do not overlap, the nucleus exactly filling a cavity within the cytoplasm. In a correctly-formed H&E (haematoxylin and eosin)-stained two-dimensional image, on the other hand, based on the differential staining of cellular components, the cytoplasm region appears as a simply connected whole, lacking a cavity, and the nucleus forms a proper part of this (see the idealised image in Fig. 1(a)). In practice, however, cell and nuclear segmentations are most often achieved independently of each other, which can result in imperfect relations, where, e.g., the nucleus partially overlaps its cytoplasm, as in Fig. 1(b). Such errors can often be corrected by a process of *resegmentation*, whereby one or both of the cell component images are manipulated using morphological or other operators so that the expected spatial relation between them holds. In Fig. 1(c) this is accomplished by eroding the nucleus; in (d), by dilating the cytoplasm. If the changes required to achieve this are too extreme, then the presence of an imaging artefact or other anomaly may be indicated.

Our approach to resegmentation uses a set of directed graphs defined on region-region relations. Each graph details transformations to the relations arising from application of specified changes to their relata. These changes use discrete topological and set-theoretic operators defined on regions and mapped to their MM equivalents. The graphs enable transitions between relations (e.g., from partial overlap to proper part) to be implemented as sequences of operations on the regions themselves. This correspondence between relations and operations underpins the model-based correction of segmentation described in this paper.

This paper builds on a series of papers that apply DM to the interpretation and segmentation of histological images [3–6]. While the mereotopological

interpretation of digitised images is not new (see e.g., [7]), what is novel is its role in enabling systematic context-based manipulation of regions and their relations in quantitative histological image processing. DM first appeared in [1], which arose out of the mereotopological theory RCC and its well-known eight-element relation set RCC8 [8]. The GRCC (Generalised Region Connection Calculus) [9] offers another way to model discrete domains using mereotopology. As a formalism for modelling cellular processes, mereotopology was used to model phagocytosis and exocytosis in [10]. A motivating example of DM used in a constraint-based graph traversal problem is given in [11]. In general, though, while RCC8 has been used extensively in the development of efficient constraint-based algorithms (for deciding the consistency of a set of constraints) the algorithms are typically restricted to operations on symbolic state-state models and not grounded in interpreted segmented digitised images as done here. For a good introduction to QSR literature and its methods, see [12].

This paper extends the framework that was originally reported in [6]. Specifically, the four directed graphs originally described in [6] have been extended to include a set of eight set-theoretical as well as topological operations on regions. The method is generic and does not favour any one particular image segmentation method over another. All that is assumed is that the segmentation of images into regions forming relations with other regions can be ultimately represented as binary regions, e.g. as masks. These relations can be used to test whether the segmented regions conform to a histological model, or not. Where they fail to conform to the segmentation model they may be either be rejected completely, or corrected by applying one of several operations on the regions in question. For this reason, we do not necessarily require a set of gold-standard segmented images with which to compare and analyse the results (though of course this could be done). We discuss this further below. In its place we consider the set of topological segmentation solutions satisfying our assumed histological model and work with those. Reasons of space and the scope of this paper mean we cannot adequately compare and contrast other segmentation methods and the various validation methods used in the literature here, though we acknowledge there is a large body of literature on the subject, e.g. [13,14]. However, none of these frameworks adopt the topological framework assumed here, nor do they address systematic methods of resegmentation to bring the results of initial segmentation into line with the expectations of histological theory, which is the main focus of the present paper.

## 2 Discrete Mereotopology (DM)

Of the different variants of DM that have been proposed, the version we adopt is the one described in [1,6]. The domain is a set of possibly empty regions which are defined as sets of pixels. We denote regions by lower-case letters ( $x, y, \dots$ ); individual pixels are denoted  $\hat{x}, \hat{y}, \dots$ . Set inclusion is defined in the standard way, the part-whole relation (P) defined as inclusion restricted to non-null regions, and overlap (O) between regions restricted to regions sharing a part in common.

Adjacency between pixels is axiomatised as a reflexive, symmetric relation. This is used to define the *contact* relation (C) between regions. The set of all pixels is defined as the universal region  $u$  and the null set as  $\emptyset$ . Singleton pixel-sized regions are treated as atoms. The *neighbourhood*  $N(\hat{x})$  of pixel  $\hat{x}$  is defined as the set of pixels in  $u$  that are adjacent to  $\hat{x}$ . In our application domain  $u$  is a rectangular pixel array, and  $N(\hat{x})$  a  $3 \times 3$  pixel array centred on  $\hat{x}$ . Implemented in MM, the function  $N(\hat{x})$  maps to a morphological  $3 \times 3$ -pixel, 8-connected *structuring element* which we now assume by default.

As is common practice in QSR when developing these spatial logics and algebras, subsets of dyadic relations forming *jointly exhaustive and pairwise disjoint* (JEPD) sets are singled out. In DM two JEPD relation sets are used: first, the eight-element set RCC8 comprising DC, EC, PO, TPP, NTPP, TPPI, NTPPI and EQ, respectively disconnected, externally connected, partial overlap, tangential proper part, non-tangential proper part, with their inverses, and equals, and second, the five-element relation set RCC5 comprising the relations DR (= DC|EC), PO, PP (= TPP|NTPP), PPI (= TPPI|NTPPI), and EQ, respectively disjoint, partial overlap, proper part, and inverse proper part. Here the symbol ‘|’ signifies disjunction, e.g.,  $DR(x, y) \leftrightarrow DC(x, y) \vee EC(x, y)$ . These respectively form the eight and five base relations of two relational subsumption lattices with top and bottom elements interpreted as the universal and null dyadic relations. Given that RCC8 is predicated on a continuous embedding space and DM on a discrete one, the relations sharing the same name are strictly not identical. For this reason, DM’s JEPD relation sets are identified in the text by the suffix ‘D’ as in ‘RCC8D’.

In [6] the discrete *interior* ( $\text{int}_D$ ) and *closure* ( $\text{cl}_D$ ) are pseudo-topological operators defined on regions; they share some but not all of the usual properties of the interior and closure operators in standard treatments of topology. A map between these operators and the MM operations *erosion* and *dilation* [5] enables one to define a notion of approximate equality that underpins transitions encoded in DM’s conceptual neighbourhood graphs, and it also enables the RCC8D relation set to be easily implemented in any image-processing programs featuring standard MM libraries. Other properties of regions can be defined in DM, e.g., regions with or without an interior, regular regions (i.e., those without pixel-wide spikes and fissures), self-connected regions, connected components.

In the histological domain, cells and their parts, groups of cells forming tissues and compartments, and the background of a digitised histological preparation can all be modelled using DM. Simple regions and arbitrary sets of pixels forming regions which may or may not be spatially contiguous all yield potential models. If a histological preparation is thresholded as a single binary image, then segmented regions of interest will form connected components, so that in any one image, the only possible relations between pairs of regions are DC and EQ. The methods presented here, however, assume that two or more independent imaging or segmentation modalities are used, e.g., separating out the contribution of the different dyes in stained sections (as in Fig. 1) or confocal microscopy channels. In this case, pairs of regions, segmented from each channel, can be compared, all RCC5D and RCC8D relations being now possible.

### 3 Conceptual Neighbourhood Graphs, Continuity and Change, and Composition Tables

In [1,6] a set of *conceptual neighbourhood diagrams* (CNDs), or graphs, were defined on dyadic relations. In RCC8, relation  $R'$  is a conceptual neighbour of relation  $R$  if some pair of regions related by  $R$  can be continuously deformed so that  $R$  changes to  $R'$  with no other relation holding during that deformation. In the discrete setting of DM, continuous deformation is recast in terms of *minimal change*. In [1], this was defined using the discrete interior ( $\text{int}_D$ ) and closure ( $\text{cl}_D$ ) operators; here we extend this to include changes to regions produced using the set-theoretic operators union ( $\text{sum}$ ), intersection ( $\text{prod}$ ) and difference ( $\text{diff}$ ) on region pairs. The universal region  $u$ , representing the image, is assumed to be self-connected, i.e.  $\text{SC}(u)$ .

In general, an *RCC8D conceptual neighbourhood* of a binary relation  $R$  can be defined as

$$\text{nbhd}_{\langle\alpha,\beta\rangle}(R) = \{R' \in \text{RCC8D} \mid \exists x,y(R(x,y) \wedge R'(\alpha,\beta))\}, \tag{1}$$

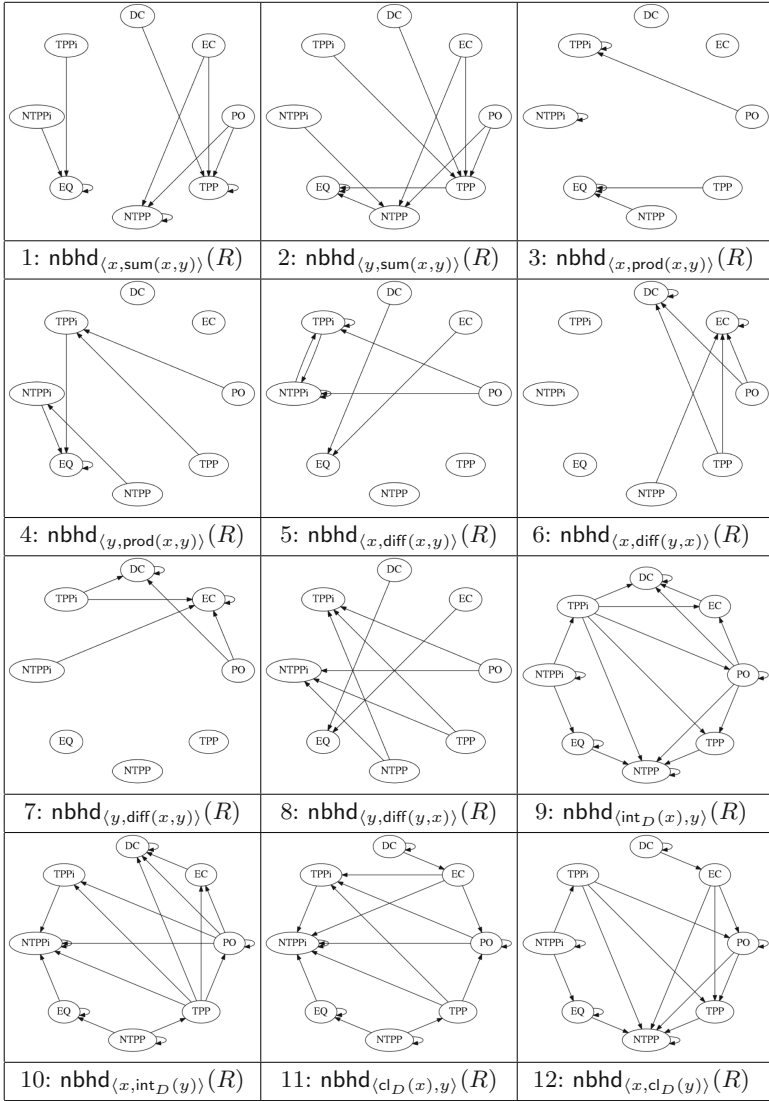
where  $\alpha$  and  $\beta$  are designated functions of the region variables  $x, y$ . The elements of  $\text{nbhd}_{\langle\alpha,\beta\rangle}(R)$  are called the  $\langle\alpha,\beta\rangle$ -neighbours of  $R$ . They are all the possible relations that can hold after regions  $x$  and  $y$  are modified in accordance with  $\alpha, \beta$ . Given a segmented image, by a *resegmentation* we understand the replacement of a set  $S$  of regions in the image by a new set  $S'$  defined from  $S$  using some sequence of conceptual neighbourhood transitions. Such a resegmentation is chosen in order to correct anomalous relations in the original segmentation so that it satisfies the constraints of the domain being modelled.

Figure 2 shows a set of graphs that encode the conceptual neighbourhood relations defined in terms of the operators  $\text{sum}, \text{prod}, \text{diff}, \text{int}_D$ , and  $\text{cl}_D$ . In the first graph, for example, showing

$$\text{nbhd}_{\langle x, \text{sum}(x,y) \rangle}(R) = \{R' \in \text{RCC8D} \mid \exists x,y(R(x,y) \wedge R'(x, \text{sum}(x,y)))\}, \tag{2}$$

the arrow from DC to TPP represents the case  $\text{nbhd}_{\langle x, \text{sum}(x,y) \rangle} \text{DC} = \{\text{TPP}\}$ , meaning that if  $\text{DC}(x,y)$  holds then we must have  $\text{TPP}(x, \text{sum}(x,y))$ . In this case only one resegmentation exists; other cases, such as  $\text{nbhd}_{\langle x, \text{sum}(x,y) \rangle} \text{PO} = \{\text{TPP}, \text{NTPP}\}$ , may allow more than one. Loops in the CND indicate where a change to a region of a designated pair does not necessarily result in a corresponding change of relation. Isolated nodes or nodes without outgoing edges arise where an operator returns null (e.g.,  $\text{diff}(x,y)$  where  $\text{PP}(x,y)$ ).

For graph  $n$ , the outgoing edges from the vertex labelled with relation  $R$  are designated with the mnemonic  $n_R$ ; for example, in the case of the pair of outgoing edges from PO to TPP and NTPP in graph 1, this is notated as  $1_{\text{PO}}$ . Note that four of the RCC8D relations are *self-inverse*, i.e.,  $R(x,y)$  implies  $R(y,x)$ ; these are DC, EC, PO, and EQ. The other four relations form two mutually inverse pairs:  $\text{TPPi}(x,y)$  if and only if  $\text{TPP}(y,x)$ , and  $\text{NTPPi}(x,y)$  if and only if  $\text{NTPP}(y,x)$ . These inverse relations will sometimes be exploited in our reasoning. In graph 4, for example, we see that  $\text{PO}(x,y)$  implies  $\text{TPPi}(y, \text{prod}(x,y))$ ;

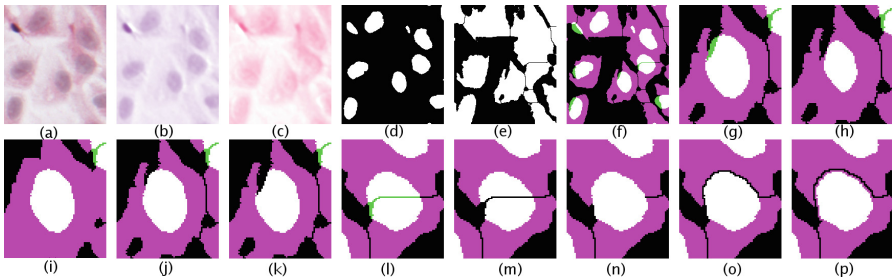


**Fig. 2.** Directed graphs encoding set theoretic and discrete topological operators. In each case the regions and the resulting operation on them are non-null.

sometimes we will find it more convenient to rewrite this as  $\text{TPP}(\text{prod}(x, y), y)$ , in which case we would cite the graph operation as  $4'_{\text{PO}}$ . An example of this is seen later in Table 1.

We also use RCC8D's composition table (RCC8D-CT). The notion of composition is well-known in AI as it provides an efficient inference mechanism for many QSR constraint satisfaction programs, where it is typically implemented as

a simple look-up table. Following [15], weak composition of DM's JEPD relation sets is defined as follows. Given relation set  $\Sigma$ , the weak composition RCC8D-CT( $R, S$ ), where  $R, S \in \Sigma$ , is defined to be the smallest subset  $\{T_i\} \subseteq \Sigma$  such that  $DM \models \forall x, y, z((R(x, y) \wedge S(y, z)) \rightarrow T_1(x, z) \vee \dots \vee T_n(x, z))$ . The elements of RCC8D-CT defined on non-null regions is identical to RCC8's composition table entailed by RCC. This was mechanically proved using the sorted theorem prover SPASS [16] to verify that all entailments of the above form were included in the composition table; and constructing a set of graphical models satisfying each  $T_i(x, z)$  disjunct. The same method was used to verify the sets of directed edges of the graphs depicted in Fig. 2.



**Fig. 3.** (a) Original RGB image; (b) Haematoxylin and (c) Eosin channels; binary segmented (d) nuclei and (e) cytoplasm; (f) RGB colour merge: magenta = cytoplasm, green = nucleus, white = cytoplasm/nucleus overlap; (g,l) cropped details of (f); (h–k, m–p) resegmentations of g,l respectively, satisfying PP(nucleus,cytoplasm). See text for explanation. (Colour figure online)

## 4 Example: Segmenting Cells in Culture

In Fig. 3, image (a) depicts an H&E-stained culture of H400 cells grown on glass. Various image pre-processing operations are done. First a Gaussian filter (kernel radius 2) is applied to the original image to remove noise and reduce the fragmentation artefacts near the region boundaries. Next, colour deconvolution [17] is used to unmix the dye contributions and identify cell nuclei (H-stain, image (b)) from the rest of the cell bodies (E-stain, image (c)). Several standard image processing operations then follow ( $k$ -means clustering on the H,E stain images using 3 clusters, Boolean compositions of thresholded clusters, binary watershed separation), which are used to generate the two binary images of cell nuclei and their associated cells (images (d) and (e) respectively). The colour composite merge depicted in (f) illustrates the extent of conformity to the assumed histological constraints; binary segmented nuclei (d) are mapped to green, cytoplasm (e) to magenta, and overlap between the two to white. Where a nucleus forms a proper part (PP) of its cytoplasm, the latter appears as magenta surrounding a white nucleus; in less common cases where EQ holds, the whole cell appears white. First we test the RCC8D relation between cell nuclei and cytoplasm,



where each typed set of spatially disjoint regions is treated as a mereological whole (regions  $n$  and  $c$  respectively).<sup>2</sup> In the case illustrated, we obtain  $\text{PO}(n, c)$ , as indicated by the presence of green regions in image (f). As this fails the test of conformity (which require nuclei to fall within their associated cytoplasm) the task is to repair the segmentation. In (f) seven candidate nuclei partially overlap (PO) cytoplasm regions and six form proper parts (PP). But these include several small ‘slivers’ adhering closely to the image boundary. Typically, regions bordering the edge of the image frame (so their true extent is not known) are removed from the analysis.

Using the directed graphs, we look for resegmentation operations on candidate nucleus/cytoplasm pairs that take us from PO to PP. Consider the enlarged detail shown in Fig. 3(g), where candidate nucleus  $n$  is PO to cytoplasm component  $c$ . One possible solution (image (h)) successively erodes  $n$  (i.e., replaces it by its discrete interior) until it becomes PP to  $c$ ; this requires three successive erosions. Another (image (i)) replaces  $c$  by its discrete closure (dilation) until the same result is achieved. In (j) we extend  $c$  to cover all of  $n$  so that once again the nucleus is PP to its cytoplasm. In (k) we achieve the same result by subtracting from  $n$  the part that lies outside  $c$ .

Image (l) shows a second enlarged detail of (f), in which another candidate nucleus  $n'$  partially overlaps *two* cytoplasm components  $c_1$  and  $c_2$ . One correction (image (m)) splits the nucleus into two by taking its intersections with  $c_1$  and  $c_2$ . Each nuclear component is now PP to one cytoplasm component. In (n), the lower cytoplasm component ( $c_2$ ) is extended to cover the whole of  $n'$ . This, though, has the effect of merging two cytoplasm components; to compensate for this, the upper component ( $c_1$ ) is reduced by subtracting from it the closure of  $n'$  (image (o)). Finally, in (p), the nucleus is completely surrounded by cytoplasm; this is achieved by extending  $c_2$  to cover not just  $n'$  but its closure; the compensatory reduction of  $c_1$  must now subtract the closure of the closure of  $n'$  in order to ensure complete separation from the extended  $c_2$ .

These resegmentations, associated graphs and inferences used to generate them are summarised in Table 1.

In detail, the steps for  $c_1$  in (p) are as follows:

1. Start with  $\text{PO}(n', c_1)$ .
2. By  $11_{\text{PO}}$  this gives  $\text{PO}|\text{NTPPi}|\text{TPPi}(\text{cl}_D(n'), c_1)$ .
3. By  $1_{\text{PO}|\text{NTPPi}|\text{TPPi}}$  this gives  $\text{EQ}|\text{NTPPi}|\text{TPPi}(\text{cl}_D(n'), \text{sum}(\text{cl}_D(n'), c_1))$ .
4. Next, from  $\text{EQ}(n', n')$  and  $12_{\text{EQ}}$  we have  $\text{EQ}|\text{NTPPi}(\text{cl}_D(n'), \text{cl}_D(n'))$ .
5. The RCC8D weak composition  $\text{EQ}|\text{NTPPi} \circ \text{EQ}|\text{NTPPi}|\text{TPPi}$  is  $\text{EQ}|\text{NTPPi}|\text{TPPi}$ .
6. Hence from 3 and 4 using 5 we have  $\text{EQ}|\text{NTPPi}|\text{TPPi}(n', \text{sum}(\text{cl}_D(n'), c_1))$ .

In step 2, we apply graph 1 to each disjunct of  $\text{PO}|\text{TPPi}|\text{NTPPi}$  separately to generate the new disjunction  $\text{EQ}|\text{TPPi}|\text{NTPPi}$  in step 3: here PO and TTPi (NTPPi)

<sup>2</sup> The justification for this step is a potential speed-up of computation in certain cases: if  $\text{DC}(n, c)$  is returned, then all pairs of connected components of  $n$  and  $c$  are DC, if  $\text{PP}(n, c)$  then every connected component in  $n$  is part of some component of  $c$  (and similarly for  $\text{PPi}$ ), and if  $\text{EQ}(n, c)$  then every connected component in  $n$  maps to an identical component in  $c$ . These are theorems of DM.

**Table 1.** Resegmentation details for Fig. 3. Here CT refers to the RCC8D composition table.

Figure	Initial relation	Relation after resegmentation	Graph operation
(h)	$PO(n, c)$	$TPP(\text{int}_D^3(n), c)$	$9_{PO}$
(i)	$PO(n, c)$	$TPP(n, \text{cl}_D^2(c))$	$12_{PO}$
(j)	$PO(n, c)$	$TPP(n, \text{sum}(n, c))$	$1_{PO}$
(k)	$PO(n, c)$	$TPP(\text{prod}(n, c), c)$	$4'_{PO}$
(m)	$\left\{ \begin{array}{l} PO(n', c_1) \\ PO(n', c_2) \end{array} \right\}$	$\left\{ \begin{array}{l} TPP(\text{prod}(n', c_1), c_1) \\ TPP(\text{prod}(n', c_2), c_2) \end{array} \right\}$	$4'_{PO}$
(n)	$PO(n', c_1)$	$TPP(n', \text{sum}(n', c_1))$	$1_{PO}$
(o)	$\left\{ \begin{array}{l} PO(n', c_1) \\ PO(n', c_2) \end{array} \right\}$	$\left\{ \begin{array}{l} TPP(n', \text{sum}(n', c_1)) \\ DC(n', \text{diff}(c_2, \text{cl}_D(n'))) \end{array} \right\}$	$1_{PO}$ $11_{PO}, 6_{PO}, 12_{EQ}, CT$
(p)	$\left\{ \begin{array}{l} PO(n', c_1) \\ PO(n', c_2) \end{array} \right\}$	$\left\{ \begin{array}{l} NTPP(n', \text{sum}(\text{cl}_D(n'), c_1)) \\ DC(n', \text{diff}(c_2, \text{cl}_D^2(n'))) \end{array} \right\}$	$11_{PO}, 1_{PO}, 12_{EQ}, CT$ $11_{PO}, 6_{PO}, 12_{EQ}, 12_{NTPP}, CT$

are mapped to  $TPP|NTPP$  and  $EQ$  by  $1_{PO}$  and  $1_{NTPPi}$  ( $1_{NTPPi}$ ) respectively; the combined operation is notated  $1_{PO|TPPi|NTPPi}$ .<sup>3</sup> The soundness and completeness of the inference procedures ensure not only that steps 1–4 encode the DM theorem  $PO(x, y) \rightarrow EQ|TPP|NTPP(x, \text{sum}(\text{cl}_D(x), y))$ , but also that the disjunctive relation  $EQ|TPP|NTPP$  is the strongest obtainable. Space limitations preclude similarly detailed analysis of the resegmentation of  $c_2$ .

## 5 Discussion

The 12 graphs reveal six resegmentation operations that take us *directly* from  $PO$  to  $PP$  (i.e.,  $TPP|NTPP$ ), hence guaranteeing the transition to  $PP$ ,<sup>4</sup> and another four that merely *allow* that possibility.<sup>5</sup> The number (and complexity) of potential resegmentations increases when several graphs are combined and node-node paths through these networks of length  $n > 2$  are considered, as in the segmentation operations used to generate the cell depicted in Fig. 3(n).

Strategies for selecting optimal resegmentations would be relatively straightforward if the segmented cells were widely separated from each other, but in Fig. 3 this is not so, since several cells are separated by a pixel-width distance and in (f) several nuclei overlap more than one cytoplasm component. Hence when applying  $12_{PO}$  to (g) to generate (i), or applying  $1_{PO}$  to (l) to generate (o), hitherto separated cytoplasm components are merged. One way to avoid this is to restrict the discrete closure operation so as to prevent merging with a neighbouring component:  $\text{cl}_D^-(x, y) =_{\text{def}} \{\hat{x} \mid O(N(\hat{x}), x) \wedge \neg O(N(\hat{x}), y)\}$ . This picks out those pixels whose immediate neighbourhoods overlap  $x$  but are disjoint from  $y$ , giving the largest subset of  $\text{cl}_D(x)$  not connected to  $y$ . It cannot be

<sup>3</sup> Other disjunctions are handled similarly.

<sup>4</sup> Namely  $1_{PO}$ ,  $2_{PO}$ ,  $3'_{PO}$ ,  $4'_{PO}$ ,  $5'_{PO}$ , and  $8'_{PO}$ .

<sup>5</sup> Namely,  $9_{PO}$ ,  $10'_{PO}$ ,  $11'_{PO}$ , and  $12_{PO}$ . In these cases the predicted models include  $PP$ , but  $PP$  is not entailed.

applied *carte blanche*, however, since histological domain constraints may *require* some regions (e.g., fragmented nuclei) to be merged.

Within DM proper, different classes of regions operate as filters, e.g., atoms, regions without interiors, regular regions (lacking spikes or fissures) or connected components; and all these give rise to varying constraints on the mereotopological relations defined on them. At the image scales typically used in digital microscopy, most segmented regions mapping to histological objects have interiors, and very few are atomic, but restricting these topological properties will constrain the possible segmentation models and relations that can be defined on them. Constraints other than those defined directly within DM can be also used to reduce the number of segmentation models. A simple example is where the range of sizes of histological objects can be used as a filter, either using MM granulometry methods or filtering by morphological thickness [1], enabling us to rule out the segmentation models depicted in Fig. 3(i), (n) (cell bodies too large), and (h) (cell nucleus too small).

Another extra-logical constraint that can be used exploits *empirical* information about the histological stains and their known selectivity in dye take-up with respect to targeted tissues and their parts. Given that the H-stain offers better segmentations for nuclei than the eosin counter-stain does for cytoplasm, resegmentations can be ranked so as to favour those that minimise the changes to *nuclei*. Other assumed empirical and ontological dependencies can also be exploited. For example, depending on microscope resolution, each cell nucleus should fall wholly within some cytoplasm component, whether this is segmentable from the original image or not; in DM this can be captured by adding the histological domain axiom  $Nuc(x) \rightarrow \exists y(Cell(y) \wedge P(x, y))$ . This constraint also justifies the assumption underlying the transition  $1_{PO}$ , where a cytoplasm component partially overlapping a nucleus is extended to cover the nucleus so that it forms part of that cell. We can also add the correspondence between histological features and stain selectivity: given a PO relation defined on a poorly segmented nucleus and its host cytoplasm, we favour  $12_{PO}$  (dilating the cytoplasm) over  $9_{PO}$  (eroding the nucleus). Also worthy of note is that the boundaries of histological objects in a greyscale image exhibit intensity gradients. This means that when binary thresholding candidate regions, subsequent erosion and dilation-based resegmentations are more likely to track gray-scale intensity levels in the original image than blind set-theoretic segmentation operations on binary images. This observation highlights a limitation of the underlying method, namely, that once the segmentation mask of the target histological object is provisionally segmented, changes subsequently made and translated back to the images may not conform to all available information in the image. However, in some cases restricting oneself to information in an image may not, even in principle lead to a histological model. For example, uneven staining may fail to reveal the full extent of the cytoplasm in the sample; in which case a model-based resegmentation solution can be used to factor out those regions in an image that need to be treated differently than the rest.

It is also perhaps useful here, to give at least some indication of empirical methods and metrics envisaged for quantifying, validating and measuring our segmentation solutions. For example, given cell nuclei are easier to segment than their associated bodies of cytoplasm, these can be used as a gold-standard reference measuring how well the segmented cytoplasm conforms to the histological model. In this case a necessary condition is that segmented nuclei form part of some overlapping body of cytoplasm, so one possible measure is the number of region pairs that form a part-whole relation divided by all possible overlap cases. These simple examples show (i) that the underlying physical model should guide the abstraction and (ii) the danger of abstracting and working with generic cases too quickly, where empirical constraints restricting valid resegmentations may be missed.

## 6 Conclusions and Future Work

We have shown how DM provides the means to model cellular and tissue structure in digitised histological images. Segmentation and resegmentation satisfying a histological model can be achieved by a set of operations on regions that satisfy a set of constraints on pairs of regions. These constraints can be encoded as a set of graphs in which topological and set-theoretic operators lead from one vertex to another. The method is generic and can be applied to any domain where it is required to segment digitised images into regions satisfying specific sets of mereotopological relations.

Several directions for future work can be suggested. First, the set of operators and associated graphs can be extended to cover all the standard topological and set-theoretic operators, including the discrete exterior, boundary, the (absolute) complement and the symmetric difference. Second, various different metrics can be defined on the conceptual neighbourhoods and their graphs, allowing optimisation of segmentation models and prediction of the most likely path to take through the graphs from a given state to a segmentation goal. These could be based on what proportion of JEPD relations reached at each step can lead to a valid model, or probability measures determined from a statistical analysis of the data sets, taking into account *a priori* and empirically derived properties such as tissue type and morphological shape and size.

**Acknowledgments.** This work was supported by the EPSRC through funding under grant EP/M023869/1, “Novel context-based segmentation algorithms for intelligent microscopy”.

## References

1. Galton, A.: The mereotopology of discrete space. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 251–266. Springer, Heidelberg (1999). doi:[10.1007/3-540-48384-5.17](https://doi.org/10.1007/3-540-48384-5.17)

2. Galton, A.: Discrete mereotopology. In: Calosi, C., Graziani, P. (eds.) *Mereology and the Sciences: Parts and Wholes in the Contemporary Scientific Context*, vol. 371, pp. 293–321. Springer, Cham (2014). doi:[10.1007/978-3-319-05356-1\\_11](https://doi.org/10.1007/978-3-319-05356-1_11)
3. Landini, G., Randell, D.A., Galton, A.: Discrete mereotopology in histological imaging. In: Claridge, E., Palmer, A.D., Pitkeathly, W.T.E. (eds.) *Proceedings of the 17th Conference on Medical Image Understanding and Analysis*, pp. 101–106 (2013). ISBN 1-901725-48-0
4. Landini, G., Randell, D.: The complexity of cellular neighbourhoods. *Fifth International Symposium on Fractals in Biology and Medicine, Locarno, Switzerland (2008)*. *Riv. di Biol./Biol. Forum*, **101**(1), 129–158 (2008)
5. Randell, D., Landini, G.: Discrete mereotopology in automated histological image analysis. In: *Proceedings of the Second ImageJ User and Developer Conference, Luxembourg*, pp. 151–156 (2008)
6. Randell, D., Landini, G., Galton, A.P.: Discrete mereotopology in automated histological image analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 568–581 (2013)
7. Bloch, I.: Spatial reasoning under imprecision using fuzzy set theory, formal logics and mathematical morphology. *Int. J. Approx. Reason.* **41**, 77–95 (2006)
8. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, pp. 165–176 (1992)
9. Li, S., Ying, M.: Generalised region connection calculus. *Artif. Intell.* **160**(1–2), 1–34 (2004)
10. Cui, Z., Cohn, A.G., Randell, D.A.: Qualitative simulation based on a logic of space and time. In: *Proceedings of the AAAI 1992*, pp. 679–684 (1992)
11. Sioutis, M., Condotta, J.-F., Salhi, Y., Mazure, B., Randell, D.A.: Ordering spatio-temporal sequences to meet transition constraints: complexity and framework. In: Chbeir, R., Manolopoulos, Y., Maglogiannis, I., Alhajj, R. (eds.) *AIAI 2015. IAICT*, vol. 458, pp. 130–150. Springer, Cham (2015). doi:[10.1007/978-3-319-23868-5\\_10](https://doi.org/10.1007/978-3-319-23868-5_10)
12. Chen, J., Cohn, A.G., Liu, D., Wang, S., Ouyang, J., Yu, Q.: A survey of qualitative spatial representations. *Knowl. Eng. Rev.* **30**(1), 106–136 (2015)
13. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009)
14. Sirinukunwattana, K., Pluim, J.P.W., Chen, H., Qi, X., Heng, P.-A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., Böhm, A., Ronneberger, O., Cheikh, B.B., Racoceanu, D., Kainz, P., Pfeiffer, M., Urschler, M., Snead, D.R.J., Rajpoot, N.M.: Gland segmentation in colon histology images: the glas challenge contest. *Med. Image Anal.* **35**, 489–502 (2016)
15. Li, S., Ying, M.: Region connection calculus: its models and composition table. *Artif. Intell.* **145**(1–2), 121–146 (2003)
16. Weidenbach, C., Dimova, D., Fietzke, A., Kumar, R., Suda, M., Wischniewski, P.: SPASS version 3.5. In: *Proceedings of the 22nd International Conference on Automated Deduction*, pp. 140–145 (2009)
17. Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**(4), 291–299 (2001)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

