Uncertainty Estimation in Vascular Networks

Markus Rempfler[1,2], Bjoern Andres[3], and Bjoern H. Menze[1,2]

[1] Institute for Advanced Study, Technical University of Munich, Germany
[2] Department of Informatics, Technical University of Munich, Germany
[3] Bosch Center for Artificial Intelligence (BCAI)

Abstract. Reconstructing vascular networks is a challenging task in medical image processing as automated methods have to deal with large variations in vessel shape and image quality. Recent methods have addressed this problem as constrained maximum a posteriori (MAP) inference in a graphical model, formulated over an overcomplete network graph. Manual control and adjustments are often desired in practice and strongly benefit from indicating the uncertainties in the reconstruction or presenting alternative solutions. In this paper, we examine two different methods to sample vessel network graphs, a perturbation and a Gibbs sampler, and thereby estimate marginals. We quantitatively validate the accuracy of the approximated marginals using true marginals, computed by enumeration.

1 Introduction

Vessel segmentation and centerline extraction is a longstanding problem in computer vision [1]. From a medical perspective, segmenting and tracking vessels is crucial for planning and guiding several types of interventions. Several recent methods, however, have focussed on reconstructing vessel network graphs [2, 3, 4, 5, 6]. Analysing vascular graphs is expected to give insights into various biological properties, e.g. the relation between vascular remodeling processes and neurological diseases or pharmaceutical treatments [7]. These methods formulate the task as MAP inference in a constrained probabilistic model over a (super-)graph of candidate vasculature, where the solution encodes the subgraph that is most likely to represent the underlying vasculature. Variations of this approach include joint-tasks such as anatomical labeling of vasculature [6] or artery-vein separation [5].

As in many applications, exploring multiple solutions or even marginal distributions would be preferable over mere point estimates – either to present local uncertainty to the end user or to pass it over to the next stage of the processing pipeline. An automated reconstruction can be inspected and, if needed, edited by an expert. In such a workflow, the controlling expert benefits from an indication of the uncertainty in the presented reconstruction (cf. Fig. 1). To this end, recent work investigated how to find the m-best diverse solutions to the MAP problem in conditional random fields (CRFs) to explore a variety of highly probable assignments [8, 9]. This approach, however, increases the computational complexity of the discrete optimization further. Alternatively, Markov

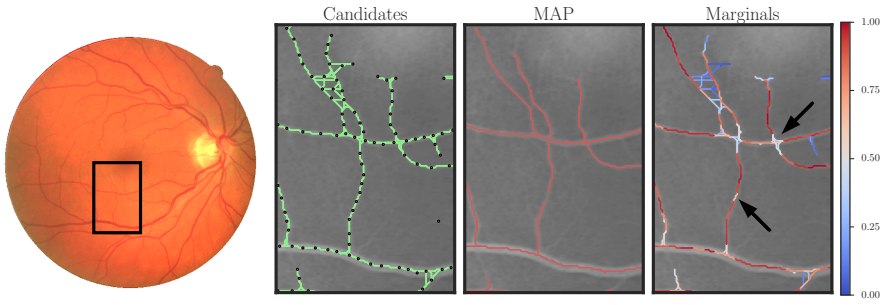

Fig. 1. Illustration of the uncertainty quantification in vasculature graphs from a 2D retinal image (**left**). Recent methods reconstruct the network from an overcomplete graph of candidate vessels (**second**, graph in green) by calculating the MAP state (**third**, graph in red) in a probabilistic model. Approximating marginal distributions (**right**) enables us to quantify the uncertainty in the network graph, which is valuable information for manual inspection and correction. Two examples are indicated with black arrows: In the first, the model is uncertain whether it is a furcation or a crossing, while in the second, a connection is not contained in the MAP but still has a high marginal probability.

chain Monte Carlo (MCMC) methods can be used to sample from probabilistic models [10, 11]. While being well established for many statistical inference tasks, they are often considered expensive and difficult to parametrize for typical problems in computer vision. Papandreou and Youille [12] presented the idea to introduce local perturbations and solve for the MAP estimate of the perturbed model repeatedly to generate samples. They identify a perturbation distribution which allows to estimate marginal densities of the original Gibbs distribution while leveraging the computational efficiency of available discrete solvers. This idea was extended to a broader problem class in [13], while the theoretical framework was further developed in [14, 15, 16, 17]. A few empirical studies investigated the effectiveness of such perturbation models in typical segmentation problems [15, 18, 19].

In this paper, we extend recent graph-based methods for reconstructing vascular networks that rely on integer progamming. We adapt two sampling approaches for the underlying probabilistic model, a perturbation sampler based on [12, 14, 15, 13] and a Gibbs sampler based on [10, 20]. They enable estimates of marginal distributions and a straight-forward way to quantify uncertainty in properties calculated from the resulting network graphs. To deal with the difficulty of validating the quality of the approximated marginals, we compare the approximated marginals to the true marginals, calculated by enumeration.

2 Background

Several recent methods for vessel network reconstruction pose the problem as MAP inference in a (constrained) probabilistic model over a supergraph com-

posed of candidate vessels [2, 3, 4, 5, 6]. In short, such a candidate supergraph is typically constructed by detecting points that are likely to lie on a vessel centerline, composing the nodes $v \in V$ of the graph, and then inserting an edge $e \in E$ for each path that connects two nodes in close proximity. The MAP state then encodes a subgraph and thereby represents which parts of the candidate supergraph are present in the reconstruction. Calculating this MAP state can be formulated as an integer linear program (ILP) and solved by a branch-and-cut procedure. In the remainder of this section, we first describe such probabilistic model for vessel graphs and its MAP estimator. Details on the particular choice of candidate graph construction used in this study can be found in Sec. 4.

Probabilistic Model. Given a (directed) candidate graph $G = (V, E)$, we define a measure of probability $P(\mathbf{X} = \mathbf{x}|\Omega, I, \boldsymbol{\Theta})$ over possible vessel networks within G, encoded by $\mathbf{x} \in \{0, 1\}^E$. These indicator variables then encode whether an edge e is present in the solution ($x_e = 1$) or not ($x_e = 0$). We denote the set of *feasible* solutions as Ω, the image evidence as I and the model parameters as $\boldsymbol{\Theta}$. The measure of probability can be defined as:

$$P(\mathbf{x}|\Omega, I, \boldsymbol{\Theta}) \propto P(\Omega|\mathbf{x}) \prod_{ij \in E} P(x_{ij}|I, \boldsymbol{\Theta}) \prod_{C \in \mathcal{C}(G)} P(x_C|\boldsymbol{\Theta}), \tag{1}$$

$$\text{where } P(\Omega|\mathbf{x}) \propto \begin{cases} 1 & \text{if } \mathbf{x} \in \Omega, \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

We identify three parts: First, $P(\Omega|\mathbf{x})$ is the uniform prior over all feasible solutions. Second, $P(x_e|I, \boldsymbol{\Theta})$ is the local evidence for an edge, i.e. the *unaries*. Third, $P(x_C|\boldsymbol{\Theta})$ corresponds to joint-events C that form higher-level potentials, and $\mathcal{C}(G)$ denotes the set of all events at any possible location within G. $x_C = 1$ indicates that the particular event C occurred.

In [2, 3, 4, 5, 6], these different parts have been chosen depending on the particular image datasets and target application of the reconstructed vasculature. For this study, we will impose the following constraints: each node can have at most one incoming edge and at most two outgoing edges. Furthermore, we do not allow the solution to contain circles. These three types of constraints define our Ω. As higher-level events x_C, we consider *appearance*, *termination* and *bifurcation* in each node, leaving us with at most $3|V|$ possible events in $\mathcal{C}(G)$. These events can be represented with binary indicator variables x_C and a set of $3|V|$ auxiliary constraints that tie their state to the original edge variables $\mathbf{x}$ upon which they depend. Note that the number of involved edge variables of a particular type of event varies with its location within G: For example, a bifurcation event at node v involves all x_e of potential outgoing edges $e \in \delta^-(v)$. We denote the set of auxiliary constraints necessary for higher-level events as Ω_A in the remainder of this section. The description of both Ω and Ω_A in terms of linear inequalities can be found in the supplement.

MAP Estimator. Using the bilinear representation of the pseudo-boolean probability functions $P\left(x_{ij}|I,\boldsymbol{\Theta}\right)$ and $P\left(x_C|\boldsymbol{\Theta}\right)$, we can formulate the MAP estimator to (1) as ILP:

$$\text{minimize} \qquad \sum_{(i,j)\in E} w_{ij}x_{ij} + \sum_{C\in\mathcal{C}(G)} w_C x_C \tag{3}$$

$$\text{s.t.} \qquad \mathbf{x}\in\Omega,\ \ [\mathbf{x},\mathbf{x}_C]\in\Omega_A,\ \ x\in\{0,1\}\ , \tag{4}$$

where $w_{ij} = -\log\frac{P(x_{ij}=1|I,\boldsymbol{\Theta})}{1-P(x_{ij}=1|I,\boldsymbol{\Theta})}$ and $w_C = -\log\frac{P(x_C=1|\boldsymbol{\Theta})}{1-P(x_C=1|\boldsymbol{\Theta})}$. The constraint $\mathbf{x}\in\Omega$ is due to $P\left(\Omega|\mathbf{x}\right)$ and $[\mathbf{x},\mathbf{x}_C]\in\Omega_A$ ties auxiliary variables for the events to the edge variables $\mathbf{x}$. Finally, all variables are binary. This ILP can be optimized with the branch-and-cut algorithm. Certain types of constraints contained in Ω may consist of an extensive number of inequalities (e.g. the cycle-free constraint). In this case, we employ a lazy constraint generation strategy: Whenever the solver arrives at an integral solution $\mathbf{x}'$, we check for violated constraints in the corresponding solution, add them if required and reject $\mathbf{x}'$. If no violation is found, i.e. $\mathbf{x}'$ is already a feasible solution, then it is accepted as new current solution $\mathbf{x}^*$. For our set of constraints Ω, we use this scheme for the cycle constraints, where we identify strongly connected components efficiently with [21] and add the violated constraints for the cycles within them. All other constraints for incoming and outgoing edges, as well as auxiliaries can be added to the optimization model from the start.

3 Uncertainty Estimation by Means of Sampling

3.1 Perturbation Sampler

Following the work of [12, 14, 15], a perturbation model is induced by perturbing the energy function of a random field and solving for its (perturbed) MAP state:

$$P\left(\hat{\mathbf{x}}|I,\boldsymbol{\Theta}\right) = P_\gamma\big(\hat{\mathbf{x}}\in\operatorname*{arg\,min}_{\mathbf{x}\in\Omega} E(\mathbf{x};I,\boldsymbol{\Theta})+\gamma(\mathbf{x})\big)\ , \tag{5}$$

where $E(\mathbf{x},I,\boldsymbol{\Theta})$ is the energy function of the random field and $\gamma\left(\mathbf{x}\right)$ is the perturbation. It was shown that if the full potential table is perturbed with IID Gumbel-distributed samples of zero mean, then the perturbation model and the Gibbs model coincide [12]. In practice, this is not feasible. The full potential table may be too large and it destroys local Markov structure, rendering optimization difficult. However, it was shown in several studies that even first order Gumbel perturbations yield sufficiently good approximations [12, 15]. In this case, only the unary potentials are perturbed and hence, the perturbation $\gamma\left(\mathbf{x}\right)$ becomes:

$$\gamma\left(\mathbf{x}\right) = \sum_{i=1}^{N}\sum_{l\in\mathcal{L}} \gamma_i^l \mathbb{1}(x_i=l)\ , \tag{6}$$

with γ_i^k being IID samples from the Gumbel distribution [22] with zero mean and variance $\frac{\pi^2}{6}$, and $\mathbb{1}(.)$ is the indicator function. Sampling from the perturbation

model then boils down to drawing a new perturbation $\gamma(\mathbf{x})$ and determining the new MAP state. Having a procedure to sample efficiently from the model enables us to estimate marginal distributions of variables (and variable subsets) as well as derived measures of uncertainty. We refer the interested reader to [12, 13, 14, 15, 16] for further information on perturbation models.

We next derive the first-order perturbed objective for the MAP estimator in (3). First, we note that two states will need two independent gumbel samples $\gamma^1_{ij}, \gamma^0_{ij}$ according to (6). Our MAP estimator, however, uses only one binary variable to encode both states. We use again the bilinear representation of the pseudo-boolean functions to find that perturbing the unaries adds a difference of the two independent gumbel samples, i.e. $\Delta\gamma_{ij} = (\gamma^1_{ij} - \gamma^0_{ij})$, to the original weight w_{ij}. The first-order perturbed objective of (3) is thus:

$$\sum_{(i,j)\in E} (w_{ij} + \Delta\gamma_{ij})x_{ij} + \sum_{C\in\mathcal{C}(G)} w_C x_C \quad . \tag{7}$$

Drawing a sample from our probabilistic model therefore boils down to constructing a new perturbed objective (with a new set of $\Delta\gamma_{ij}$) and optimizing the according ILP with the original constraints (4) and (7) instead of (3). This can be implemented by changing the coefficients of the optimization problem for each new perturbation. We note that we can warm-start the optimization with the previous solution and that we can keep previously generated constraints since they are not depending on the weights but only on the structure of G and thus, remain valid.

3.2 Gibbs Sampler

As alternative to the perturbation sampling, we employ a Gibbs sampler [10], a method of the MCMC family. We apply the following two modifications described in [20] to obtain a *metropolized* variant of the Gibbs sampler, which is expected to be more efficient for discrete problems. 1) variables are sampled in random-scan fashion within each sweep, and 2) the acceptance probability is replaced with the Metropolis-Hastings acceptance probability

$$\alpha = \min\left(1, \frac{1-\pi(x_e|\mathbf{x}_{\setminus e})}{1-\pi(x'_e|\mathbf{x}_{\setminus e})}\right) \quad , \tag{8}$$

where $\pi(x_e|\mathbf{x}_{\setminus e})$ and $\pi(x'_e|\mathbf{x}_{\setminus e})$ are the conditional probabilities of current and proposed state. To cope with the extra constraints of Ω, we can employ the same procedures to identify violated constraints as within the branch-and-cut algorithm. In this case, however, it suffices to check only those constraints which involve the changed variable(s). Changes that render the state infeasible with respect to Ω have a zero probability and will thus always be rejected. Auxiliary variables x_C for higher-level events need not to be sampled but can be determined directly from the current state $\mathbf{x}$ using the relationship encoded by the auxiliary constraints Ω_A. After a burn-in period of 1000 sweeps, we run one sweep for each sample.

4 Experiments & Results

We conduct our experiments on retinal images [23]. In the first part of this section, we detail on the preprocessing, i.e. the candidate vessel graph construction. In the second part, we then present both quantitative and qualitative results of the two sampling approaches. We address the difficulty of validating marginal distribution estimates by computing exact marginals on smaller problem instances, where brute-force enumeration of all states is computationally possible.

Candidate Graph Construction. As a first step, we need to propose vasculature in terms of an overcomplete candidate graph $G = (V, E)$. We rely on the following scheme to achieve this, which is mainly based on [2, 3, 24]:

1. *Centerline detection.* We compute a centerline score $f_{\text{cl}}(I)$ for the entire image using a regression approach based on [24]. High centerline scores indicate the presence of the centerline of the vessel.
2. *Candidate node selection.* We construct a collection of candidate nodes V by iteratively selecting the locations with the highest value in the centerline measure map and suppressing its neighbourhood within a radius r_{sup} until no more locations with a value larger than θ_T are left.
3. *Connection of candidates.* Next, we reconnect previously selected candidate nodes to its N closest neighbours using Dijkstra's algorithm on the centerline score map. A connection between two nodes $i, j \in V$ then forms an edge $(i, j) \in E$ in the vessel candidate graph. Connections that pass through a third candidate node are discarded as they would introduce unnecessary redundance. To save computation time, we limit the maximum search radius to r_s.

In these experiments, we set $r_{\text{sup}} = 5\,\text{px}$ and $\theta_T = 0.3 \max f_{\text{cl}}(I)$ for the candidate selection, and $N = 4$ and $r_s = 30\,\text{px}$ for the edge construction. We use a discriminative path classifier to estimate $P\left(x_{ij} = 1|I\right)$, i.e. how likely edge $ij \in E$ belongs to the graph or not, which is then used to calculate the weights w_{ij}. To this end, we use gradient boosted decision trees with 5 features calculated along the path: length, tortuosity, cumulative f_{cl}, $\min f_{\text{cl}}$ and standard deviation of f_{cl}. Additional details on both centerline regressor and path classifier can be found in the supplement. For each class of events, appearance, termination and bifurcation, we introduce one parameter θ^a, θ^t and θ^b as constant weight for the respective event happening at a given node, and set them to $\theta^a = 0.5$, $\theta^t = 0.1$ and $\theta^b = 0.1$.

Comparison. In order to quantitatively validate the marginals that we approximate by using the perturbation sampler, we set up a series of 15 small test graphs with $|E| \leq 20$ from the test images of [23], such that we are able to enumerate all feasible states and thereby obtain exact marginals by brute force. We then compare these exact marginals to the approximate marginals obtained

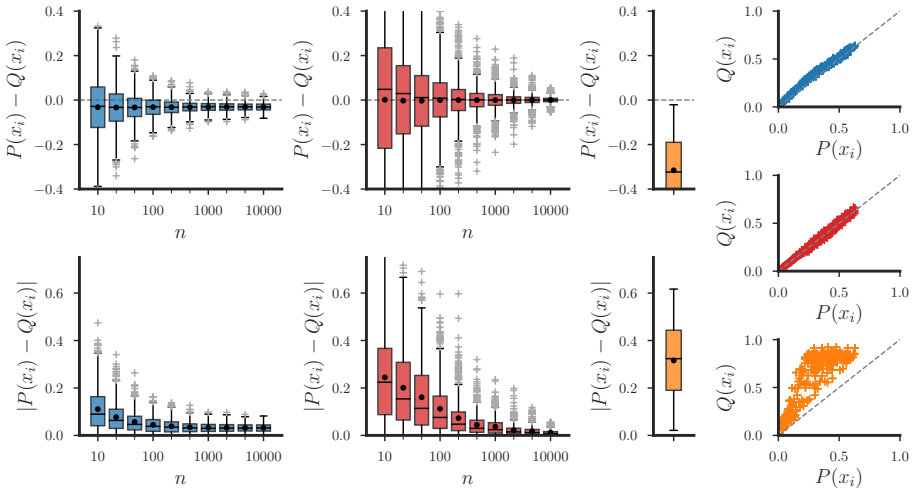

Fig. 2. Comparison of the approximated marginals $Q(x_i)$ with the exact marginals $P(x_i)$ calculated by brute force enumeration of all states. Approximates are obtained from the perturbation sampler (**blue**), from the Gibbs sampler (**red**), and from the raw classifier probability (**orange**). The figure shows deviation $P(x_i) - Q(x_i)$ (**top row**) and absolute deviation $|P(x_i) - Q(x_i)|$ (**bottom row**) of the marginal estimate with increasing number of samples n. Boxplots denote the median with a black bar, the mean value with a black dot and outliers with a grey cross. **Right column:** Scatter plot of exact marginal probabilities $P(x_i)$ versus approximated marginal probabilities $Q(x_i)$. We observe that the perturbation sampler converges to an absolute bias of about 0.032 on average and has the tendency to overestimate the marginal probabilites slightly. The Gibbs sampler does not exhibit such a systematic bias, but needs more samples to reduce its variance. Using the probabilistic output of the local classifier as an approximate to the marginals is considerably less accurate than both sampling approaches.

by both perturbation and Gibbs sampler. We solve the ILP of our MAP estimator by the branch-and-cut algorithm of [25] and implement the lazy constraint generation as callback. We use the default relative optimality gap of 10^{-4}.

In Fig. 2, we compare the approximated marginals from our perturbation sampler with exact marginals. We sample 10000 samples per case in total and repeat the experiment 5 times. We observe that the absolute deviation of the approximated from the exact marginals converges already at about 1000 samples to an absolute error of $|P(x_i) - Q(x_i)| \approx 0.032$ on average and the perturbation sampler shows a tendency to overestimate the marginal probabilities. Such a systematic bias is to be expected, as we apply a low-order perturbation instead of the (intractable) full perturbation. The Gibbs sampler does not exhibit such systematic bias, yet shows a larger variance when fewer samples are aquired. With 10000 samples, its mean absolute approximation error is 0.012 and therefore better than the perturbation sampler. Wilcoxon signed-rank tests for each

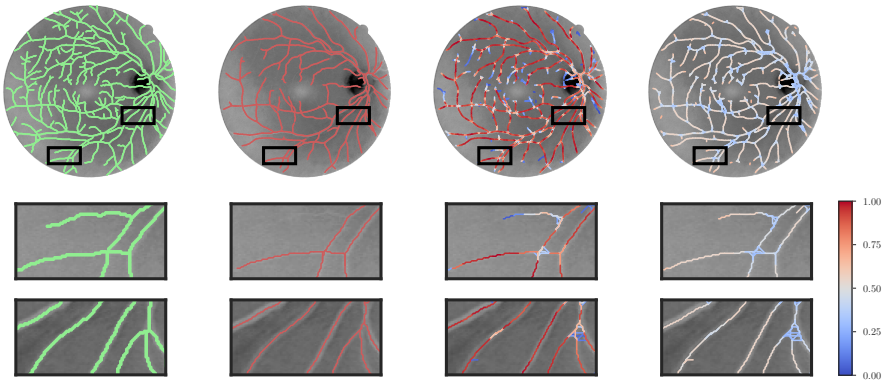

Fig. 3. Visualisation of the vascular network graphs overlaid on the (grey-scale) input image. **From left to right:** Pixel-based centerline obtained by skeletonizing the ground truth segmentation, MAP reconstruction, approximated marginals using the perturbation sampler, and the Gibbs sampler. The colorbar applies only to the marginals of two right columns, where we show the marginal $P(x_{ij} = 1 \vee x_{ji} = 1)$, i.e. the probability of either edge being active, for better visibility. We find that the marginals of the perturbation sampler indicate uncertainty in small bifurcations and point out the (possible) presence of weak terminal branches, which would be discarded if we only consider the MAP solution. The Gibbs sampler displays overall a higher uncertainty on such large graphs.

fixed number of samples n indicate that the approximation errors of perturbation and Gibbs sampler are significantly different ($p < 0.001$), with the exception of $n = 1000$ where both show similar errors. Using the probabilities of the path classifier directly as an approximate marginal is considerably worse than both sampling approaches. Note that the exact marginals for our test cases do not exhibit very high values (cf. Fig. 2, right column) due to the fact that for these small graphs, often no direction is strongly dominating and thus, several solutions that contain similar physical paths but in different orientations are competing.

A qualitative visualisation of the approximated marginals on complete graphs is given in Fig. 3. We draw 100 perturbation samples, which we found a reasonable trade-off between computation time and informativeness of the marginals, and slightly increase the relative optimality gap to $5 \cdot 10^{-3}$ to prevent the branch-and-cut solver from spending too much time proofing optimality. From the Gibbs sampler, we draw 10000 samples after a burn-in period of 1000. We find that the marginals from the Gibbs sampler display overall a higher uncertainty in the graph than the perturbation samples, which could be due to more difficult transitions between different modi of the distribution and would likely require adapted sampling parametrization or even an extension of the set of allowed transformations. In both cases, thresholding the marginal distributions $P(x_e)$ has no guarantee to satisfy all constraints and is therefore not recommended for obtaining a single reconstruction. To improve a reconstruction, an interactive

procedure using the uncertainties (and individual samples) would be advisable, and for downstream analysis, metrics of interest should be calculated on each sample. Regarding computation time, the average runtime per sample is 7.85 s for the perturbation and approximately 0.01 s for the Gibbs sampler (not including any additional overhead caused by the burn in period). The perturbation sampler spends on average 0.5 % of its runtime in the lazy constraint generation where violated cycle inequalities are identified.

5 Conclusion

We adapted two sampling approaches for vascular network graph reconstruction models, a perturbation sampler and a Gibbs sampler. Our experiments confirm the expected systematic bias of the perturbation sampler due to the computationally cheaper low-order perturbations. The Gibbs sampler, on the other hand, exhibits an unbiased behaviour but instances with varying properties might require an appropriately adapted parametrization. The perturbation approach benefits from not having a burn in period, which renders it considerably easier to use on large instances. Both approaches were shown to be more informative than the predictive probabilities of local classifier and can be used to approximate marginals or determine the uncertainty in network graph properties. Beyond this, the two sampling procedures could be employed within a Bayesian model selection framework or for maximum-likelihood hyperparameter estimation.

Acknowledgements. With the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative (and the European Union Seventh Framework Programme under grant agreement n° 291763).

References

1. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Medical Image Analysis 13(6), 819–845 (2009)
2. Türetken, E., Benmansour, F., Andres, B., Glowacki, P., Pfister, H., Fua, P.: Reconstructing curvilinear networks using path classifiers and integer programming. IEEE Transactions on Pattern Analysis and Machine Intelligence 38(12), 2515–2530 (2016)
3. Rempfler, M., Schneider, M., Ielacqua, G.D., Xiao, X., Stock, S.R., Klohs, J., Székely, G., Andres, B., Menze, B.H.: Reconstructing cerebrovascular networks under local physiological constraints by integer programming. Medical Image Analysis 25(1), 86 – 94 (2015)
4. Rempfler, M., Andres, B., Menze, B.H.: The minimum cost connected subgraph problem in medical image analysis. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III. pp. 397–405. Springer International Publishing, Cham (2016)

5. Payer, C., Pienn, M., Blint, Z., Shekhovtsov, A., Talakic, E., Nagy, E., Olschewski, A., Olschewski, H., Urschler, M.: Automated integer programming based separation of arteries and veins from thoracic ct images. Medical Image Analysis 34, 109 – 122 (2016)
6. Robben, D., Türetken, E., Sunaert, S., Thijs, V., Wilms, G., Fua, P., Maes, F., Suetens, P.: Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Medical Image Analysis 32, 201–215 (2016)
7. Klohs, J., Baltes, C., Princz-Kranz, F., Ratering, D., Nitsch, R.M., Knuesel, I., Rudin, M.: Contrast-enhanced magnetic resonance microangiography reveals remodeling of the cerebral microvasculature in transgenic arcaβ mice. Journal of Neuroscience 32(5), 1705–1713 (2012)
8. Batra, D., Yadollahpour, P., Guzman-Rivera, A., Shakhnarovich, G.: Diverse m-best solutions in Markov Random Fields. In: ECCV 2012, LNCS, vol. 7576, pp. 1–16. Springer Berlin Heidelberg (2012)
9. Kirillov, A., Savchynskyy, B., Schlesinger, D., Vetrov, D., Rother, C.: Inferring m-best diverse labelings in a single one. In: IEEE International Conference on Computer Vision (ICCV). pp. 1814–1822 (2015)
10. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6(6), 721–741 (1984)
11. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. MIT press (2009)
12. Papandreou, G., Yuille, A.L.: Perturb-and-MAP random fields: Using discrete optimization to learn and sample from energy models. In: International Conference on Computer Vision 2011. pp. 193–200 (2011)
13. Tarlow, D., Adams, R.P., Zemel, R.S.: Randomized optimum models for structured prediction. In: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, vol. 22, pp. 1221–1229 (2012)
14. Hazan, T., Jaakkola, T.: On the partition function and random maximum a-posteriori perturbations. In: Proceedings of the 29th International Conference on Machine Learning (ICML-12), pp. 991–998 (2012)
15. Hazan, T., Maji, S., Jaakkola, T.: On sampling from the gibbs distribution with random maximum a-posteriori perturbations. Advances in Neural Information Processing Systems pp. 1268–1276 (2013)
16. Orabona, F., Hazan, T., Sarwate, A., Jaakkola, T.: On measure concentration of random maximum a-posteriori perturbations. In: International Conference on Machine Learning. pp. 432–440 (2014)
17. Gane, A., Hazan, T., Jaakkola, T.: Learning with maximum a-posteriori perturbation models. In: Artificial Intelligence and Statistics. pp. 247–256 (2014)
18. Alberts, E., Rempfler, M., Alber, G., Huber, T., Kirschke, J., Zimmer, C., Menze, B.H.: Uncertainty quantification in brain tumor segmentation using CRFs and random perturbation models. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). pp. 428–431 (2016)
19. Meier, R., Knecht, U., Jungo, A., Wiest, R., Reyes, M.: Perturb-and-MPM: Quantifying segmentation uncertainty in dense multi-label CRFs. CoRR abs/1703.00312 (2017), http://arxiv.org/abs/1703.00312
20. Liu, J.S.: Monte Carlo strategies in scientific computing. New York: Springer (2001)
21. Mehlhorn, K., Näher, S., Sanders, P.: Engineering DFS-based graph algorithms. CoRR abs/1703.10023 (2017), http://arxiv.org/abs/1703.10023
22. Gumbel, E.J.: Statistical theory of extreme values and some practical applications: a series of lectures. No. 33, US Govt. Print. Office (1954)

23. Staal, J.J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging 23(4), 501–509 (2004)
24. Sironi, A., Türetken, E., Lepetit, V., Fua, P.: Multiscale centerline detection. IEEE Transactions on Pattern Analysis & Machine Intelligence 1, 1–14 (2015)
25. Gurobi Optimization, I.: Gurobi Optimizer Reference Manual (2017), http://www.gurobi.com