Tree-LSTM: Using LSTM to Encode Memory in Anatomical Tree Prediction from 3D Images

Mengliu Zhao(✉) and Ghassan Hamarneh

Medical Image Analysis Lab, School of Computing Science,
Simon Fraser University, Burnaby, Canada
{mengliuz,hamarneh}@sfu.ca

Abstract. Extraction and analysis of anatomical trees, such as vasculatures and airways, is important for many clinical applications. However, most tracking methods so far intrinsically embedded a first-order Markovian property, where no memory beyond one tracking step was utilized in the tree extraction process. Motivated by the inherent sequential construction of anatomical trees, vis-à-vis the flow of nutrients through branches and bifurcations, we propose Tree-LSTM, the first LSTM neural network to learn to encode such sequential priors into a deep learning based tree extraction method. We also show that mathematically, by using LSTM, the variational lower bound of a higher order Markovian stochastic process could be approximated, which enables the encoding of a long term memory. Our experiments on a CT airway dataset show that, by adding the LSTM component, the results are improved by at least 11% in mean direction prediction accuracy relative to state-of-the-art, and the correlation between bifurcation classification accuracy and evidence is improved by at least 15%, which demonstrate the advantage of a unified deep model for sequential tree structure tracking and bifurcation detection.

1 Introduction

Vascular and airway trees, i.e., anatomical trees, of circulatory and respiratory systems are extremely important clinically as they are related to fatal diseases like ischaemic heart disease, stroke, chronic obstructive pulmonary disease, and lower respiratory infection, which were identified as the top four causes of death globally, according to a report from World Health Organization in 2018 [16]. In the field of medical image analysis, anatomical tree extraction from 3D medical imaging data is a crucial task, as accurate extraction of anatomical trees could be further utilized in tasks such as surgery planning [12], early diabetic diagnosis [6], and tumor type classification [2].

Most tree extraction approaches fall into one of the following categories [8,22]: (a) active contour/surface methods (or deformable energy minimizing models), which expand, commonly via a variational framework, a curve or surface to fit the

H.-I. Suk et al. (Eds.): MLMI 2019, LNCS 11861, pp. 637–645, 2019.
https://doi.org/10.1007/978-3-030-32692-0_73

desired vessel/airway boundaries; (b) tracking methods (described below); (c) minimal path approaches, which determine the path of a vessel/airway branch as the route with minimal energy between given points at both ends of the branch; (d) graph based methods, which represent the anatomical tree as a tree graph and model the tree extraction as a combinatorial optimization problem; (e) machine learning and deep learning methods that learn various aspects of the tree extraction process (e.g., bifurcation detection) from training data; and (f) other hybrid methods.

The tracking methods cover a large portion of the tree extraction literature. At a high level, this class of approaches involves starting from one or multiple seed points, followed by iteratively detecting tracking directions, and identifying new candidate branch points (usually along the branch centerline), until the whole tree is tracked. Cetin et al. [3] proposed to use a tensor based direction prediction technique for coronary vessel tracking with a "shape prior" encoding the tube-like geometry of branches. Wang et al. [15] proposed to use Bayesian estimation in the tracking process and the vessel likelihood was estimated iteratively using gradient flux on the cross sectional plane. Lesage et al. [9] proposed to use particle filtering, which adopted the sequential importance sampling technique to sample candidate particles in a recursive way. Lee et al. [7] also proposed to use particle filtering for vessel tracking, only that they leveraged a Chan-Vese model at each prediction. There have also been a rising number of works encoding machine/deep learning techniques [17,20,21] in the tree tracking process.

Examining most tracking methods from the stochastic modelling perspective, we note that they share a common underlying first-order Markovian model assumption, which means the current prediction is only directly affected by the last tracking step (commonly the image features therein), ignoring direct relations with previous points along the tracked trajectory [3,7,9,15,17,20,21]. Even though a perfect Monte-Carlo estimation in the particle filtering approaches is supposed to involve all previous particles for importance sampling, however, in most particle updating scheme implementations [7,9], a first-order Markovian assumption is usually adopted to reduce complexity. Very limited works have tried to encode higher-order Markovian models or global information into the tracking process. Zhao et al. [22] proposed adopting a Bayesian inference formulation to encode branch-wise statistics (tree topology, branch length, and angle geometrical priors), however, in the form of a prior term instead of the interaction among previous predictions.

Deep learning architectures have been explored extensively in recent years in the field of computer vision and medical image analysis in general, and to a lesser degree for anatomical tree extraction. The work of Wu et al. [18] used an AlexNet classifier with PCA and nearest neighbor search to classify whether the current point in a 2D retinal image is a bifurcation. Charbonnier et al. [4] proposed to combine the results of three ConvNets on orthogonal planes to detect leaks in airway segmentation. Wolterink et al. [17] used a seven-layer dilated convolutional neural network (DCNN) to predict coronary vessel directions in 3D patches. Zhao et al. [21] proposed to use a modified V-net [10] for airway direction

prediction, by using a custom multi-loss function. Selvan et al. [13] proposed to use a K-nearest neighbour voxel-wise classifier for airway segmentation, then a 10-layer mean field network was used later for tree graph refinement.

Although long short-term memory (LSTM) architectures have been applied to computer vision and medical image analysis problems to perform inference on sequential data by exploiting temporal information in 2D+time and 3D+time data [5,14], no prior work has been done using LSTM for vasculature and airway structure prediction. To the best of our knowledge, the only work that adopted LSTM for anatomical tree analysis is the work of Wu et al. [19], in which they labelled an already segmented abstract representation of a coronary vessel tree using a bidirectional LSTM per branch, whereas the focus of our work is to segment or extract trees form a 3D images.

In this work, we argue that the tree tracking process goes beyond a first-order Markovian and leverage the global information along the pseudo-temporal (sequence) dimension in combination with the proven capabilities of deep networks to automatically learn appearance features. To this end, we propose Tree-LSTM, the first neural network LSTM architecture to learn to encode such sequential priors into a deep learning tree extraction method. Our proposed method is the first work in the field of anatomical tree analysis where: (a) a formal derivation of the higher-order Markovian process is given; (b) an LSTM-equipped convolutional neural network (CNN) is used for approximating the higher-order Markovian process; (c) a novel evaluation method is proposed for inspecting the correlation between bifurcation classification accuracy and sequential image evidence collected along branches. By using the proposed architecture, we show that the bifurcation and direction prediction accuracy could be improved by a large margin.

2 Methodology

2.1 Problem Definition

Tree tracking is predominantly formulated as a centerline tracking problem, wherein the vessel/airway direction and bifurcation existence at each tracked point is inferred, and used to estimate the next candidate centerline point, and child branches are spawned whenever a bifurcation point is encountered. Specifically, we focus on the inference problem of each tracking step.

Let C_t be a random variable whose value, e.g., branch direction or bifurcation presence, needs to be estimated at the branch centerline point of step t. A realization of C_t is denoted C_t^*, and I_t be the corresponding image feature learnt from a sequence of image patches. Then the Maximum A Posteriori formulation for estimating C_t is given by:

$$C_t^* = \arg\max_{C_t} P(C_t|C_{t-1}, C_{t-2}, ..., C_1, I_t, I_{t-1}, ..., I_1). \quad (1)$$

In the following, we show that although the maximization of this posterior probability can be intractable (Sect. 2.2), we are able to maximize its variational

lower bound instead (Sect. 2.3), a common and effective trick utilized in variational inference [1]. We find the solution to this lower bound problem using LSTM (Sect. 2.4).

2.2 Central Theorem

First we prove Lemma 1 and Theorem 1, which essentially state that the optimization in (1) could be reformulated as a simpler optimization (that of the right hand side of (4)).

Lemma 1. *If X and Y are conditionally independent variables given Z, $P(X)$, $P(Y)$, $P(Z)$ and $P(X,Y)$ are prior probabilities, and $P(Z)$ is constant, then*

$$P(Z|X,Y) \propto P(Z|X)P(Z|Y). \tag{2}$$

Proof. By using Bayes' theorem and the definition of conditional independence,

$$\begin{aligned} P(Z|X,Y) &= \frac{P(X,Y|Z)\cdot P(Z)}{P(X,Y)} \propto P(X,Y|Z) = P(X|Z)\cdot P(Y|Z) \\ &= \frac{P(Z|X)\cdot P(X)}{P(Z)} \cdot \frac{P(Z|Y)\cdot P(Y)}{P(Z)} \propto P(Z|X)\cdot P(Z|Y). \end{aligned} \tag{3}$$

□

Please note the proof in (3) does have a strong assumption that $P(Z)$ is constant. However this assumption is intrinsically true in our application, as we use $P(Z)$ to predict the existence of bifurcation or branch direction, modelled by a uniform distribution in the absence of prior knowledge.

Defining $\mathbb{I}_t = \{I_t, I_{t-1}, ..., I_1\}$ and $\mathbb{C}_t = \{C_t, C_{t-1}, ..., C_1\}$, and replacing X, Y, Z with I_t, $(\mathbb{I}_{t-1}, \mathbb{C}_{t-1})$, C_t, respectively, we attain Theorem 1.

Theorem 1. *If I_t, $\mathbb{I}_{t-1}$ and $\mathbb{C}_{t-1}$ are conditionally independent given C_t, then*

$$P(C_t|I_t, \mathbb{I}_{t-1}, \mathbb{C}_{t-1}) \propto P(C_t|I_t)\cdot P(C_t|\mathbb{I}_{t-1}, \mathbb{C}_{t-1}), \tag{4}$$

i.e, the inference of candidate C_t could be separated into inference from image features at time step t and the inference from all previous states and their respective image features. Since our goal is to find a feasible algorithm to approximate (1), we must approximate $P(C_t|\mathbb{I}_{t-1}, \mathbb{C}_{t-1})$ instead. This could be achieved by calculating its variational lower bound (ELBO) using Jensen's inequality.

2.3 ELBO Calculation

We first introduce hidden variables $\mathbb{H}_t = \{h_t, h_{t-1}, ..., h_1\}$ to encode non-observable variables (i.e., beyond $\mathbb{I}$ and $\mathbb{C}$ of the tracking process). We denote

the prior distribution of $\mathbb{H}_{t-1}$ by q. Then, since the logarithm function is concave, by Jensen's inequality, $\mathbb{E}(\log(X)) \leq \log(\mathbb{E}(X))$, which we use to arrive at (rewriting $\mathbb{I}_{t-1}, \mathbb{C}_{t-1}$ as Θ for readability):

$$\log P(C_t|\Theta) = \log \int_{\mathbb{H}_{t-1}} P(C_t, \mathbb{H}_{t-1}|\Theta) = \log \int_{\mathbb{H}_{t-1}} P(C_t, \mathbb{H}_{t-1}|\Theta) \frac{q(\mathbb{H}_{t-1})}{q(\mathbb{H}_{t-1})}$$
$$= \log \left(\mathbb{E}_q \left[\frac{P(C_t, \mathbb{H}_{t-1}|\Theta)}{q(\mathbb{H}_{t-1})} \right] \right) \geq \mathbb{E}_q \left[\log \left(\frac{P(C_t, \mathbb{H}_{t-1}|\Theta)}{q(\mathbb{H}_{t-1})} \right) \right] = \mathcal{LB}. \quad (5)$$

Since $\mathbb{E}_q[\log P(C_t, \mathbb{H}_{t-1}|\Theta)] = \mathbb{E}_q[\log(P(C_t|\mathbb{H}_{t-1}, \Theta) \cdot P(\mathbb{H}_{t-1}|\Theta))]$ and $\log(x/y) = \log x - \log y$, the variational lower bound $\mathcal{LB}$ of (1) can be further simplified to:

$$\mathcal{LB} = \mathbb{E}_q[\log(P(C_t|\mathbb{H}_{t-1}, \Theta))] + \mathbb{Q}(\mathbb{H}_{t-1}, \Theta), \quad (6)$$

where $\mathbb{Q}(\mathbb{H}_{t-1}, \Theta) = \mathbb{E}_q[\log(P(\mathbb{H}_{t-1}|\Theta))] - \mathbb{E}_q[\log(q(\mathbb{H}_{t-1}))]$.

2.4 Adopting LSTM

Given their ability to encode sequential data, we adopt LSTM to learn the sequential information within the tracking process. Unlike the memoryless Markovian assumptions, the adoption of the LSTM network naturally encodes all information from previous steps in h_{t-1}, which approximates corresponding terms in (4) and (6) as $P(C_t|I_t) \approx P(C_t|I_t, h_{t-1})$ and $P(C_t|\mathbb{H}_{t-1}, \Theta) \approx P(h_{t-1}|\mathbb{H}_{t-1}, \Theta)$. See the proposed architecture in Fig. 1.

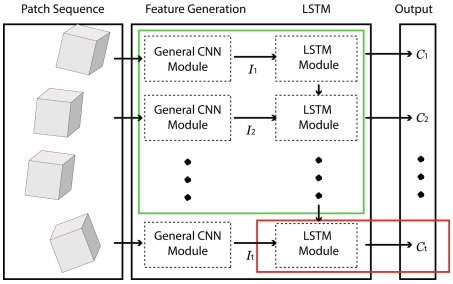

Fig. 1. (color figure) Schematic representation of Tree-LSTM. Left column: A sequence of patches from an image. Green box: predicts $P(h_{t-1}|\mathbb{H}_{t-1}, \Theta)$. Red box: predicts $P(C_t|I_t, h_{t-1})$. Right column: final output. (Color figure online)

2.5 Implementation Details

We use two CNN implementations, TreeNet [21] and DCNN [17], to extract features I_t, which are then fed into the LSTM cell of Sak et al. [11]. We use two fully connected layers (with 1024 and 64 nodes respectively) before the output

layer of the network and use the 1024-dim vector as I_t. The CNN model is first trained then kept fixed during LSTM training. For the LSTM network, a hidden state vector length of 32 is used with a sequence length $L = 10$. Our model uses a separate LSTM for each task as C_t encodes either the presence of bifurcation ($\mathcal{L}_{\text{CLASS}}$) or the tracking direction ($\mathcal{L}_{\text{DIR}}$), with corresponding losses:

$$\mathcal{L}_{\text{CLASS}} = \sum_{t=1}^{L} |c_t - c_t^g| \qquad \mathcal{L}_{\text{DIR}} = -\sum_{t=1}^{L} \langle\, \vec{v}_t, \vec{v}_t^g \rangle, \tag{7}$$

where c is the predicted class ($c = 0$ for single branch, $c = 1$ for bifurcations) and $\vec{v}$ is the predicted direction(s) (1 direction predicted for $c = 1$ and 3 directions predicted for $c = 2$). Superscript g implies ground truth values. We use stochastic gradient descent for optimization with momentum 0.999, exponential decay ratio 0.6, and initial learning rate $1e^{-8}$ for direction prediction and $1e^{-5}$ for bifurcation classification. The implementation is based on TensorFlow v1.12 on Nvidia GTX 1080Ti GPUs.

3 Experiments

Datasets. We use the EXACT[1] challenge public training dataset with 4-fold cross validation. Three major airway branches – trachea, left main bronchi (LMB) and right main bronchi (RMB) are extracted for evaluation.

Competing Methods. For evaluation, we use three state-of-the-art works: Zhao et al. [20] (RF); TreeNet [21]; and DCNN [17] (the last layers of TreeNet and DCNN are modified for classification and regression purposes). To isolate the benefit of adding LSTM, we evaluate the two CNN-based methods, TreeNet and DCNN, without and with LSTM. We change the input patch size to 32 and use the same loss function ([21]) for both TreeNet and DCNN, and remove the middle layer of TreeNet for computational speed up. For evaluation, we measure bifurcation classification accuracy ($\mathcal{ACC}_{CLASS}$) between output class (c) and ground truth class (c^g), and direction prediction accuracy ($\mathcal{ACC}_{DIR}$) between output direction ($\vec{v}$) and ground truth class ($\vec{v}^g$):

$$\mathcal{ACC}_{\text{CLASS}} = 1 - |c_t - c_t^g| \qquad \mathcal{ACC}_{\text{DIR}} = \langle\, \vec{v}_t, \vec{v}_t^g \rangle. \tag{8}$$

Prediction Evaluation Results. From Table 1a, we see that TreeNet+LSTM outperforms TreeNet by 21%, and DCNN+LSTM outperforms DCNN by 17%. RF performs better than TreeNet and DCNN alone on almost all cases, and performs no worse than TreeNet+LSTM on the trachea. This is not surprising as RF was designed for the purpose of bifurcation classification whereas TreeNet and DCNN were both designed for direction prediction. From Table 1b, adding LSTM, boosted TreeNet's performance by 11% and DCNN by 18%, on average.

Assessing LSTM's Ability to Leverage Sequential Data. In Sect. 2, we hypothesized that LSTM is applicable to higher-order Markovian inference due

[1] http://image.diku.dk/exact/.

Table 1. Detection accuracy. †Not applicable as RF does not predict direction.

Branch	Trachea	LMB	RMB
RF [20]	0.74 (±0.49)	0.67 (±0.45)	0.61 (±0.45)
TreeNet [21]	0.35 (±0.49)	0.54 (±0.50)	0.57 (±0.50)
TreeNet+LSTM	0.61 (±0.39)	0.74 (±0.44)	0.75 (±0.44)
DCNN [17]	0.74 (±0.46)	0.61 (±0.50)	0.45 (±0.50)
DCNN+LSTM	0.80 (±0.41)	0.78 (±0.43)	0.72 (±0.45)

(a) Bifurcation classification accuracy (mean± std)

Branch	Trachea	LMB	RMB
RF [20]	n/a†	n/a†	n/a†
TreeNet	0.71 (±0.12)	0.79 (±0.11)	0.80 (±0.08)
TreeNet+LSTM	0.82 (±0.12)	0.92 (±0.06)	0.90 (±0.05)
DCNN [17]	0.76 (±0.15)	0.57 (±0.21)	0.84 (±0.13)
DCNN+LSTM	0.87 (±0.12)	0.93 (±0.05)	0.92 (±0.07)

(b) Direction prediction accuracy (mean± std)

to $P(C_t|\mathbb{H}_{t-1}, \mathbb{I}_{t-1}, \mathbb{C}_{t-1}] \approx P(C_t|h_{t-1}, I_t)$, which suggests the model directly learns information from $\mathbb{H}_{t-1}$ as a whole. To this end, we wish to validate that Tree-LSTM prediction accuracy improves with increased evidence along an L-long sequence of patches. So, we define an evidence support measure within the sequence as $\mathcal{B} = \sum_{i=1}^{L} \beta_i$ where $\beta_i = 1$ indicates the presence of a bifurcation in the ground truth data and 0 otherwise. Now, we bin our data based on $\mathcal{B} \in \{1, 2, \cdots, 10\}$ and measure, for every evidence bin, the average bifurcation classification accuracy at the last or 10^{th} patch, i.e. $P(C_{10}|\mathbb{H}_9, \mathbb{I}_9, \mathbb{C}_9)$. Table 2 records the correlation values between average bifurcation classification accuracy and evidence, which clearly shows how adding LSTM improves the correlation substantially between ∼15% and 67%. Intuitively speaking, an increased correlation value ρ means, by seeing more evidence in the sequence (e.g., bifurcations found in $\mathbb{C}_9$), the accuracy predicting C_{10} would be increased, which is consistent with our initial assumption in (1).

Table 2. Correlation values ρ between average classification accuracy and degree of evidence for different branches. The percentage improvement in ρ when using LSTM (with TreeNet and DCNN) is also reported (calculated as $(\rho_{LSTM} - \rho_{no})/2$, denominator 2 is the max possible change in ρ).

Branch	Trachea				LMB				RMB			
Method	TreeNet		DCNN		TreeNet		DCNN		TreeNet		DCNN	
with LSTM?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
ρ	0.09	0.61	0.12	0.91	0.50	0.81	0.10	0.67	−0.26	0.87	−0.52	0.83
Improvement	25.74%		39.54%		15.98%		28.25%		56.62%		67.24%	

4 Conclusion

To utilize predictions of all previous points along a tracked vessel/airway centerline, we extended the commonly adopted first-order tree branch tracking assumption to a higher-order Markovian process. We estimated the Bayesian variational

lower bound of the proposed formulation and used the CNN+LSTM architecture to optimize tracking. We showed the advantage of using LSTM in tracking real clinical data where the proposed method outperformed the state-of-the-art by at least 11%. The improvement in correlation values between bifurcation classification accuracy and amount of branch sequence evidence is improved by at least 15%.

Acknowledgments. Partial funding for this project is provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors are grateful to the NVIDIA Corporation for donating a Titan X GPU used in this research.

References

1. Blei, D.M., et al.: Variational inference: a review for statisticians. J. Am. Stat. Assoc. **112**(518), 859–877 (2017)
2. Caresio, C., et al.: Quantitative analysis of thyroid tumors vascularity: a comparison between 3-D contrast-enhanced ultrasound and 3-D power doppler on benign and malignant thyroid nodules. Med. Phys. **45**(7), 3173–3184 (2018)
3. Cetin, S., et al.: Vessel tractography using an intensity based tensor model with branch detection. TMI **32**(2), 348–363 (2013)
4. Charbonnier, J.P., et al.: Improving airway segmentation in computed tomography using leak detection with convolutional networks. MedIA **36**, 52–60 (2017)
5. Chen, H., et al.: Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part I. LNCS, vol. 9349, pp. 507–514. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_62
6. Eladawi, N., et al.: Early diabetic retinopathy diagnosis based on local retinal blood vessels analysis in optical coherence tomography angiography (OCTA) images. Med. Phys. **45**(10), 4582–4599 (2018)
7. Lee, S., et al.: Enhanced particle-filtering framework for vessel segmentation and tracking. CMPB **148**, 99–112 (2017)
8. Lesage, D., et al.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. MedIA **13**(6), 819–845 (2009)
9. Lesage, D., et al.: Adaptive particle filtering for coronary artery segmentation from 3D CT angiograms. CVIU **151**, 29–46 (2016)
10. Milletari, F., et al.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision, pp. 565–571 (2016)
11. Sak, H., et al.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: International Speech Communication Association (2014)
12. Selle, D., et al.: Analysis of vasculature for liver surgical planning. TMI **21**(11), 1344–1357 (2002)
13. Selvan, R., Welling, M., Pedersen, J.H., Petersen, J., de Bruijne, M.: Mean field network based graph refinement with application to airway tree extraction. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018, Part II. LNCS, vol. 11071, pp. 750–758. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_83

14. Stollenga, M.F., et al.: Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: NeurIPS, pp. 2998–3006 (2015)
15. Wang, X., et al.: Statistical tracking of tree-like tubular structures with efficient branching detection in 3D medical image data. Phys. Med. Biol. **57**(16), 5325 (2012)
16. WHO: Global health estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016. World Health Organization, Geneva (2018)
17. Wolterink, J.M., et al.: Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier. MIA **51**, 46–60 (2019)
18. Wu, A., et al.: Deep vessel tracking: a generalized probabilistic approach via deep learning. In: ISBI, pp. 1363–1367 (2016)
19. Wu, D., et al.: Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs. IJCARS **14**(2), 271–280 (2019)
20. Zhao, M., Hamarneh, G.: Bifurcation detection in 3D vascular images using novel features and random forest. In: ISBI, pp. 421–424 (2014)
21. Zhao, M., Hamarneh, G.: TreeNet: multi-loss deep learning network to predict branch direction for extracting 3D anatomical trees. In: Stoyanov, D., et al. (eds.) DLMIA 2018/ML-CDS 2018. LNCS, vol. 11045, pp. 47–55. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_6
22. Zhao, M., et al.: Leveraging tree statistics for extracting anatomical trees from 3D medical images. In: Computer and Robot Vision, pp. 131–138 (2017)