Vestibular schwannoma growth prediction from longitudinal MRI by time-conditioned neural fields

Yunjie Chen[1], Jelmer M. Wolterink[2], Olaf M. Neve[3], Stephan R. Romeijn[1], Berit M. Verbist[1], Erik F. Hensen[3], Qian Tao[4], and Marius Staring[1]

[1] Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands
[2] Department of Applied Mathematics, Technical Medical Center, University of Twente, Enschede, the Netherlands
[3] Department of Otorhinolaryngology and Head & Neck Surgery, Leiden University Medical Center, Leiden, the Netherlands
[4] Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands

Abstract. Vestibular schwannomas (VS) are benign tumors that are generally managed by active surveillance with MRI examination. To further assist clinical decision-making and avoid overtreatment, an accurate prediction of tumor growth based on longitudinal imaging is highly desirable. In this paper, we introduce DeepGrowth, a deep learning method that incorporates neural fields and recurrent neural networks for prospective tumor growth prediction. In the proposed model, each tumor is represented as a signed distance function (SDF) conditioned on a low-dimensional latent code. Unlike previous studies, we predict the latent codes of the future tumor and generate the tumor shapes from it using a multilayer perceptron (MLP). To deal with irregular time intervals, we introduce a time-conditioned recurrent module based on a ConvLSTM and a novel temporal encoding strategy, which enables the proposed model to output varying tumor shapes over time. The experiments on an in-house longitudinal VS dataset showed that the proposed model significantly improved the performance ($\geq$ 1.6% Dice score and $\geq$ 0.20 mm 95% Hausdorff distance), in particular for top 20% tumors that grow or shrink the most ($\geq$ 4.6% Dice score and $\geq$ 0.73 mm 95% Hausdorff distance). Our code is available at `https://github.com/cyjdswx/DeepGrowth`.

Keywords: Tumor growth prediction · neural fields · signed distance function · ConvLSTM

1 Introduction

Vestibular schwannomas (VS) are intracranial tumors arising from the balance and hearing nerves, of which approximately 40% are progressive and ultimately become life-threatening [2]. In current clinical practice, VS are generally managed by active surveillance with MRI examination and manual tumor diameter

measurements [16,7]. Once significant growth (> 2mm difference between two consecutive MRI scans) is detected, the tumors are treated with either radiotherapy or surgery [16,12]. However, research shows that although 80% of VS shows certain growth during observation, only half of them are truly progressive, indicating that many patients suffer from overtreatment [12]. On the other hand, late treatment of a larger tumor can also damage the prognosis after treatment, which requires a timely clinical decision [9]. Hence, to avoid overtreatment and sequelae associated with the treatment of large tumors, early and precise prediction of tumor growth based on longitudinal imaging is highly desirable.

Early studies on image-driven tumor growth prediction typically utilized biomechanical models, such as reaction-diffusion equations, to derive physiological parameters related to tumor progression [11,13]. However, most of the models require specific imaging modalities that are unfortunately not available in clinical routine for VS. Recently, deep learning models have shown promising performance for longitudinal tumor shape modeling. Inspired by the neural process framework, Petersen et al. proposed to learn a distribution of possible future shapes of glioma using a self-attention mechanism [19]. Instead of generative models, Zhang et al. [26] applied a spatio-temporal ConvLSTM for pancreatic tumor growth modeling. Elazab et al. [4] proposed a 3D GP-GAN that utilizes multiple stacked generative adversarial networks to predict glioma growth. Subsequently, Wang et al. [23] applied a Transformer model to longitudinal CT for 4D lung cancer tumor modeling. Although promising results were demonstrated, most models assume unified time intervals between consecutive scans, which is unfortunately uncommon in the clinic. Moreover, future prediction in high-dimensional image space has large memory requirements, which could limit application [26], and may also introduce spatial redundancy that potentially damage performance [10].

One way to tackle this problem is compressing the input into a low-dimensional latent code utilizing an autoencoder and performing predictions in the latent space [5,20]. In line with this, we propose to perform future tumor prediction with neural field representations [17,24]. The key idea of neural fields is to represent a function describing an image or object in the spatial or spatio-temporal domain as a neural network with trainable weights [25]. The neural network can be conditioned on latent codes to represent a distribution of objects. Recently, Agro et al. [1] successfully predicted future occupancy maps using spatio-temporal neural fields. However, the method requires sufficient frames over time, while longitudinal medical imaging usually contains only few measurements.

To address these limitations, we propose DeepGrowth, a model that incorporates neural fields and recurrent neural networks for tumor growth prediction. Specifically, DeepGrowth encodes prior images and tumor masks into latent codes and parameterizes the tumor as a signed distance function (SDF). To deal with irregular time intervals between scans, we apply a time-conditioned recurrent module to predict the latent code, on which the reconstruction of the future tumor shape is conditioned. The main contributions of this work are: (1) In contrast to previous studies that perform tumor prediction directly in image

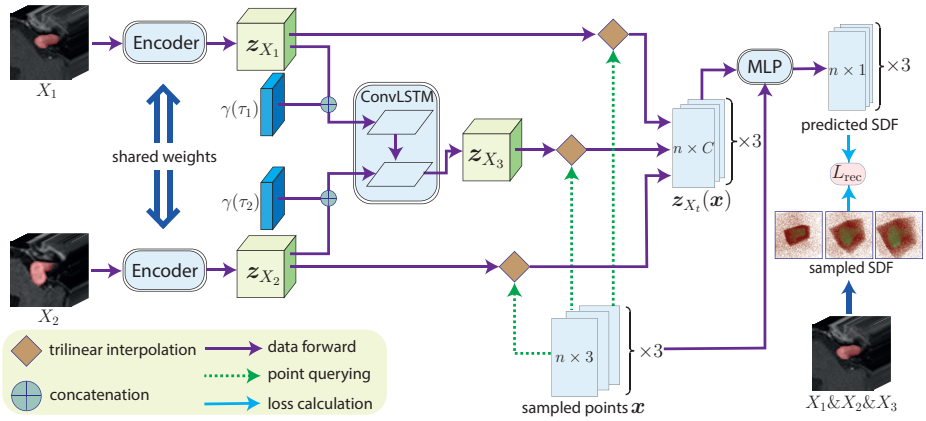

Fig. 1. The overall architecture of DeepGrowth ($N = 3$). Prior scans are encoded into latent codes, which are concatenated with temporal encoding. The MLP reconstructs the future tumor as an SDF conditioned on the output of the ConvLSTM. L_{rec} is calculated between the predictions and SDF sampled from all three tumor masks.

space, for the first time, we represent tumor shapes as neural fields and predict the future based on learned latent codes. (2) We introduce a time-conditioned recurrent module with a novel temporal encoding strategy that enables us to query tumor shapes at specific time intervals. (3) The proposed model was evaluated on an in-house longitudinal VS dataset, showing a significantly better performance than other models, in particular for relatively fast growing tumors.

2 Methods

Given a patient with N longitudinal images with corresponding segmentations, denoted as $X_t = \{I_t, M_t, D_t\}$, $t = 1, 2, ..., N$, where I_t is the image at time t, M_t is the corresponding tumor mask and D_t the normalized scan date ranging from 0 to 1, our goal is to find a function Φ:

$$\Phi : \{X_1, X_2, ..., X_{N-1}, D_N\} \rightarrow M_N. \tag{1}$$

Instead of performing prediction directly in image space, we encode X_t into a low-dimensional latent code and predict future by a time-conditioned recurrent module. See Fig. 1 for an overview of the model architecture when $N = 3$.

2.1 3D tumor shape as signed distance function

In the proposed model, each tumor is encoded into a low-dimensional latent code, which can be used to condition a neural field for tumor shape reconstruction. More specifically, we concatenate $I_t \in \mathbb{R}^{D \times H \times W}$ and $M_t \in \mathbb{R}^{D \times H \times W}$, and encode them via a convolution-based encoder with a downsampling factor s. The

latent code is denoted as $\boldsymbol{z}_{X_t} \in \mathbb{R}^{C\times d\times h\times w}$, where $d = D/s$, $h = H/s$, $w = W/s$, and C is the feature dimension. Unlike studies that use a single vector to represent the entire object [17,24], our latent code contains $d\times h\times w$ vectors, encoding the local information in a more expressive representation [18].

To reconstruct tumor shapes from the latent code, we represent each tumor shape using an SDF [17]. For clarity, we use c_t to denote the tumor contour of M_t, which is a closed 2D manifold embedded in 3D space. Hence, for each M_t, the SDF of the tumor can be defined as:

$$\mathrm{SDF}_{M_t}(\boldsymbol{x}) = \begin{cases} \min_{u\in c_t}\|x-u\|_2, & \text{if } x \text{ inside } c_t \\ 0, & \text{if } x \text{ belonging to } c_t \\ -\min_{u\in c_t}\|x-u\|_2, & \text{if } x \text{ outside } c_t \end{cases} \tag{2}$$

where $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$. Different from voxelized or meshed representations, the SDF and therefore $\boldsymbol{x}$ is defined over the entire space. In the proposed model, we approximate the SDF by a multilayer perceptron (MLP) f. Similar to [18,3], we apply a local conditioning strategy, in which $\mathrm{SDF}_{M_t}(\boldsymbol{x})$ is conditioned on the local latent code $\boldsymbol{z}_{X_t}(\boldsymbol{x})$. $\boldsymbol{z}_{X_t}(\boldsymbol{x})$ is a vector of size C queried from the entire latent code $\boldsymbol{z}_{X_t}$ using trilinear interpolation [18]. For each point $\boldsymbol{x}$, we concatenate the coordinates $\boldsymbol{x}$ with $\boldsymbol{z}_{X_t}(\boldsymbol{x})$ as the input of the MLP, which can then be denoted as:

$$\mathrm{SDF}_{M_t}(\boldsymbol{x}) \approx f_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}_{X_t}(\boldsymbol{x})), \tag{3}$$

where $\boldsymbol{\theta}$ are the parameters of the MLP. Hence, each tumor contour is described by the zero-level set of the SDF estimated by the MLP.

2.2 Time-conditioned recurrent module

Earlier studies on tumor prediction usually assume unified time intervals between consecutive scans [23] and predict a more distant future with additional recurrent steps [4]. However, patients frequently receive follow-up scans with irregular time intervals. We therefore introduce a time-conditioned recurrent module, which consists of temporal encoding and a small 3D ConvLSTM, to predict future tumor shapes. The 3D ConvLSTM takes the input of $\boldsymbol{z}_{X_t}$, $t = 1, 2, ..., N-1$ with the study dates D_t, $t = 1, 2, ..., N$ and predicts $\boldsymbol{z}_{X_N}$. To better encode the temporal information, we apply sinusoidal functions to the time intervals similar to positional encoding [14], which we call temporal encoding. Given the time interval $\tau_i = D_{i+1} - D_i$, where $i = 1, 2, ..., N-1$, the temporal encoding is expressed as follows:

$$\gamma(\tau_i) = [\sin(2^0\pi\tau_i), \cos(2^0\pi\tau_i), \ldots, \sin(2^{l-1}\pi\tau_i), \cos(2^{l-1}\pi\tau_i)], \tag{4}$$

where l is the order of the temporal encoding. To avoid overfitting, a dropout layer is added to the temporal encoding. We then concatenate $\gamma(\tau_i)$ to all vectors of $\boldsymbol{z}_{X_i}$ as the input of the ConvLSTM. Given the output $\boldsymbol{z}_{X_n}$ of the ConvLSTM, we can obtain SDF_{M_n} of the future tumor via Eq. (3).

Table 1. Quantitative comparison results on a vestibular schwannoma dataset using 5-fold cross-validation. The mean and standard deviation of Dice, 95% HD, and RVD are reported. The highest values per column are indicated in bold; † indicates a significant difference ($p < .05$) compared to the proposed method.

Method	#params	Dice ↑	95% HD (mm) ↓	RVD ↓
Stable tumor		$0.766 \pm 0.143^{\dagger}$	$1.95 \pm 2.55^{\dagger}$	$\mathbf{0.490 \pm 2.99}$
ST-ConvLSTM [26]	0.6 M	$0.758 \pm 0.141^{\dagger}$	$2.07 \pm 2.65^{\dagger}$	$0.611 \pm 3.62^{\dagger}$
3D ConvLSTM [21]	4.4 M	$0.784 \pm 0.139^{\dagger}$	$1.91 \pm 2.50^{\dagger}$	$0.564 \pm 3.47^{\dagger}$
DeepGrowth (proposed)	4.9 M	$\mathbf{0.800 \pm 0.115}$	$\mathbf{1.71 \pm 2.23}$	0.521 ± 3.48

Table 2. Quantitative comparison results of the top 20% fastest growing or shrinking VS using 5-fold cross-validation. The mean and standard deviation of Dice, 95% HD and RVD are reported. The highest values per column are indicated in bold; † indicates a significant difference ($p < .05$) compared to the proposed method.

Method	#params	Dice ↑	95% HD (mm) ↓	RVD ↓
Stable tumor		$0.697 \pm 0.182^{\dagger}$	$4.18 \pm 3.30^{\dagger}$	$0.413 \pm 0.323^{\dagger}$
ST-ConvLSTM [26]	0.6 M	$0.707 \pm 0.188^{\dagger}$	$4.28 \pm 3.42^{\dagger}$	0.398 ± 0.470
3D ConvLSTM [21]	4.4 M	$0.736 \pm 0.176^{\dagger}$	$3.87 \pm 3.22^{\dagger}$	0.366 ± 0.332
DeepGrowth (proposed)	4.9 M	$\mathbf{0.782 \pm 0.120}$	$\mathbf{3.14 \pm 2.22}$	$\mathbf{0.321 \pm 0.315}$

2.3 End-to-end network training

All components are optimized together end-to-end. For training, we randomly sample n points from each tumor volume with 80% of the points sampled near the contour and the rest sampled from the entire space. We apply an ℓ_1 reconstruction loss that maximizes the similarity between the real SDF and the estimations, as suggested in [17], for all N tumors:

$$L_{\text{rec}} = \frac{1}{nN} \sum_{t=1}^{N} \sum_{i=1}^{n} \| f_\theta(\boldsymbol{x}_i, \boldsymbol{z}_{X_t}(\boldsymbol{x}_i)) - \text{SDF}_{M_t}(\boldsymbol{x}_i) \|_1, \tag{5}$$

where $\boldsymbol{x}_i$ are the sampled points. To stabilize the training, we apply the ℓ_2 norm to the latent codes as the regularization: $L_{\text{reg}} = \frac{1}{N} \sum_{t=1}^{N} \|\boldsymbol{z}_{X_t}\|_2$. As a result, the overall loss function of the proposed model is $L = \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{reg}} L_{\text{reg}}$, where λ_{rec} and λ_{reg} are the weights of each loss function.

3 Experiments

3.1 Dataset

To evaluate the proposed method, 131 vestibular schwannoma patients were selected from our previous study [15,16]. Each patient in the dataset has three consecutive contrast enhanced T1 (T1ce) scans, separated by 87 to 2157 days. The spatial resolution of the T1ce ranges from $0.254 \times 0.254 \times 0.81$ mm to

$1.17 \times 1.17 \times 1.20$ mm and the in-plane resolution ranges from 256×192 to 640×520. Of all scans the tumor masks were generated using a segmentation model developed in our previous study based on nnUNet [6,15]. We aligned all scans of each patient by rigid registration using elastix [8]. All images were then resampled to an isotropic resolution of $0.58 \times 0.58 \times 0.58$ mm. To avoid the influence of background, $64 \times 64 \times 64$ cropping was performed around the centroid of the tumor. The intensities of T1ce were normalized to $[-1, 1]$ and D_t was normalized to $[0, 1]$ within the original range of 0 to 10 years.

3.2 Implementation details

We adapted a 3D U-Net from [18] as the encoder with two extra convolutional blocks for downsampling. The ConvLSTM in the time-conditioned recurrent module consists of three 32-channel layers and the MLP contains five 64-channel layers with sine as the activation function [22]. Due to the limited dataset size and diversity in tumor growth trends, we perform five-fold cross-validation and report the average results of the five folds to avoid bias. We set the downsampling factor $s = 4$ and temporal encoding order $l = 6$ for best performance (see Section 3.4). Little difference was observed between different loss weights, which were set to $\lambda_{\text{rec}} = 1.0$ and $\lambda_{\text{reg}} = 0.1$. The model was optimized using Adam with an initial learning rate of $1e - 4$. During inference, the tumor masks were generated from the zero level-set of the predicted SDF and evaluated using the Dice, 95% Hausdorff distance (95% HD), and relative volume difference (RVD). All experiments were conducted using Python 3.10 and PyTorch 1.12.1 on a machine equipped with Nvidia Quadro RTX 6000 and Nvidia Tesla V100 GPUs.

3.3 Future tumor shape prediction

We first evaluate the proposed model by predicting the third future tumor shape from the first two scans and time intervals. We compare our model against three baselines. The first baseline assumes the tumor remains stable after the second scan, which is reasonable due to the slow growth of VS, so we simply take the tumor mask from the second time point as a prediction, which we call "stable tumor" in the experiments. The second and third baselines are two ConvLSTM-based models: ST-ConvLSTM [26] and 3D ConvLSTM [21]. ST-ConVLSTM is a smaller 2D model where we use the same architecture as described in the original paper. 3D ConvLSTM, which contains a comparable number of parameters to the proposed model, consists of three layers with (64, 128, 64) channels respectively. Unlike the original papers that use the ℓ_1 loss to train the model to generate binary maps, we used a weighted sum of Dice loss and binary cross-entropy loss, which performed better on our data, to train the baselines. Wilxocon signed rank tests were performed between the proposed model and each baseline.

The quantitative results are listed in Table 1 with visualizations in Fig. 2. The proposed model performed significantly better than all baselines in terms of Dice and 95%HD. The proposed model obtained a higher RVD due to an extreme outlier (see last row in Fig. 2). When removing this outlier, the proposed method

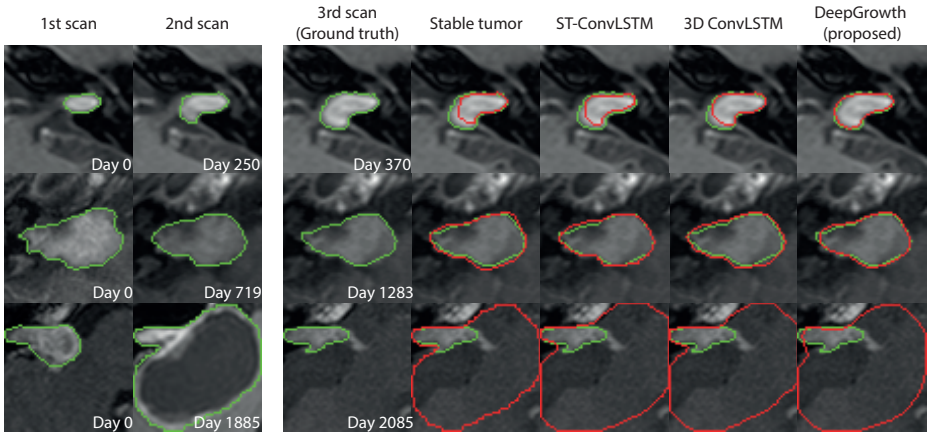

Fig. 2. Example results of the different models. The first two columns are the input of the models, followed by the ground truth in the third column, and model predictions in subsequent ones. Predicted tumors are depicted in red and the ground truths in green. The dates are the study dates. The last row depicts a tumor that suddenly shrank after the second scan, which was difficult to predict for all models.

Table 3. Quantitative results of top 20% growers when varying temporal encoding.

methods	order	Dice ↑	95% HD (mm) ↓	RVD ↓
w/o time		0.765 ± 0.143	3.37 ± 2.49	0.341 ± 0.314
with time		0.774 ± 0.126	3.23 ± 2.50	0.316 ± 0.278
time + temporal encoding	$l = 4$	0.773 ± 0.144	3.33 ± 2.53	0.316 ± 0.306
	$l = 6$	$\mathbf{0.782 \pm 0.119}$	$\mathbf{3.14 \pm 2.22}$	0.321 ± 0.315
	$l = 8$	0.773 ± 0.142	3.23 ± 2.39	$\mathbf{0.315 \pm 0.363}$

obtained an RVD of 0.218 ± 0.248, which outperformed all baselines (0.229 ± 0.230, 0.296 ± 0.304 and 0.261 ± 0.266, respectively).

We noticed that the stable tumor method obtained comparable quantitative scores, which is on par with the fact that many VS grow slowly or even remain stable. Focusing on the top 20% of tumors that grow (or shrink) the most, see Table 2, we observe a larger gap between the proposed model and the baselines, indicating the improved capability of modeling tumor growth.

3.4 Ablation study

To examine the impact of temporal encoding, we trained two additional models: one without time factors at all, and one using time intervals τ_i directly as suggested in [26]. We also compare the models using different orders l for temporal encoding. The results of the top 20% growers are shown in Table 3. Direct use of τ_i barely improved results, while temporal encoding improved the results for all metrics. Best results were obtained for $l = 6$, with higher l leading to overfitting.

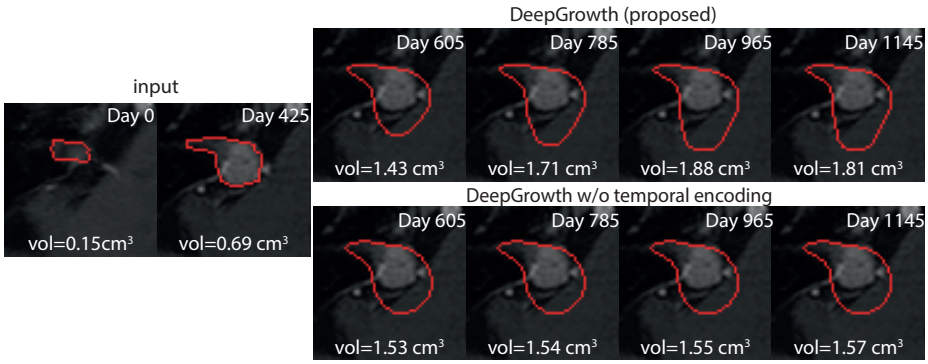

Fig. 3. Querying the proposed model at different time points (increments of 180 days). We overlaid predictions on I_2 for visualization in columns 3-6. The proposed model can output varied tumor shapes given different time intervals, while the model without temporal encoding outputs almost the same results regardless of the time intervals.

Table 4. Quantitative results of DeepGrowth using different downsampling factors.

downsampling factor s	Dice ↑	95% HD (mm) ↓	RVD ↓
s=1	0.788 ± 0.127	1.87 ± 2.39	0.544 ± 3.68
s=2	0.796 ± 0.122	1.78 ± 2.40	0.577 ± 4.18
s=4	$\mathbf{0.800 \pm 0.115}$	$\mathbf{1.71 \pm 2.23}$	$\mathbf{0.52 \pm 3.48}$
s=8	0.784 ± 0.125	1.85 ± 2.35	0.598 ± 4.10

As our model allows us to query arbitrary future time points, we show predictions given different τ_2 (with a step of 180 days) in Fig. 3. We can see that the model using τ_i without temporal encoding outputs almost the same results regardless of the time intervals. On the contrary, by using temporal encoding, the proposed model can output varied tumor shapes given different time intervals, from which we can view how tumors grow over time.

Models using high-resolution feature maps were more difficult to train, while lower-resolution feature maps potentially degraded performance due to lowered expressive capability [5,20]. We, therefore, varied the downsampling factors s, see Table 4, and concluded that $s = 4$ resulted in the best performance.

4 Discussion and Conclusion

In this paper, we proposed DeepGrowth, a deep learning model that incorporates neural fields and recurrent neural networks for tumor growth prediction. Unlike conventional models that predict image or segmentation masks directly in the image space [26,4], we encode tumors into a latent space and predict future latent codes. The future tumor shape is reconstructed as the zero-level set of an SDF conditioned on the predicted latent code via an MLP. A comparison on a longitudinal VS dataset showed improved performance of the proposed model,

in particular for more challenging growing or shrinking tumors. We applied temporal encoding to the study intervals, which helped the model to encode time information and output varied tumor shapes given different time intervals. However, it remains to be investigated if tumor growth derived from our predictions can be used to aid clinical decision making. In conclusion, we showed that neural fields hold great promise for information compression, which can facilitate longitudinal tumor modeling.

Acknowledgments. This study was supported by the China Scholarship Council (grant 202008130140), and by an unrestricted grant of Stichting Hanarth Fonds, The Netherlands (project MLSCHWAN).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agro, B., Sykora, Q., Casas, S., Urtasun, R.: Implicit occupancy flow fields for perception and prediction in self-driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1379–1388 (2023)
2. Carlson, M.L., Link, M.J.: Vestibular schwannomas. New England Journal of Medicine **384**(14), 1335–1348 (2021)
3. Chen, Y., Staring, M., Neve, O.M., Romeijn, S.R., Hensen, E.F., Verbist, B.M., Wolterink, J.M., Tao, Q.: CoNeS: Conditional neural fields with shift modulation for multi-sequence MRI translation. Machine Learning for Biomedical Imaging **2**, 657–685 (2024)
4. Elazab, A., Wang, C., Gardezi, S.J.S., Bai, H., Hu, Q., Wang, T., Chang, C., Lei, B.: GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR images. Neural Networks **132**, 321–332 (2020)
5. Hu, T., Chen, F., Wang, H., Li, J., Wang, W., Sun, J., Li, Z.: Complexity matters: Rethinking the latent space for generative modeling. In: Advances in Neural Information Processing Systems. vol. 36 (2024)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
7. Kanzaki, J., Tos, M., Sanna, M., Moffat, D.A.: New and modified reporting systems from the consensus meeting on systems for reporting results in vestibular schwannoma. Otology & neurotology **24**(4), 642–649 (2003)
8. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2009)
9. Li, D., Tsimpas, A., Germanwala, A.V.: Analysis of vestibular schwannoma size: A literature review on consistency with measurement techniques. Clinical neurology and neurosurgery **138**, 72–77 (2015)
10. Liu, B., Chen, Y., Liu, S., Kim, H.S.: Deep learning in latent space for video prediction and compression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 701–710 (2021)

11. Liu, Y., Sadowski, S.M., Weisbrod, A.B., Kebebew, E., Summers, R.M., Yao, J.: Patient specific tumor growth prediction using multimodal images. Medical image analysis **18**(3), 555–566 (2014)
12. Marinelli, J.P., Link, M.J., Carlson, M.L.: Size threshold surveillance-a revised approach to wait-and-scan for vestibular schwannoma. JAMA Otolaryngology–Head & Neck Surgery **149**(8), 657–658 (2023)
13. Meghdadi, N., Soltani, M., Niroomand-Oscuii, H., Yamani, N.: Personalized image-based tumor growth prediction in a convection–diffusion–reaction model. Acta Neurologica Belgica **120**, 49–57 (2020)
14. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
15. Neve, O.M., Chen, Y., Tao, Q., Romeijn, S.R., de Boer, N.P., Grootjans, W., Kruit, M.C., Lelieveldt, B.P., Jansen, J.C., Hensen, E.F., et al.: Fully automated 3D vestibular schwannoma segmentation with and without Gadolinium-based contrast material: a multicenter, multivendor study. Radiology: Artificial Intelligence **4**(4), e210300 (2022)
16. Neve, O.M., Romeijn, S.R., Chen, Y., Nagtegaal, L., Grootjans, W., Jansen, J.C., Staring, M., Verbist, B.M., Hensen, E.F.: Automated 2-Dimensional measurement of vestibular schwannoma: Validity and accuracy of an artificial intelligence algorithm. Otolaryngology–Head and Neck Surgery **169**(6), 1582–1589 (2023)
17. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 165–174 (2019)
18. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional occupancy networks. In: Proceedings of the European Conference on Computer Vision. pp. 523–540. Springer (2020)
19. Petersen, J., Isensee, F., Köhler, G., Jäger, P.F., Zimmerer, D., Neuberger, U., Wick, W., Debus, J., Heiland, S., Bendszus, M., et al.: Continuous-time deep glioma growth models. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 83–92. Springer (2021)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
21. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems. vol. 28 (2015)
22. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: Advances in Neural Information Processing Systems. vol. 33, pp. 7462–7473 (2020)
23. Wang, H., Xiao, N., Zhang, J., Yang, W., Ma, Y., Suo, Y., Zhao, J., Qiang, Y., Lian, J., Yang, Q.: Static-dynamic coordinated transformer for tumor longitudinal growth prediction. Computers in Biology and Medicine **148**, 105922 (2022)
24. Wiesner, D., Suk, J., Dummer, S., Nečasová, T., Ulman, V., Svoboda, D., Wolterink, J.M.: Generative modeling of living cells with SO(3)-equivariant implicit neural representations. Medical image analysis **91**, 102991 (2024)
25. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond. In: Computer Graphics Forum. vol. 41, pp. 641–676. Wiley Online Library (2022)

26. Zhang, L., Lu, L., Wang, X., Zhu, R.M., Bagheri, M., Summers, R.M., Yao, J.: Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. IEEE Transactions on Medical Imaging **39**(4), 1114–1126 (2019)