Learning Tumor Growth via Follow-Up Volume Prediction for Lung Nodules

Yamin Li[1,2,3], Jiancheng Yang[1,2,3], Yi Xu[1,2,3(✉)], Jingwei Xu[1,2,3], Xiaodan Ye[1,4], Guangyu Tao[1,4], Xueqian Xie[1,5], and Guixue Liu[1,5]

[1] Shanghai Jiao Tong University, Shanghai, China
xuyi@sjtu.edu.cn
[2] Shanghai Institute for Advanced Communication and Data Science, Shanghai, China
[3] MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China
[4] Shanghai Chest Hospital, Shanghai, China
[5] Shanghai General Hospital, Shanghai, China

Abstract. Follow-up serves an important role in the management of pulmonary nodules for lung cancer. Imaging diagnostic guidelines with expert consensus have been made to help radiologists make clinical decision for each patient. However, tumor growth is such a complicated process that it is difficult to stratify high-risk nodules from low-risk ones based on morphologic characteristics. On the other hand, recent deep learning studies using convolutional neural networks (CNNs) to predict the malignancy score of nodules, only provides clinicians with black-box predictions. To this end, we propose a unified framework, named Nodule Follow-Up Prediction Network (*NoFoNet*), which predicts the growth of pulmonary nodules with high-quality visual appearances and accurate quantitative results, given any time interval from baseline observations. It is achieved by predicting future displacement field of each voxel with a WarpNet. A TextureNet is further developed to refine textural details of WarpNet outputs. We also introduce techniques including Temporal Encoding Module and Warp Segmentation Loss to encourage time-aware and shape-aware representation learning. We build an in-house follow-up dataset from two medical centers to validate the effectiveness of the proposed method. *NoFoNet* significantly outperforms direct prediction by a U-Net in terms of visual quality; more importantly, it demonstrates accurate differentiating performance between high- and low-risk nodules. Our promising results suggest the potentials in computer aided intervention for lung nodule management.

Keywords: Lung nodule · Follow-up · Tumor growth prediction

Y. Li and J. Yang—These authors have contributed equally.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59725-2_49) contains supplementary material, which is available to authorized users.

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12266, pp. 508–517, 2020.
https://doi.org/10.1007/978-3-030-59725-2_49

1 Introduction

Pulmonary nodule management strategy influences the cost-effectiveness of a lung cancer screening program [3]. It remains difficult to differentiate high-risk nodules from low-risk ones based on morphologic characteristics [13]. In order to help radiologists and clinicians to make precise clinical decision for each patient, researchers have made several categorical management recommendation and scoring systems according to morphology, diameters or volume in recent years, *e.g.*, NCCN [16], Fleischner [9], Lung-RADS [12]. However, tumor growth is such a complicated progress that more advanced strategies are worth exploring to facilitate precision medicine. Emerging deep learning technology suggests a potential alternative to develop end-to-end lung nodule management system in a data-driven fashion. Although numerous studies have explored end-to-end approaches to predict malignancy scores [5,17,20] or categories [19,22,23] of lung nodules, while only a few studies [1,4] address the lung nodule follow-up problem. Nevertheless, these studies only provide black-box predictions without intuitive explanations. There is also study [11] on predicting tumor growth with a model-free appearance modeling approach using a probabilistic U-Net [14], however it could not provide any quantitative assessment on the risk of tumors.

In this study, we aim at a unified approach to predict growth of lung nodules, with both high-quality visual appearances and accurate quantitative results. The core of our approach is based on a WarpNet, predicting displacement field $\mathbf{u}$ (or motion [6]) on a future volume from a baseline volume. With the field $\mathbf{u}$, we could obtain not only the predicted **visual appearance** of the future volume by warpping the baseline, but also the feature segmentation mask from the baseline mask, which could be used for **quantitative assessment** of tumor growth. This approach is inspired from VoxelMorph [14], where the displacement field for registration is conditional on both the baseline and future volumes; instead, our predictive displacement field is conditional only on the baseline volume and could be dynamically estimated. Moreover, a TextureNet is designed to refine textural details of the outputs from WarpNet. We introduce techniques including Temporal Encoding Module and Warp Segmentation Loss to encourage time-aware and shape-aware representation learning. The whole network, named Nodule Follow-Up Prediction Network (*NoFoNet*), establishes a unified framework to produce both high-quality visual appearances and accurate quantitative assessment for lung nodule follow-up. Our in-house follow-up dataset from two medical centers validates the effectiveness of *NoFoNet*.

2 Materials and Methods

2.1 Task Formalization and Dataset

We aim at a unified framework to predict future volume of a lung nodule, given any time interval and a baseline volume. An in-house dataset is collected, containing 622 LDCT scans from 246 patients (114 males and 132 females) with a total of 315 long-standing pulmonary nodules. Each patient has at least two time

points of thin layer LDCT (slice thickness ≤ 1.25 mm), with the time interval of 30–1351, 136 days (min-max, median). We select nodules at every two time points as a sample (for example if a nodule has 3 follow-up scans at time points $t_1t_2t_3$, we choose time points $t_1\&t_2$, $t_1\&t_3$, $t_2\&t_3$ as 3 samples), resulting in 731 pairs. The age of the patients at first examination is 23–97, 62 years. The segmentation VOI of each selected nodule (diameter 3 mm to 30 mm) is delineated by an expert radiologist and checked by another.

We pre-process the data as follows [20]: CT scans are resampled isotropically into 1 mm × 1 mm × 1 mm. The voxel intensity is normalized to $[-1, 1)$ from the Hounsfield unit (HU), using the mapping function $I = \lfloor \frac{I_{HU}+1024}{400+1024} \times 255 \rfloor / 128 - 1$. Each data sample is a cubic volume image with the size of $48 \times 48 \times 48$, which covers the size of all nodules in our study.

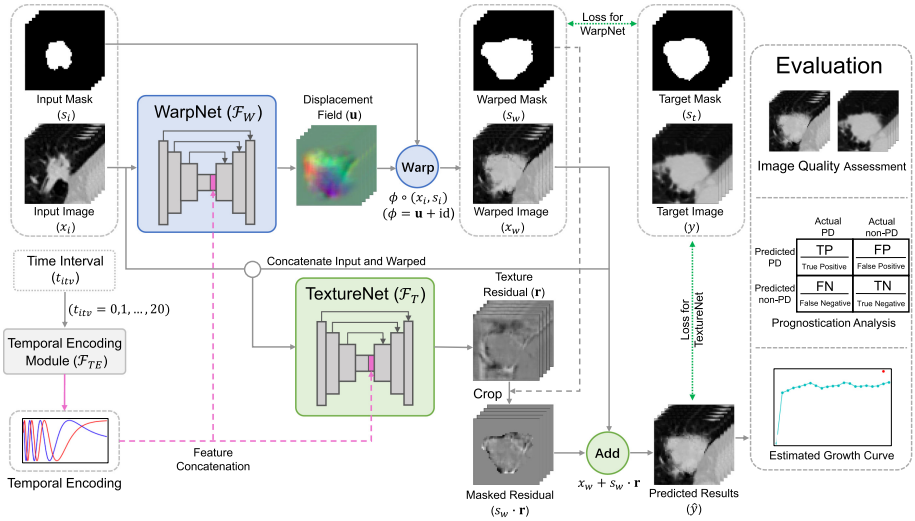

Fig. 1. Overview of the proposed *NoFoNet* architecture. *NoFoNet* consists of a WarpNet and a TextureNet, each with a Temporal Encoding Module (TEM), which encodes follow-up time interval into the nodule representation. WarpNet and TextureNet are two 3D CNNs based on U-Net [14], modeling the spatial and texture transformation for nodule growth respectively. During evaluation, we perform the image quality assessment, prognostication analysis and growth estimation. Note PD means progressive disease, i.e., significant growth of nodule in our study. The red point in the growth curve represents actual volume of the nodule in the future time. (Color figure online)

2.2 *NoFoNet*: Nodule Follow-Up Prediction Network

To model the growth of nodules, we develop a Nodule Follow-Up Prediction Network (*NoFoNet*, see Fig. 1) consisting of a WarpNet $\mathcal{F}_W$ and a TextureNet

$\mathcal{F}_T$ for spatial and texture (intensity) transformations [21] respectively, where an integrated temporal encoding module (TEM) $\mathcal{F}_{TE}$ is addressed to encode different follow-up time interval information into the lesion representation. As we will show later, the WarpNet and TextureNet are able to model the shape and texture variation of nodule growth well.

Given a pair of follow-up input and target images[1] with time interval t_{itv}, each of which has corresponding nodule segmentation map $\{x_i, s_i\}$ and $\{y, s_t\}$, the WarpNet $\mathcal{F}_W$ with parameter θ_w first predicts a smooth voxel-wise displacement field $\mathbf{u} = \mathcal{F}_W(x_i, \mathcal{F}_{TE}(t_{itv}); \theta_w)$ for spatial transformation. Following the registration literature [2], we have the warp function $\phi = \mathbf{u} + \mathrm{id}$, where id is identity function. We apply the warp function ϕ to x_i to get the warped image x_w, and denote this as $x_w = \phi \circ x_i$. Similarly, the warped segmentation map $s_w = \phi \circ s_i$. The TextureNet $\mathcal{F}_T$ with parameter θ_t takes the concatenation of x_i and x_w as inputs and generates a voxel-wise residual $\mathbf{r} = \mathcal{F}_T(x_i, x_w, \mathcal{F}_{TE}(t_{itv}); \theta_t)$. Then we get the results $\hat{y} = x_w + s_w \cdot \mathbf{r}$, where $s_w \cdot \mathbf{r}$ denotes the residual cropped by warped segmentation. The overall formulation of our *NoFoNet* is as follows:

$$\begin{aligned} &\mathbf{u} = \mathcal{F}_W(x_i, \mathcal{F}_{TE}(t_{itv}); \theta_w), \quad \phi = \mathbf{u} + \mathrm{id}, \quad x_w = \phi \circ x_i, \quad s_w = \phi \circ s_i; \\ &\mathbf{r} = \mathcal{F}_T(x_i, x_w, \mathcal{F}_{TE}(t_{itv}); \theta_t), \quad \hat{y} = x_w + s_w \cdot \mathbf{r}. \end{aligned} \tag{1}$$

2.3 Temporal Encoding Module (TEM)

Since the time interval between two follow-up scans can be rather different, inspired by positional encoding [15] we develop a Temporal Encoding Module (TEM) to embed time interval information into the prediction model. Due to the limitation of dataset size, we discretize the interval using time mapping function $t_{itv} = \lceil t_{day}/30 \rceil$ with an upper cut-off value 20, for most of the intervals are less than 600 days. Sine and cosine functions with different frequencies are used in the TEM to generate values of different dimensions of the encoded temporal feature vector:

$$\begin{aligned} \mathcal{F}_{TE}(t_{itv}, 2i) &= \sin(t_{itv}/100^{2i/d_{fm}}), \\ \mathcal{F}_{TE}(t_{itv}, 2i+1) &= \cos(t_{itv}/100^{2i/d_{fm}}), \end{aligned} \tag{2}$$

where t_{itv} is the discretized time interval, d_{fm} is the total number of channels of the encoded feature vector and i is the dimension. That is, the even/odd dimensions of the temporal encoding are generated by sin/cos function with different wavelengths (2π to $100 \times 2\pi$), which makes the relative time information encoded in a redundant way. Besides, the value range of the encoding result is within a certain numerical interval due to the boundedness of sinusoid. These two points ensure that the temporal encoding method can generate a more meaningful high-dimensional representation space. Then the feature vector is expanded repetitively and concatenated with the bottom feature map of WarpNet and TextureNet.

[1] If no otherwise specified, image mentioned here and later in this article refers to $48 \times 48 \times 48$ cubic volume image with a nodule in the center.

2.4 WarpNet for Spatial Transformation

As the core of our method, WarpNet predicts a displacement field $\mathbf{u}$ to model the shape variation of nodule growth, which is similar to the motion prediction in video tasks [6,8,18]. The architecture of WarpNet is based on a CNN similar to U-Net [14] with skip connections, and the temporal encoding from TEM is connected to the bottom of WarpNet.

The loss function for training WarpNet θ_w has four terms: similarity loss $\mathcal{L}_{sim}$ between warped images x_w and target images y, segmentation loss $\mathcal{L}_{seg}$ between warped segmentation maps s_i and target maps s_t, smoothness loss $\mathcal{L}_{smooth}$ for the deformation field and regularization loss $\mathcal{L}_{reg}$ for the output of WarpNet when $t_{itv} = 0$. In summary, the learning of WarpNet is formulated as:

$$\hat{\theta_w} = \underset{\theta_w}{\operatorname{argmin}}\{\mathcal{L}_{sim}(x_w, y) + \lambda_1 \mathcal{L}_{seg}(s_w, s_t) + \lambda_2 \mathcal{L}_{smooth}(\mathbf{u}) + \lambda_3 \mathcal{L}_{reg}(\phi_0)\} \quad (3)$$

with weights $\lambda_1, \lambda_2, \lambda_3 > 0$, where ϕ_0 is the predicted spatial warp function when time interval $t_{itv} = 0$. All loss functions are designed as follows:

a) similarity loss and regularization loss: In our experiments we find that for spatial transformation normalized cross correlation (NCC) loss leads to more reasonable and robust results than MSE loss. The NCC loss between warped image x_w/target image y and the regularization loss for ϕ_0 is defined as:

$$\begin{aligned} \mathcal{L}_{sim}(x_w, y) &= 1 - NCC(x_w, y) = 1 - NCC(\phi \circ x_i, y), \\ \mathcal{L}_{reg}(\phi_0) &= \mathcal{L}_{sim}(\phi_0 \circ x_i, x_i) + \mathcal{L}_{sim}(\phi_0 \circ y, y). \end{aligned} \quad (4)$$

b) segmentation loss: We use Dice loss to constrain the similarity between warped segmentation mask s_w and target mask s_t:

$$\mathcal{L}_{seg}(s_w, s_t) = 1 - \frac{2 \cdot \sum_{p \in \Omega} s_w(p) s_t(p)}{\sum_{p \in \Omega} s_w(p) + \sum_{p \in \Omega} s_t(p)}. \quad (5)$$

c) smoothness loss: Considering that the contour of nodule changes continuously as it grows, we use a diffusion regularization loss to encourage the smoothness of displacement field $\mathbf{u}$:

$$\mathcal{L}_{smooth}(\mathbf{u}) = \frac{1}{|\Omega|} \sum_{p \in \Omega} ||\nabla \mathbf{u}(p)||^2. \quad (6)$$

where finite differences between neighboring voxels are used to approximate the spatial gradients $\nabla \mathbf{u}(p)$ (for x, y, z 3 dimensions).

2.5 TextureNet for Texture Transformation

In addition to the shape variation, there is also a texture variation in nodule growth caused by the change of CT value distribution of nodules. So a TextureNet is needed to estimate the residual between warped image x_w and target

Table 1. Quantitative results of multiple models. We choose U-Net w/ or w/o TEM and WarpNet w/ or w/o segmentation loss as comparisons of *NoFoNet*. The performance is estimated by PSNR, PSNR* (PSNR in the nodule parts), dice coefficient between warped/target segmentation maps and sensitivity/specificity/G-mean for PD/non-PD classification. We evaluate the performance of our models on our in-house dataset (see Sect. 2.1) with 5-fold cross validation.

Method	PSNR	PSNR*	Dice	Sensitivity	Specificity	G-mean
Baseline (U-Net)	4.1213	29.7490	–	–	–	–
+TEM	6.0380	31.8821	–	–	–	–
WarpNet	18.0915	43.1140	0.6301	0.7656	**0.9083**	0.8339
+Warp Seg Loss	18.1952	43.2464	0.6474	0.8594	0.8805	0.8699
+TextureNet	**18.2089**	**43.4904**	**0.6474**	**0.8594**	0.8805	**0.8699**

image y. TextureNet follows the architecture of WarpNet. To train TextureNet we need an intensity similarity loss $\mathcal{L}_{sim}^{'}$ between textured images $\hat{y}$ (see Eq. 1) and target images y, and a regularization loss $\mathcal{L}_{reg}^{'}$ for the predicted residual $\mathbf{r_0}$ when $t_{itv} = 0$. So the texture transformation learning is formulated as:

$$\hat{\theta}_t = \underset{\theta_t}{\operatorname{argmin}}\{\mathcal{L}_{sim}^{'}(\hat{y}, y) + \lambda_1^{'}\mathcal{L}_{reg}^{'}(\mathbf{r_0})\} \tag{7}$$

with weight $\lambda_1^{'} > 0$. We choose MSE loss to encourage maximal intensity similarity. The loss functions of TextureNet are defined as:

$$\begin{aligned} \mathcal{L}_{sim}^{'}(\hat{y}, y) &= \frac{1}{|\Omega|}\sum_{p\in\Omega}[\hat{y}(p) - y(p)]^2, \\ \mathcal{L}_{reg}^{'}(\mathbf{r_0}) &= \frac{1}{|\Omega|}\sum_{p\in\Omega}|\mathbf{r_0}(p)|^2. \end{aligned} \tag{8}$$

2.6 Implementation Details

NoFoNet can use any CNN architecture for WarpNet and TextureNet, and we use the network design of Appendix Fig. A.1 in this work. All of the experiments in this study are implemented on an NVIDA Titan X GPU and an Intel i7-6700 CPU. Our codes are based on Python 3.7.3 and PyTorch-1.2.0 [10]. We use $\lambda_1 = 0.5, \lambda_2 = 10$ and $\lambda_3 = \lambda_1^{'} = 1$ for the loss weights in Eq. 3 and Eq. 7. Online data augmentation methods, including rotation and flipping along a random axis, are applied on the input images. Each part of *NoFoNet* is trained using Adam optimizer [7] with an initial learning rate of 0.001 for 200 epochs. Specifically, we emphasize the similarity loss inside the segmentation map to put more attention on the nodule.

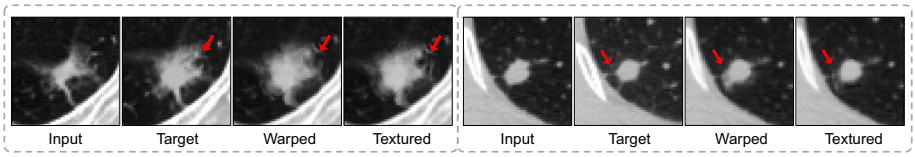

Fig. 2. Comparison of warped images and textured images for a PD case (left) and a non-PD case (right). Areas where intensity is changed significantly by TextureNet are indicated by red arrows. (Color figure online)

3 Experiments

3.1 Evaluation Protocol

Our *NoFoNet* is trained to predict what the nodule may be visually like after a certain time interval, then we can determine whether it is a PD (progressive disease, i.e., significant growth of nodule in size) case. Since some nodules in our dataset have multiple follow-ups, we stipulate that a nodule is judged as a PD case as long as one of its follow-up pairs (see Sect. 2.1) meets specific criterion, which is determined with the help of two senior radiologists.

Define V_1, V_2 (mm^3) as the two nodule volumes of a follow-up pair with time interval T (d), the criterion is as follows: (1) Considering that the fast-growing nodules have higher risks, we calculate the average volume growth rate $\text{AVGR} = (V_1 - V_2)/T$, and set a threshold of 1; (2) Some cases may have AVGR less than 1 but eventually grow significantly in size, we set a threshold of 200 for volume difference $\text{VD} = V_1 - V_2$ and a threshold of 50% for relative volume difference $\text{RVD} = (V_1 - V_2)/V_1$. In summary, a nodule is classified as PD case if one of its observed $\text{AVGR} \geq 1\,\text{mm}^3/\text{d}$, or $\text{VD} \geq 200\,\text{mm}^3$ and $\text{RVD} \geq 50\%$.

The 315 nodules are divided into two parts according to the aforementioned criterion, resulting in 64 positive cases (PD, significant growth) and 251 negative cases (non-PD, stable or shrinking). We split our dataset randomly into 5 groups based on patients (i.e., all nodules of one patient must be in same subset) and perform 5-fold cross validation to evaluate our models.

3.2 Performance Analysis

In this section we will present some quantitative results and qualitative results. Table 1 shows the performance of our models and baselines using 5-fold cross validation method. Note that U-Net w/ or w/o TEM predicts output images directly so it only has PSNR and PSNR* (PSNR in the nodule parts) for output/target images. As is shown in Appendix Fig. A.2, U-Net baselines generate predicted images with low visual quality. It is noticeable that when added segmentation loss for warped/target images, WarpNet predicts more accurate displacement fields, resulting in higher dice coefficient between warped/target images and better performance for PD/non-PD classification than WarpNet without segmentation loss. We use the geometrical mean (G-mean) of sensitivity (TP/TP+FN) and

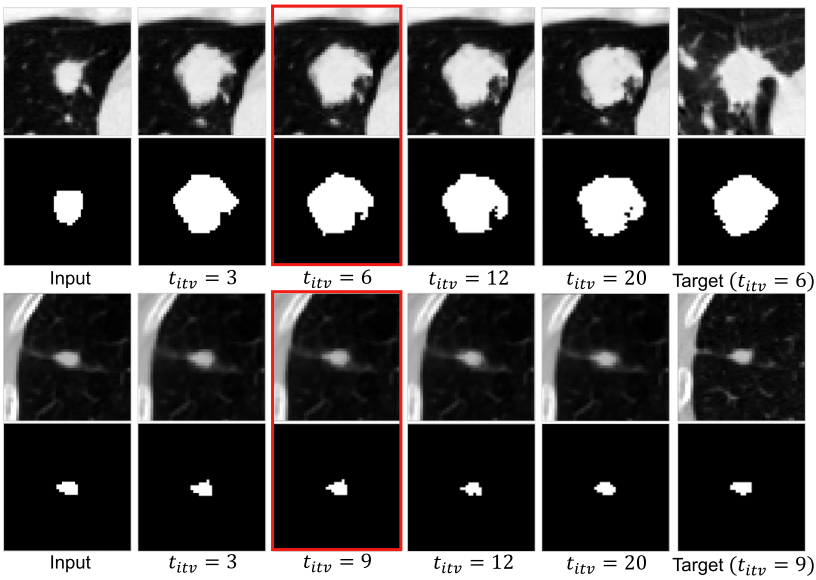

Fig. 3. Continuous prediction results of a PD case (top) and a non-PD case (bottom) by WarpNet. The first and last columns are the input image/segmentation and the target image/segmentation, and columns in the middle are the warped images/segmentations by WarpNets with different temporal encodings. Warped results that have the same time interval as the targets are highlighted in red. (Color figure online)

specificity (TN/TN+FP) as main evaluation index for the unbalanced dataset. The TextureNet in *NoFoNet* improves the visual quality of the warped images and achieves higher PSNR/PSNR* scores, as visually shown in Fig. 2.

Figure 2 shows the results of spatial transformation for input images by WarpNet and voxel-wise texture addition for warped images by TextureNet. We select a PD case and a non-PD case to demonstrate the performance of *NoFoNet* on different types of nodules. It can be seen that TextureNet is able to refine the warped images from WarpNet and increase the intensity similarity between the predicted and target nodules. Please refer to Appendix Fig. A.2 for more comparison results (including results from U-Net).

Figure 3 illustrates the continuous prediction results of two nodules using WarpNet. Note that results with the same follow-up time interval as the targets are highlighted in red. We choose a PD (progressive disease) case and a non-PD case for contrast to show that our WarpNet can represent both significant growth and stabilization of nodules in size well. For PD case it can also be seen that the model is able to generate reasonable nodules as time interval changes and the variation tendency is plausible, indicating the effectiveness of TEM.

4 Conclusion

We develop the *NoFoNet*, a unified network to predict the tumor growth for lung nodules. By explicitly learning spatial transformation and texture transformation, it yields high-quality visual appearances and accurate quantitative results, with validated effectiveness on an in-house dataset from two clinical centers. To the best of our knowledge, this is one of the first study to predict nodule growth quantitatively (size) and visually (appearance) given any time interval during.

A limitation of this study is that we only model the tumor growth as the indicator of nodule risk. However, according to TNM tumor staging system, tumor size (T), lymph node (N) and metastasis (M) are considered in tumor prognosis assessment. In future studies, we will address the N and M information to develop a more advanced risk stratification system for lung nodule follow-up. Besides, we will expand the dataset from cooperative hospitals and explore more effective architectures and temporal encoding methods for our framework.

Acknowledgment. This work was supported in part by National Natural Science Foundation of China (61671298), 111 project (BP0719010), Shanghai Science and Technology Committee (18DZ2270700) and Shanghai Jiao Tong University Science and Technology Innovation Special Fund (ZH2018ZDA17).

References

1. Ardila, D., et al.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. **25**, 954–961 (2019)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
3. Cressman, S., et al.: Resource utilization and costs during the initial years of lung cancer screening with computed tomography in Canada. J. Thorac. Oncol. **9**, 1449–1458 (2014)
4. Huang, P., et al.: Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. Lancet Digit. Health **1**, e353–e362 (2019)
5. Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 249–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_20
6. Jin, X., et al.: Predicting scene parsing and motion dynamics in the future. In: NIPS, pp. 6915–6924 (2017)
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Luc, P., Neverova, N., Couprie, C., Verbeek, J., LeCun, Y.: Predicting deeper into the future of semantic segmentation. In: ICCV, pp. 648–657 (2017)
9. MacMahon, H., et al.: Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. Radiology **284**(1), 228–243 (2017)
10. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)

11. Petersen, J., et al.: Deep probabilistic modeling of glioma growth. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 806–814. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_89
12. Pinsky, P.F., et al.: Performance of lung-RADS in the national lung screening trial. Ann. Intern. Med. **162**, 485–491 (2015)
13. Pinsky, P.F., Gierada, D.S., Nath, P., Kazerooni, E.A., Amorosa, J.: National lung screening trial: variability in nodule detection rates in chest CT studies. Radiology **268**(3), 865–73 (2013)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
15. Vaswani, A., et al.: Attention is all you need. In: NIPS, pp. 5998–6008 (2017)
16. Wood, D.E.: National Comprehensive Cancer Network (NCCN) clinical practice guidelines for lung cancer screening. Thorac. Surg. Clin. **25**(2), 185–97 (2015)
17. Xie, Y., Xia, Y., Zhang, J., Feng, D.D., Fulham, M., Cai, W.: Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 656–664. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_75
18. Xu, J., Ni, B., Yang, X.: Video prediction via selective sampling. In: Advances in Neural Information Processing Systems, pp. 1705–1715 (2018)
19. Yang, J., Deng, H., Huang, X., Ni, B., Xu, Y.: Relational learning between multiple pulmonary nodules via deep set attention transformers. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1875–1878. IEEE (2020)
20. Yang, J., Fang, R., Ni, B., Li, Y., Xu, Y., Li, L.: Probabilistic radiomics: ambiguous diagnosis with controllable shape analysis. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 658–666. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_73
21. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: CVPR, pp. 8543–8553 (2019)
22. Zhao, W., et al.: Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning. Cancer Med. **8**(7), 3532–3543 (2019)
23. Zhao, W., et al.: 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. Cancer Res. **78**(24), 6881–6889 (2018)

Appendix from *Learning Tumor Growth via Follow-Up Volume Prediction for Lung Nodules*

A Appendix Figures

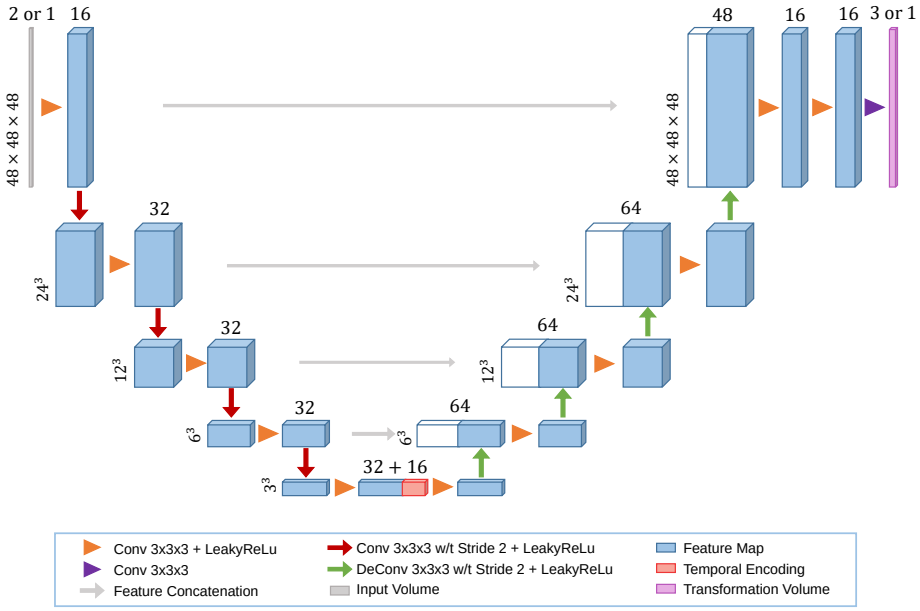

Fig. A.1. Architecture of WarpNet and TextureNet with temporal encoding. The channel of input volume and transformation volume is 1 and 3 in WarpNet, and the corresponding channel is 2 and 1 in TextureNet. 2-stride convolution is used for down-sampling and 2-stride deconvolution is used for up-sampling. The U-Net architecture in our work is quite simple and could be easily extensible.

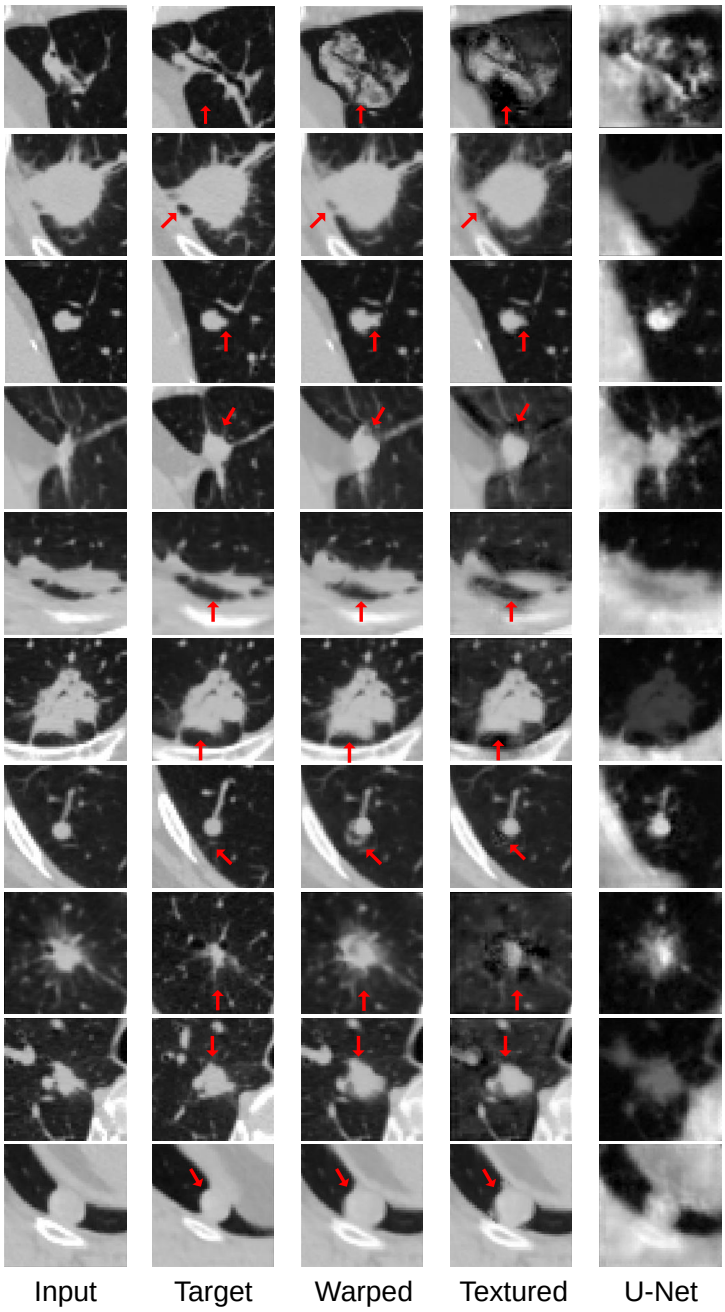

Fig. A.2. Comparison of results from WarpNet, TextureNet and U-Net with TEM. We highlight (by red arrows) the areas where intensity is changed significantly by TextureNet. It can be seen that TextureNet is able to refine the warped images from WarpNet significantly, making the textured images more visually similar to the target images, even though the average PSNR metric over the whole volume images in Table of outputs by WarpNet and TextureNet is close. On the other hand, U-Net generates images of low visual quality, which is consistent with corresponding PSNR in Table 1.