



Functional-realistic CT image super-resolution for early-stage pulmonary nodule detection

Hongbo Zhu^{a,e}, Guangjie Han^b, Yan Peng^{c,*}, Wenbo Zhang^a, Chuan Lin^b, Hai Zhao^{d,e}

^a School of Information Science and Engineering, Shenyang Ligong University, Shenyang, 110159, China

^b Key Laboratory for Ubiquitous Network and Service Software of Liaoning province, School of Software, Dalian University of Technology, Dalian, 116024, China

^c Research Institute of USV Engineering, Shanghai University, Shanghai, 200444, China

^d School of Computer Science and Engineering, Northeastern University, Shenyang, 110169, China

^e Neusoft Research of Intelligent Healthcare Technology, Neusoft Corporation, Shenyang, 110179, China



ARTICLE INFO

Article history:

Received 10 January 2020
Received in revised form 28 August 2020
Accepted 19 September 2020
Available online 3 October 2020

Keywords:

Super resolution
Pulmonary nodule detection
Semantic patterns
CADs
Data augmentation

ABSTRACT

Early-stage pulmonary nodule detection is challenging for Computer-aided Diagnosis systems (CADs) in clinical practice. It always relies on large-scale annotated pathological images. Unfortunately, the limited voxels of earlier-stage nodules can aggravate the risk of escaping diagnosis. Due to high-dose CT and bronchoscope potential threats, CT image super-resolution has become a suboptimal way to tackle the problem. Therefore, we proposed a deep generative adversarial network (GAN) architecture based on a deep grammar model, called FRGAN (Functional-Realistic GAN). By using region proposal network (RPN), the bottom semantic features are recommended and classified as the basic units of functional structure. Local pathological images can be hierarchically aggregated with corresponding to different semantic patterns as parse trees. Refer to their EMRs, and we use TreeGAN to generate the correct syntax patterns for each early-stage pulmonary nodule candidates. We report the generated results of the super-resolution images and feed them into a convolutional network to assess the functional loss of the generated results along to different parse trees. Within the contextual and generative losses, we rebuild a novel objective function paralleling with TreeGAN. The aim is to boost the sensibility of pulmonary nodule detection with more functional-realistic data augmentation. Experimental results show that our generator can faster generate more realistic SR images with pathological features. Moreover, it could be a data augmentation tool for some deep architecture to overcome sample imbalance.

© 2020 Published by Elsevier B.V.

1. Introduction

Lung cancer is one of the fatal disease, which has killed 1.6 million people, about 19% of all cancer deaths in 2017 worldwide [1]. It is known that earlier-stage lung cancers have more possibility of being cured, but only 15% of all the diagnosed cases are preliminary, almost detected with chest computed tomography (CT) in the routine checkup. So far the CT is the best way to detect early-stage pulmonary nodules because of non-invasive and inexpensive.

Recall the entire reading process of CT image. The radiologists always screen all the CT sequences images, in which the early-stage instances are just imaging with the limited pixels, even if increasing X-ray dose [2]. When the radiologists find a suspected lesion with low resolution (LR), they always advise the examinees

to review invasive examinations (e.g., bronchoscope) to get more rigorous diagnosis. However, the examinations are uncomfortable and risky (e.g., infection and asphyxia). Hence, the image enhancement is a suboptimal solution for early-stage pulmonary nodule detection, and worth being boosted in the applications of Computer-Aided Diagnosis/Detection system (CADs).

uper-resolution (SR) is a conventional solution for enhancing image quality. The methods utilize the pixel and features of low-resolution (LR) image to reconstruct its high-resolution (HR) versions, such as linear and cubic-spline interpolation methods. For most of the SR methods, their optimization targets minimize the mean squared error (MSE) between the recovered HR image and the ground truth (GT). This category of methods is ill-posed for capturing high texture detail. The smaller MSE can only increase the peak signal-to-noise ratio (PSNR), but not improve the perceptual experience. To overcome the drawback, SRGAN (Super Resolved Generative Adversarial Networks) [3] defines a perceptual loss function for its generator to focus on better perceptual information of the generated HR counterparts. Its results fulfill

* Corresponding author.

E-mail address: pengyan@shu.edu.cn (Y. Peng).

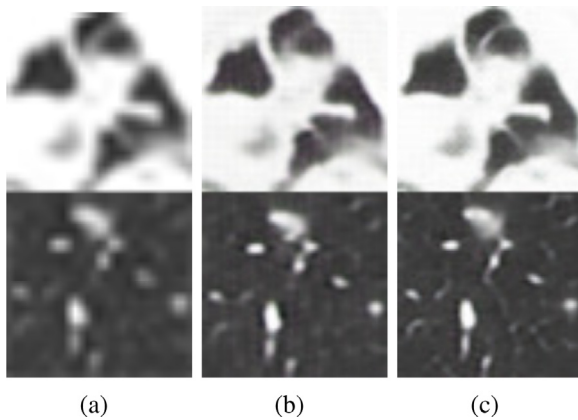


Fig. 1. (a) Original LR images; (b) SRGAN-based SR images and (c) FRGAN-based SR images, generated from the LR counterparts after 20 training epochs.

the user demands for photo-realistic but are also questionable. Are the generated images objective enough for more rigorous generative tasks, such as nodule detection? As the exploratory works, some medical SR methods are used for data augmentation to solve the imbalance ratio of the positive/negative samples in the original dataset. They can be roughly divided into two representative methods, including noise-to-image methods and image-to-image methods. The former group is fully referred to as the idea of the original GAN [4], generating the pathological image by random noise samples. Another is always used to produce a malignant derivative from a benign image by appending some malignant pathological conditions. Compared with the noise-to-image methods, the latter group can significantly reduce computational cost. Moreover, its procedure can be understood as a semantic editing process. Different from the photo-realistic constraints, the editing the process needs the higher-level semantic constraints to map the features extracted from the input LR images to those in the generated HR counterparts. Fig. 1 shows a series of the generative cases, in which the first column contains two original low-resolution inputs, the other columns list the generative $4\times$ SR results of SRGAN and the proposed work, respectively.

We observe that the SRGAN-based super-resolved images are realistic enough to make them indistinguishable from their ground truths, which reflect the same visual response as the definition of photo-realism [5]. Besides, we believe another realism need to be followed to emphasize the functional cognition of the generated HR images, called functional-realism. It could quantify the medical image's functional structures to supply more reliable imaging information as the visual criteria for suspected nodule detection. We use functional-realism as an editing criterion of an image-to-image GAN architecture called functional-realistic GAN (FRGAN) to ensure the more exact visual semantic information for the paper's generated HR images. We try to generate high-resolution counterparts for an input early-stage nodule image. Meanwhile, we also learn the semantic rules of the generator. For the first goal alone, it has lots of SISR solutions, such as SRResnet, SRGAN [3] and WGAN-GP. Both for PSNR and SSIM, these methods have reached higher performance while gaining unsatisfactory MOS. Some improved image-to-image methods are proposed to reconstruct ROI/VOI with the additional, conditional setting, including size, location, and background context. However, the generator cannot yield its desired results than its expensive cost. The main reason for the non-ideal results come the conditional setting without semantic production rules which fit with the doctors' knowledge. Therefore, we supplement

an assistant generating process for structural-semantic feature sequences.

We also need an explicable mode of the fundamental semantic unit and the corresponding grammar model for a better structural-semantic representation. To meet the requirements of the model's scalability and flexibility, attributed relational graph. (ARG) [6] has almost become the optimal solution for the editable semantic representation. Its interesting follow-up study, named AOGNets [7] can split the input feature map into different feature channels as a sentence of words, and then hierarchically select and integrate the words into a word group (e.g., phrase/sentence) based on a structure/dependency grammar, respectively. A more convenient and logical editable word group offers better performance for feature exploration and reuse. Due to the natures of the pulmonary nodule, its image features and the semantic features of its EMR are the different ways to describe the identical image in a standard diagnostic procedure, where the parse tree for semantic structure, the AOGs for visual feature structure. FRGAN is also an image-to-image GAN, parallels two generators for semantics and visual features, respectively. Its generative stage likes a sentence making, and the discriminator is used to point out the ill-formed sentences. The two main contributions of this paper include:

(1) It proposes a novel image-to-image GAN architecture for early-stage pulmonary nodule detection, called FRGAN. Its generative stage is abstracted as a process of sentence completion based on AOGNets. More complicated features are explored, reused, and then gradually socketed into the original inputs' suitable fields in a hierarchical and compositional way. Due to the grammar-guided functional combination, the generator owns a better performance to fabricate the HR nodules' image with the potential malignant pathologies of its LR inputs;

(2) It implies a proper data augmentation (DA) method to balance the positive/negative ratio (e.g., 1:1), which may boost the detection performance of the CNN-based discriminator for the false positive reduction task in pulmonary nodule detection. On 1034 scans merging with the self-collected dataset and LUNA16 dataset, the proposed architecture can achieve a CPM score of 0.915.

The remainder of this paper is organized as follows: we detail our methodology in Section 3, and report our experimental results and discuss some critical options in Section 4, followed by concluding our contribution in Section 5.

2. Related works

CADs had been rapidly developed in the last two decade [8]. Especially after the introduction of Artificial Intelligence (AI), continuous progress was emerged, which gains more accuracy and convenience. Moreover, the earlier-stage nodules were detected via more intelligent architectures. The methods can be roughly divided into two categories: hand-crafted feature and learning-based feature. The former relies on feature engineering [9–13], including feature preprocessing, extraction, selection, and dimension reduction analysis. The features are extracted from the grayscale, gradient, spectral channels, and transforms. As the learnable features, part of them are selected out to form the subset of important information as the inputs of many classification architectures, such as KNN [14], SVM [15] and so on. These methods strongly depend on the effectiveness of feature descriptors without massive data resources. However, the descriptors are always customized for certain data classification, which can limit their more applications. The latter is dominant in the current methods due to the research upsurge of DL(Deep Learning)-based methods [16–20]. These methods can automatically extract the raw data feature and obtain high-level representation by multi-layer convolutional and pooling operators [17].

They encourage the diversity of features via deeper network architecture. Some problems with feature quantization and interpretation give a dramatic leap both in performance and usability. However, more complexity of deeper architectures demands more amount of parameter tuning. Data-driven gives a new challenge for data quality and sample balance. The aforementioned drawbacks cause the rocketed increasing of human and computational resource costs. So combining the effective feature engineering and deep learning architecture is a necessary solution in further research. Deep learning-based methods have been carried out for detecting/classifying both natural and medicine images with the development of computational cognitive science in the recent 5-years. An army of their outstanding performances promotes the novel stride of computer vision. Though the maturing technologies have been advanced at a breathless pace, the different input image qualities limit them in clinical application.

For super-resolution for medical images, the existing methods inherit the outstanding performance of similar works in the field of natural image processing. Jebaveerasingh et al. [21] proposed a super-resolution (SR) algorithm to process the retinal images captured using smartphone funduscopy based on a multi-kernel support vector regression (SVR), which can improve the quality of the captured images. Ren et al. designed a Deep Resnet (deep residual network)-based multi-scale reconstruction model for different medical lesion images. Chen et al. [22] proposed a trusted deep convolutional neural network to generate SR images for medical images, and their method used a feedback connection between the information of the output images and the low-level features, which can reduce the feature redundancy by using an adaptive weighted dense block. Moreover, some optimizer for high-dimensional data analysis method [23] captured our attention, which can offer a guideline for multiple ROI detection.

3. Problem formulation

3.1. Function aware sequence generation

We focus on the semantic feature extraction from pulmonary nodule images at first, due to the necessary connection between semantic features and function. For each function, we can consider them as a latent pattern that covers structural-semantic features. Refer to the nine semantic features of nodule patterns listed in [24], we could initialize the functional attribute set, and follow the text tokens “Tex”, “Sub”, “Spi”, “Sphe”, “Mar”, “Mal”, “Lob”, “IS”, and “Cal” for “texture”, “subtlety”, “spiculation”, “sphericity”, “margin”, “malignancy”, “lobulation”, “internal structure”, and “calcification”, respectively. We note that a nodule pattern may be composed of more than one attribute and the attributes are organized as an hierarchical feature structure functioning in the nodule patterns. Specifically, “Mal” is located at the top of the tree structure, followed with “Tex”, “Mar” and “Sub”, which are defined as global descriptions for a whole nodule pattern. The others are the bottom features conditionally integrating various of local descriptors for their superior attributes. Again, some local semantic features can be discovered in the different layers of a multi-scale local image. Hence, the function aware sequence generation problem is defined as follows.

Definition 1. Given a dataset of real HR images with a structural feature sequences $X = \{X_1, \dots, X_N\}$, where all the $X_n \in X$ follows a grammar \mathcal{G} . A NN (Neural Network)-based generator G_θ with a generative model θ can be trained to construct a T -component sequence $\mathcal{Y} = \{Y_1, \dots, Y_T\}$ based on a structural feature sequences $\mathcal{X} = \{X_1, \dots, X_M\}$ from a LR images, where Y is the vocabulary set with $\mathcal{Y}_t \subseteq Y$.

3.2. Grammar

Recall the descriptions of the malignant/benign nodules from the radiologists recorded in EMR (Electronic Medical Record), the nine semantic features are organized into a hierarchical formal language. The entire generation problem can be formulated as a grammar $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$, where \mathcal{V} is a set of non-terminal semantic features, $\mathcal{T} = \mathcal{V} \cup \epsilon$ is the set of terminal features.

3.3. Parse tree

According to the overlaps of different ROI from a given nodule, which bring the possible vocabulary reuses, we use a RPN-based architecture [25] to recommend the bounding boxes of the latent nine semantic features, and annotate them with 1–4 degrees by three expert radiologists. Each sub-image is a derivation of the whole sequence, its features were recorded in the EMR from three radiologists, including local and global pathological descriptions. A parse tree construction [26] can be executed following context-free grammar $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$. For a global pathological description of real HR/LR sub-image, the grammar owns the following properties:

- (1) The root node S is a complex description of any global feature.
- (2) The interior node belongs to a variable in \mathcal{V} .
- (3) All the leaves are covered by the terminal set \mathcal{T} , corresponding to 6 possible local semantic features.
- (4) If a local feature labeled \mathcal{A} , its children are flatted by an order of the box sizes from left to right, and followed a production rule in \mathcal{P} .

By comparing the different-stage patients' EMRs, the descriptions of the early nodules tend to the comprehensive representations, and those of the later ones are close to the more detailed descriptions with local representations of pathological features and comprehensive representations of the nodules. In clinical practice, the medical advice for an early nodule is always pointed out its malignancy or potential malignancy for a caution to the patient and his family. The doctors give the therapeutic schedules based on nodule size, position, feature, and class for the later-stage cases, instead of the further malignant assessments. So we can transform the function-aware sequence generation problem into a parse tree generation problem.

Definition 2. Given a dataset of real HR images with a structural feature sequences $X = \{X_1, \dots, X_N\}$, where all the $X_n \in X$ follows a grammar \mathcal{G} , a NN (Neural Network)-based generator G_θ with a generative model θ can be trained to construct a SR derivation of a T -component sequence $Y = \{Y_1, \dots, Y_T\}$, $Y_t \subseteq \mathcal{Y}$ from a structural feature sequences $\mathcal{X} = \{X_1, \dots, X_M\}$ of LR images with a sequence of production rules $\Gamma_T = \{\gamma_1, \dots, \gamma_T\}$, $\gamma_t \in \mathcal{P}$.

3.4. Function-to-image

For any production rules, a series of mapping is built from a structural-semantic feature sequence of LR image to that of HR images. However, the semantic inference process has some apparent differences from a hierarchical construction for the real image representation, which has various spatial topologies. So a hierarchical encoding mechanism need be introduced to initialize a context graph to deeply integrates hierarchical and compositional grammar and deep architecture. For local semantic features, ARG (Attribute Relational Graph) [6] is a suitable solution for fast building the structural features. In a visual recognition task, it is always followed by mining AOGs (And-Or Graphs) from the positive/negative ARG templates for object/module detection. Refer to AOG [7], terminal nodes are activation units on feature

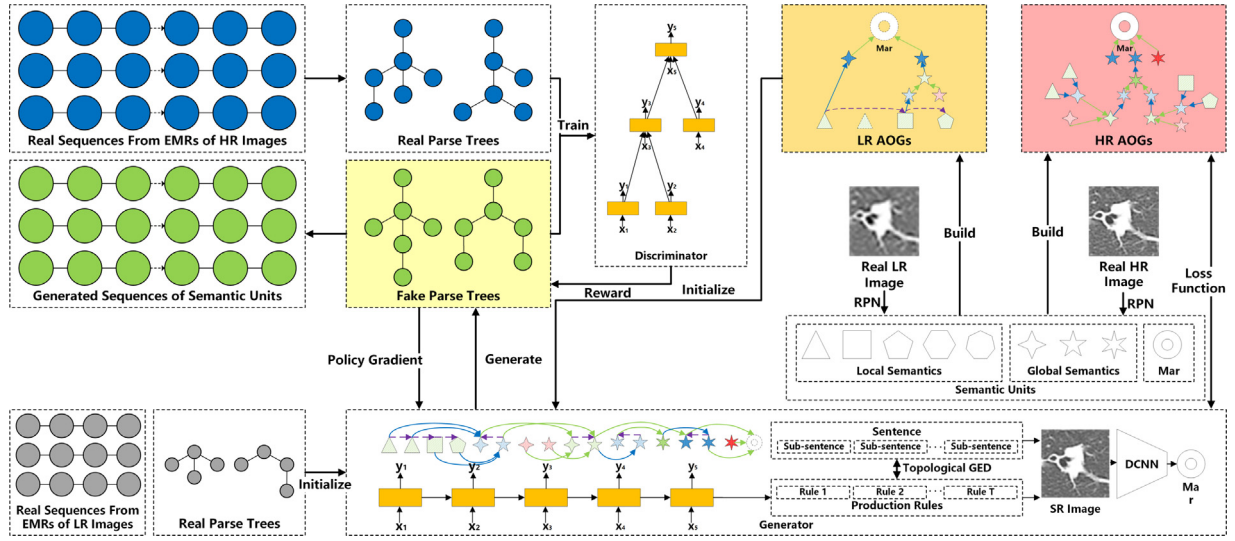


Fig. 2. The overall of FRGAN architecture.

maps of a CNN-like filter in our work. Part of the activated units could be naturally selected and converged into an OR node to represent a constituent region or contextual region. While appending one of its children unit, a new geometric deformation is generated for a latent pattern. Many latent patterns are concatenated to further encode a constituent region as a candidate for the top semantic part. Finally, the multi-layer AOG blocking implicit latent patterns in raw CNN feature maps to explicit semantic parts.

4. Methodology

We show the proposed framework in Fig. 2. As a derivative from the original GAN formulation [4], a generator and a discriminator are adversarial to gain a synchronized evolution for high performance. In the existing outstanding methods, the former is used to learn a mapping from a series of random noise vector to an output image I (noise-to-image generator) or redraw a new counterpart of the referenced image by appending some noise vector (image-to-image generator), as the defined problem in Section 3.1, and its optimization objective is

$$L(\theta, \phi) = \mathbb{E}_X[\log D_\phi(X)] + \mathbb{E}_Y[\log(1 - D_\phi(Y))] \quad (1)$$

Both of the above generative patterns are maximizing the likelihood estimation of the generated distribution to close to the distribution of the real HR images, in which some Gaussian noise units are assembled via a generator G_θ . The SR outputs of the model are repeatedly tried to fool the discriminator D_ϕ , aiming to produce enough realistic outputs. Meanwhile, due to the more realistic generated counterpart, the adversarial process can synchronously promote the discriminator performance. For an image-to-image generator, the generating step always depends on a conditional setting to composite nodule image with surrounding tissues to generate more realistic images with fixed attributes. Because the limited conditional setting is generally the top feature-oriented, we replace the setting with a semantic guideline built from its EMR for performing a sequence of production rules as Definition 2 in Section 3.3. Hence, we set an assistant generator to construct the corresponding production rules for guiding a generating process of SR images with more believable topological attributes.

4.1. Preprocessing

A RPN-based architecture [25] is used to recommend the bounding boxes for the six local semantic features (i.e., words), which uses five-group convolutional layers of VGG-16Net as multi-layer feature extractor due to the much smaller size of pulmonary nodules. Besides, a deconvolutional layer is followed with the extractor to enhance the limited descriptions for each image feature. To generate ROIs and their predicted labels, a small network uses a 3×3 sliding window over the feature map of the deconvolutional layer to produce the inputs for a couple of fully-connected layers output the bounding box and the predicting objectness distribution for words. The same work is beforehand manually finished by three expert radiologists as the referenced labels. To fit the size of the word and avoid the overlaps, we set 3 anchors with different sizes of sliding windows 2×2 , 4×4 , and 6×6 . A deep 2D DCNN with ROI pooling layer [19] is used to fine adjust the classes and locations of the visual words.

4.2. Functional-consistency generator

After the preprocessing in Section 4.1, an input nodule image can transform a set of visual words. However, the words are unordered, which cannot be suitably arranged into a structural feature sequence. However, in a radiologist's view, the scattered visual words could be divided into different channels, following a grammar $\mathbb{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, \mathcal{S})$ as the defined problem in Section 3.2. Hence, we can follow the approach proposed in TreeGAN [26] to use the policy gradient to guide the learning of G_θ .

4.2.1. Semantic guideline generator

For an EMR of a given real LR nodule image, G_θ starts from its global semantic features as a series of root nodes. Depend on choosing different production rules from the EMRs of similar HR images with an additional pathological feature set. The trees are growing to hierarchically expand the branches filled with local semantic features until its leaf nodes are generated terminal tokens to stop growing. The generation proceeds in DFS and

left-right order to feed a vanilla LSTM unit as a tree generator:

$$\begin{aligned} i_t &= \sigma(\mathbf{W}^{(i)}x_t + \mathbf{U}^{(i)}h_{t-1} + b^{(i)}), \\ f_t &= \sigma(\mathbf{W}^{(f)}x_t + \mathbf{U}^{(f)}h_{t-1} + b^{(f)}), \\ o_t &= \sigma(\mathbf{W}^{(o)}x_t + \mathbf{U}^{(o)}h_{t-1} + b^{(o)}), \\ u_t &= \tanh(\mathbf{W}^{(u)}x_t + \mathbf{U}^{(u)}h_{t-1} + b^{(u)}), \\ c_t &= i_t \odot u_t + f_t \odot c_{t-1}, \\ h_t &= o_t \odot \tanh(c_t), \end{aligned} \quad (2)$$

where $x_t = (a_{t-1}, p_t)$ is a two-tuples, including the action embedding vector for the data d_{t-1} of a data sample $D = d_1, \dots, d_{t-1}, d_t, \dots, d_T$ at time step $t - 1$ and its parent embedding vector p_t for d_t . By using the generation state tracking algorithm [26], we can generate a set of the potential semantic topologies for an input LR image.

4.2.2. Image-to-image generator

When we take a generated rule of a parse tree from EMR, we note that it cannot be immediately used as a guideline for encoding the image features to a SR image. Because the generated tree is considered a parse-making process without any consideration, such as its unary, pairwise, and topological features. Hence, we use AOG-based phrase grammar [7] naturally package spatial and contextual features for multi-layer structure.

More specifically, the phrase structure grammars are defined by three rules to describe three states of its sub-sentences, called Termination rule, And rule, and Or rule, respectively.

For a non-terminal symbol $S_{i,j}$, the Termination rule is used to assign the whole symbol to a sub-sentence $Ter_{i,j}$ starting at the i th word and ending at the j th word. i.e. The symbol maps all its words to a feature channel which are propagated in forwarding computation.

In fact, for a better description of a more complicated pathology, its feature is always integrated with two or more than two simple features. Thus, the And rule is defined as a binary decomposition for a non-terminal symbol $S_{i,j}$. The m th word ($i < m < j$) is a breaking point to decompose $S_{i,j}$ into two child non-terminal symbols. It is aimed to concatenate two symbols to match feature channels.

Due to the different views of CT scans, an identical pathology may show various forms, which is defined by an alternative way of decomposing $S_{i,j}$ as the Or rule, denoted by $S_{i,j}(0)|S_{i,j}(1)|\dots|S_{i,j}(j-i)|Ter_{i,j}$.

Furthermore, we note the generated products from larger image regions with high-level semantic features appear more natural than those based on the same regions' local semantic features. However, the local semantic features contain more functional semantic constituents, which are the key for a functional-consistency generator. See from Section 3.4, the corresponding sub-image of an OR node is determined by a selectable pattern of assembling some activated units on a CNN-like filter. Though the different patterns are pointed out with an identical semantic label, the sub-images of its followed And node show various high-level semantic features. A consistent limitation is considered as hierarchical attribute distances between the topology of the sub-images from the referenced image and the generated image. The related attributes are calculated as follows:

Unary attribute: For each pairwise local topologies G_p and G_Q , they respectively contains the sub-sentences of the nodes v_a and v_α , and w_a^p and w_α^q denote their word sets. By fulfilling with dummy nodes, G_p and G_Q are isometric. We define the cosine distance as their unary attribute, and it can be calculated by

$$U_{w_a^p, w_\alpha^q} = \frac{w_a^p \cdot w_\alpha^q}{|w_a^p|_2 \cdot |w_\alpha^q|_2} \quad (3)$$

Pairwise attribute: The distance between the edges from a couple of nodes is defined as a pairwise attribute. For a pair of given edges e_{ab} and $e_{\alpha\beta}$, the detail calculation can be denoted as

$$E_{e_{ab}^p, e_{\alpha\beta}^q} = \exp(-|\lambda_{ab}^p e_{ab}^p - \lambda_{\alpha\beta}^q e_{\alpha\beta}^q|^2 / \epsilon^2) \quad (4)$$

where e_{ab}^p and $e_{\alpha\beta}^q$ are the edges from G_p and G_q , and λ_{ab}^p and $\lambda_{\alpha\beta}^q$ are two constants to label their edge types, respectively. ϵ^2 is the scaling factor between two topologies.

Local topological attribute: In the final stage of similarity calculation, we integrate both unary and pairwise attributes into the local topological attribute for template selection, which can be formulated in the cubic form:

$$\begin{aligned} \Omega(G_p, G_Q) &= \{ \langle E_p, E_Q \rangle + \sum_{v_a v_\alpha} z_{v_a v_\alpha} U_{v_a v_\alpha} \\ &\quad + \sum_{v_a \neq v_b, v_\alpha \neq v_\beta} z_{v_a v_b} z_{v_\alpha v_\beta} E_{e_{ab}^p, e_{\alpha\beta}^q} \} \\ \text{s.t. } Z \cdot 1_{K_p \times 1} &\leq 1_{K_Q \times 1}, \quad Z^T \cdot 1_{K_Q \times 1} \leq 1_{K_p \times 1} \end{aligned} \quad (5)$$

where E_p and E_Q are respectively the edge matrices of the topological graphs G_p and G_Q , and $\langle \cdot \rangle$ denotes their GED. Z is a matrix of the node correspondence, i.e., $z_{v_b v_\beta} = 1$ means the b th node of the graph G_p corresponds to the β -th node of the graph G_Q . The node number of two graphs is denoted by K_p & K_Q .

4.3. Functional-realistic discriminator

Based on grammar-guided network generation, the input LR and HR images are abstracted into the sentences of N words. Due to grammar interpretability, a sentence's construction process can be backtracked as a tree topology with several Terminal nodes, nodes, and nodes. For pairwise trees between a LR image and its SR counterpart, their topological graphs have local similarity, which can be exploited as the criteria to select the subsentences of existing HR images as the templates of its SR generation. The entire generative process can be considered a graph editing that sockets the robust sub-sentences from the networks of HR images into the word/sub-sentence of the LR image network. Naturally, graph edit distance (GED) is used to calculate the graph similarity for template selection. More specifically, for a sequential input map, we track each graph edit process to form a candidate for tree production rules. Compared with a real parse tree from EMR, the candidate is considered as an isomorphic tree with dummy nodes from the root node to the leaf nodes. Many local semantic words/parts replace the dummy nodes to offer enough OR nodes for the tree generation in the generating step. Some newborn subtrees break the real parse tree's syntax, which can guide the discriminator to learn the correct syntax patterns generated from a set of production rules. The Child-sum Tree-LSTM [26] is used as a model of the discriminator, which is formulated as follows:

$$\begin{aligned} \tilde{h}_j &= \sum_{k \in Ch(j)} h_k, \\ i_j &= \sigma(\mathbf{W}^{(i)}x_j + \mathbf{U}^{(i)}\tilde{h}_j + b^{(i)}), \\ f_{jk} &= \sigma(\mathbf{W}^{(f)}x_j + \mathbf{U}^{(f)}h_k + b^{(f)}), \\ o_j &= \sigma(\mathbf{W}^{(o)}x_j + \mathbf{U}^{(o)}\tilde{h}_j + b^{(o)}), \\ u_j &= \tanh(\mathbf{W}^{(u)}x_j + \mathbf{U}^{(u)}\tilde{h}_j + b^{(u)}), \\ c_j &= i_j \odot u_j + \sum_{k \in Ch(j)} f_{jk} \odot c_k, \\ h_j &= o_j \odot \tanh(c_j), \end{aligned} \quad (6)$$

where h_k is the hidden state of node $k \in Chj$ which is the k th child node of the node j . When node j locates in the root of the given tree r , the output layer can obtain a probability of the

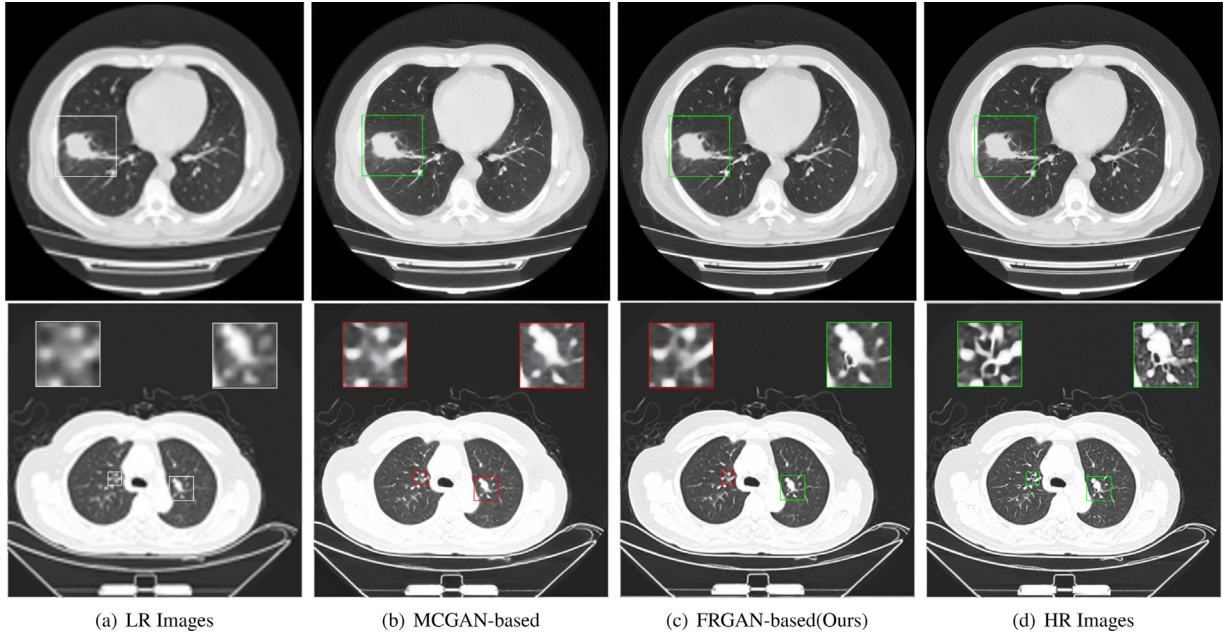


Fig. 3. Performance comparison among (a) LR images, (b) MCGAN-based SR images, (c) FRGAN-based SR images after 50 training epochs. (d) HR images.

encoded tree \mathcal{G}_j being a real instance, activated by a sigmoid unit following with a fully-connected layer as $\Psi = \text{sigmoid}(\mathbf{W}^c h_r + b^c)$. Meanwhile, the replaced OR nodes bring a sequence of the log logistic loss output from CNN as a functional loss for the SR image.

$$I_{\text{Func}}^{\text{SR}} = \mathbb{E}_{\mathcal{G}(I^{\text{LR}}) \in I^{\text{HR}}} [I(\hat{\mathcal{G}}_{\mathcal{G}(I^{\text{LR}})}, y_{I^{\text{HR}}}^*)] \quad (7)$$

where $\hat{\mathcal{G}}_{\mathcal{G}(I^{\text{LR}})}$ and $y_{I^{\text{HR}}}^*$ are the predicted label and ground-truth label of its target HR image. Within the context and adversarial losses [3], the loss function of functional-realism can be expressed as:

$$I_{\text{SR}} = I_{\text{Cont}}^{\text{SR}} + 10^{-3} I_{\text{Gen}}^{\text{SR}} + I_{\text{Func}}^{\text{SR}} \quad (8)$$

5. Experiments

In this section, we first introduce the experimental dataset and setup. By using the parameter configure, we show the success rate of fabricating the HR counterparts, and the pulmonary nodule detection results, then compare them with those of the state-of-the-art methods. To facilitate the comparison, we dichotomize our result to form the results pulmonary nodule detection for the early-stage subsets and the entire merged dataset. All the experiments were conducted on PyTorch deep learning architecture with 2 TITAN X Pascal GPUs.

5.1. Datasets

In the experiments, we merge a benchmark dataset of LUNA16 and self-collected dataset. The LUNA16 dataset was derived from the Lung Image Database Consortium (LIDC) dataset, which removed part of the CT scans with the thickness ≥ 2.5 mm. In the remained set of 888 scans, there exists 1186 ground truth nodules. The organizers provided a set of pre-screened candidates with the positive/negative ratio of 753, 418/1, 557. The self-collected dataset covers 241 NSCLC (Non-Small-Cell Lung Cancer) nodules of the 146 scans from the lung cancer patients treated in our partner hospital from Oct. 2015 to Oct.2017. The CT slices of all the self-collected scans are imaging on a 2D plan with 512×512 resolution, and their thickness is 0.625 mm. Its annotations follows the LIDC annotation criteria for the detail

Table 1

Descriptions of different size 3D Nodules: MDN = Mean Diameters of Nodules, NN (L/A) = Numbers of Nodules from LUNA16 (L)/AH-LUTCM (A) datasets, MAS = Mean Actual Sizes.

Level	MDN (mm)	NN (L/A)	MAS (mm ³)
1	(3, 6]	502/41	129.33
2	(6, 9]	348/54	482.50
3	(9, 12]	125/45	1356.50
4	(12, 15]	84/36	2561.50
5	(15, 18]	53/35	4237.06
6	(18, 21]	33/23	5762.08
7	(21, 24]	29/4	7472.14
8	(24, ∞)	12/3	9755.20

Table 2

Quantitative comparison between the discriminative results of SR image source based on MCGAN and FRGAN.

Methods	20 epochs	50 epochs	100 epochs	200 epochs
MCGAN	33.6%	62.1%	72.6%	83.5%
FRGAN	52.3%	78.4%	83.3%	91.5%

nodule patterns including multi-graded labels of texture, subtlety, spiculation, sphericity, calcification, margin, malignancy, lobulation, and internal structure. Table 1 shows a distribution of nodule sizes, we partition the entire dataset into 8 subsets with 3.0 mm step. We note that most of the nodules are concentrated in the subset 1–3. Due to the smaller sizes, these nodules are easier to be misdiagnosed or fails to be examined.

5.2. Network architecture setup

For data preparation, the merged dataset can be divided into the HR image dataset and the LR image dataset. The former covers the subsets of Level 2–8 in Table 1. The latter is composed of two parts: (1) The original LR images in the subsets of Level 1 in Table 1; (2) The downsampled counterparts of the HR images. Both parts of the nodules are cropped/resize with a size of $32 \times 32 \times 10$. Based on the 9 semantic nodule patterns, we predefine the groups of the feature channels. The AOGNets-12M model specification is exploited to construct the sub-sentences.

Table 3

Multi-classification results of the malignancy of SR pulmonary nodules by FRGAN discriminator.

	Benign	Benign-like	Review	Malignant-like	Malignant	Total	ACC (%)
Benign(Predicted)	1916	397	23	5	1	2429	81.81
Benign-like(Predicted)	216	1475	232	10	2	1848	76.23
Review(Predicted)	75	227	734	61	4	1101	66.67
Malignant-like(Predicted)	18	25	336	1355	14	1748	77.52
Malignant(Predicted)	2	8	76	546	3442	4074	84.49
Total	2227	2132	1401	1977	3463	11200	77.34

Table 4

Performance comparisons of pulmonary nodule detection with various DA schemes.

Schemes	Datasets	1/8	1/4	1/2	1	2	4	8	Score (CPM)
Multi-View CNNs	543 real LR images	0.349	0.469	0.584	0.617	0.637	0.665	0.668	0.569
	+1 × 2D FRGAN-based DA	0.391	0.512	0.627	0.704	0.736	0.743	0.745	0.636
	+1 × 2D MCGAN-based DA	0.376	0.491	0.603	0.684	0.706	0.711	0.712	0.612
	+2 × 2D FRGAN-based DA	0.384	0.505	0.612	0.694	0.732	0.737	0.738	0.629
CUMedVis	543 real LR images	0.536	0.669	0.784	0.812	0.827	0.833	0.835	0.756
	+1 × 3D FRGAN-based DA	0.556	0.694	0.803	0.844	0.856	0.863	0.865	0.783
	+1 × 3D MCGAN-based DA	0.539	0.672	0.792	0.823	0.831	0.838	0.841	0.762
	+2 × 3D FRGAN-based DA	0.557	0.696	0.802	0.837	0.851	0.856	0.858	0.779

Table 5

Performance comparisons of false positive reduction between other schemes and ours on the LUNA16 dataset (L) and the merged dataset (L&A)

Schemes	Datasets	1/8	1/4	1/2	1	2	4	8	Score (CPM)
Multi-View CNNs	L&A	0.549	0.669	0.784	0.883	0.906	0.917	0.928	0.809
	L	0.676	0.761	0.822	0.884	0.936	0.943	0.956	0.854
CUMedVis	L&A	0.611	0.778	0.854	0.921	0.941	0.957	0.968	0.861
	L	0.702	0.831	0.933	0.971	0.973	0.975	0.975	0.908
MR-Forest	L&A	0.620	0.785	0.862	0.923	0.943	0.957	0.967	0.865
	L	0.714	0.827	0.931	0.969	0.975	0.977	0.977	0.910
FRGAN (Ours)	L&A	0.692	0.819	0.893	0.928	0.949	0.956	0.959	0.885
	L	0.729	0.871	0.923	0.958	0.970	0.978	0.982	0.915

A noise box is used to replace the original $32 \times 32 \times 10$ voxels for each LR input which is cropped into $64 \times 64 \times 20$ voxels as VOIs (Volumes of Interest), then The VOIs (Volumes of Interest) are concatenated with the matched sub-sentences of 9 semantic patterns from the network of the HR images accumulated to the corresponding locations. So we uses $64 \times 64 \times 20 \times 10$ inputs to each layer of our generator to fabricate desired HR counterparts in the noise box regions. All the layers are organized into a 3D-Unet [27]-like architecture, including 4-layer encoders and 4-lay decoders with skip connections to effectively capture the context and semantic information of the inputs. In the training step, the epoch sets 500, and each epoch contains 1000 steps with a mini-batch size of 16 and a learning rate 1.0×10^{-4} for Adam optimizer.

5.3. Results

5.3.1. Pulmonary nodules generation

In the paper, there are two goals for the functional structure generator. On the one hand, we use it to generate a set of functional-realistic HR counterparts for the early-stage pulmonary nodules' images. The HR generations could offer a high-quality images for the radiologists' inferences. On the another hand, it can be a data augmentation tools for balancing the ratio of the positive/negative samples to improve the performance of pulmonary nodule detection. Hence, we first show a comparison generative results in Fig. 3.

As shown in Fig. 3, the original LR and HR image groups are listed in Fig. 3a and Fig. 3d. The others are the generative results

based on MCGAN (Multi-Conditional GAN) [28] and the proposed network. We note that our results are obviously superior to those of MCGAN-based. Due to the additional inputs of functional structure, the proposed network can capture context/pathological information faster, and uses these information to generate realistic nodules. For a better comparison, we zoom-in part of the ROIs (2D planes of VOIs on Z-axis), and the white, red and green bounding boxes which cover the initial inputs, the fake/real SR images annotated by the logicale discriminator, respectively. In the first row, both groups of SR images are determined as real images, because the radiologists describe more detail characteristics for the SPN (Solitary Pulmonary Nodule) in its EMR. In contrast, the SR images in the second row report more complex. Its EMR only gives an incomplete description for the left ROI, which offers few semantic feature leading to a worse parse tree. Although the content loss and generative loss can drive the content of SR images close to those of real HR images, the absence of key functional structures are discovered by the discriminator. For another ROI of the group, only the generated image based on FRGAN is determined as real image. Compared with the MCGAN-based SR image, the generated image own a more logical structure of semantic attributes in according with the parse tree from its EMR to faster generate high-quality SR image (Shown in Table 2). The success rate of cheating the discriminator is rocketed by using our generator. Compared with ours, the generator of MCGAN needs more training epochs to boost its generative capacity. Moreover, to further demonstrate the effectiveness of our generator, we use a quantitative comparison of determinative results for 5-class SR images in Table 3.

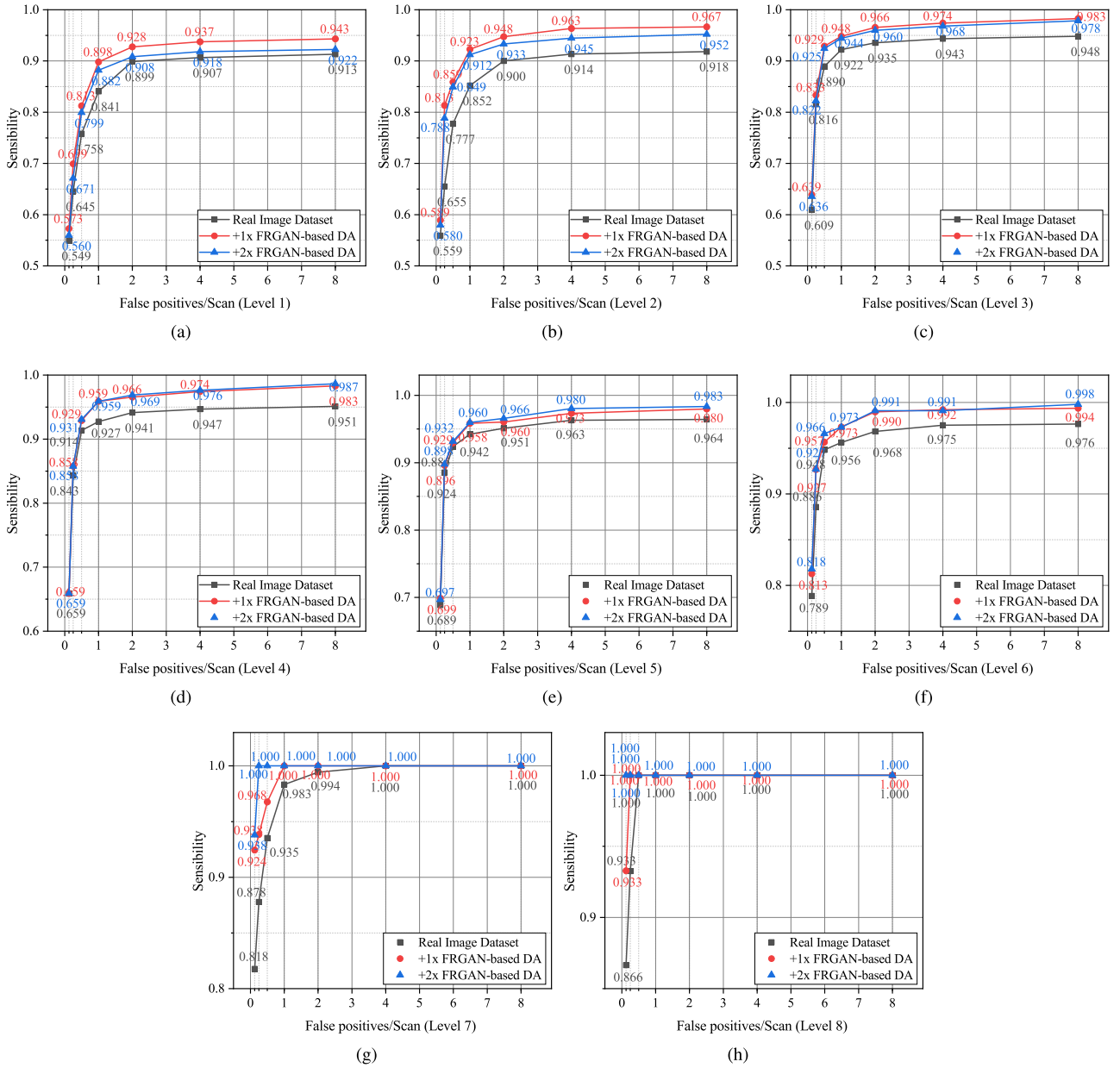


Fig. 4. FROC curves of the architectures with FRGAN-based DA with variant inputs in the different levels.

5.3.2. Pulmonary nodules detection

We employ two metrics to assess the performance of pulmonary nodules detection, including Free Receiver Operating Characteristic (FROC) [29] and Competition Performance Metric (CPM) [30] score. In the section, we design two experiments for the subsets of early-stage nodules and the entire dataset. In the former experiment, we use 543 real volumetric images from the subsets of Level 1 in Table 1 to generate $2\times$ HR images to fed into two CNNs-based detection architecture. A quantitative comparison is shown in Table 4. In this comparison, we select two architectures of the outstanding team in LUNA16 challenge. A brief description for their works is as follow:

Multi-View CNN [17]: a CAD system is based on 2D convolutional networks (ConvNets) to classify the extracted multi-view patches of pulmonary nodules candidates filtrated from three candidate detectors for solid, subsolid, and large nodules. Data augmentation and dropout are applied to avoid model overfitting.

CUMedVis [18]: Inspired by Multi-view CNN, a 3DCNNs based architecture comprises three-level network architectures for contextual information in different resolutions. The prediction probabilities from multi-level models are fused to output the final classification.

Compared with two CNN-based detection architectures, the detection performance of the 3D CNNs exceeds those of the 2D counterparts above 23%. Both the CPM scores of two group are not boosting with more multiples of SR data augmentation, because the limit real data source deepens overfitting. Moreover, we find that it is more conspicuous for the detection performance improvement of 2D CNNs by using data augmentation. The main reason is the less context information of 2D slices in the comparison with the 3D counterparts, and the limited benefit is gained for the context informations of the 3D blocks on Z-axis.

In the subsequent experiment, all the candidates of the merged dataset and their FRGAN-based SR images are fed into CUMedVis.

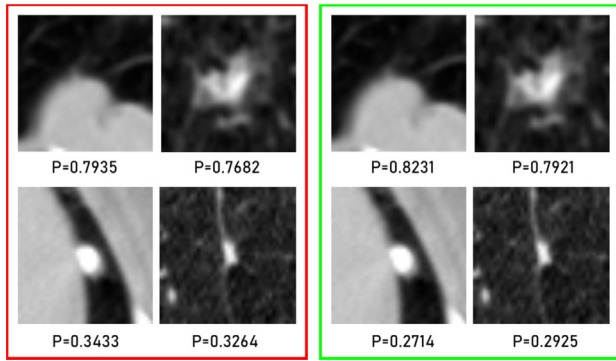


Fig. 5. Representative transverse plane of some annotated nodule and their degrees of suspicion (the p values below the them) figured out by CUMedVis [18] without/with FRGAN-based DA.

To further weaken the imbalance between positive and negative amounts, we properly modify the generative loss to reduce visual artifacts. Afterwards, the phenomenon mentioned in Section 5.3.1, which is motivated to generate more diverse positive candidates. Meanwhile, translation and rotation augmentations are conducted. In total, we obtain 0.22 million samples to train the detection network. 10-cross validation is used for a fair comparison with the other methods. Fig. 4 shows the pulmonary nodule detection results of our architecture on the 8 subsets of the merged dataset.

In a global view, the larger level subsets have higher sensitivities, especially in those results of level 6–8 nodules, the sensitivities have even reach 100% under 8 false positives per scan. The reason has great relation with the multi-pattern functional structure for generate more plenty of positive samples. It first demonstrates that our architecture is competitive for the task of pulmonary nodule detection.

Moreover, we regroup the entire results into 5 subsets refer to their recommended malignancy in original LIDC dataset in Table 3, in which the result of review subset shows the lower accuracy in the comparison with the other four subsets. In our survey for the misclassified samples, these samples are endowed a lower or higher suspected degrees by the proposed architecture, respectively classifying into the malignancy-like and benign-like classes. The similar phenomena are also discovered in the malignancy-like and benign-like subsets. We believe that these are positive, can demonstrate the proposed architecture has a better performance than human subjective inference for malignancy evaluation. The detail discussion can be found in Section 5.4.1.

In Table 5, we list three state-of-the-art methods, including Multi-view CNNs [17], CUMedVis [18], MR-Forest [31]. CPM score is the main evaluation metric in the comparison. Our results located in the top of all the results, whose CPM scores achieve 0.875 and 0.915 on the benchmark dataset and the merged dataset, respectively. It performs better than the first two compared methods, which were advanced in the pulmonary nodule detection track of LUNA16 challenge.

5.4. Discussions

To illustrate the more details and demonstrate the better performance of the proposed architecture, we discuss two significant points in this subsection, including effectiveness of the semantic patterns and effectiveness of the subset balance.

5.4.1. Effectiveness of the semantic patterns

Different from the given probability of malignancy by most of DL-based methods, the proposed architecture can implies a probability between deterioration and stabilization by functional-realism SR base on the semantic patterns. In Fig. 5, we show a simple comparison between the pulmonary nodule detection results of multi-level contextual 3D CNNs (CUMedVis) and ours. The left group contains four representative transverse planes of detection results using CUMedVis (shown in red box of Fig. 5), and the other box (green) arranges four corresponding results figured out by the proposed architecture. By comparing in the degrees of suspicion, we find that the results of our architecture tend to polarization in the range of [0, 1]. The main reason for such differences between the two groups lies in the different methodologies. The 3DCNNs-based methods are always encoding the static visual information via massive 3D convolutional operation. They encourage high-level feature diversity to mine the latent significant information for certain specific detection task. The training set consists of many labeled nodules in static sampling. However, the actual malignancy evaluation is a dynamic process in clinical practice, such as cell doubling time and cell cycle progression. Hence, the results can emerge tiny deflections after integrating the temporal factor into the spatial graph.

5.4.2. Effectiveness of the subset sample balance

In the detection experiments, we use FRGAN-based DA to increase the positive sample amount. After DA, the sample ratio approximate to 1:1 in the merged dataset. However, few studies focus on the subset sample balance in medical data because most of the existing works are binary classification for false/true positives discrimination. In the task of building a grammar structure, each structural block contains multiple pathological patterns. For each pathological pattern, the radiologists set a list of the pre-defined degrees to segment their probability related to the final predicted result. If the subsets exist severe sample imbalance, the pathological patterns related to the absent samples may cause local grammar structure monotony. In Section 5.4.1, we observe the phenomenon of result polarization due to replacing the LR structure with an improper sub-sentence with a higher degree of grammar patterns.

6. Conclusion and future works

This paper proposes a deep generative adversarial network (GAN) architecture based on a deep grammar model, called FRGAN (Functional-Realistic GAN). Different from the other GAN architecture, we focus on AOG-based functional semantic graph construction. First of all, we build a parse tree based on the EMR of each input image for mining the potential production rules. A region proposal network (RPN) is used to recommend the bounding boxes and label their semantics for local features. The functional structure of local pathological images can be hierarchically aggregated with corresponding to different semantic patterns. A novel generator is proposed to collect the functional structure from the semantic graphs of high-resolution images and then sequentially socket them into the low-resolution image graph. The low-quality images and nine potential semantic patterns are fed into the architecture to generate realistic/diverse 64×64 volumetric images placed naturally on its super-resolved counterpart to boost the sensibility in pulmonary nodule detection. Our discriminator learns the source (real/synthetic) and classes (benign/ malignancy) of various images synchronously to ensure a better generative result. Experimental results show that our generator can faster generate more realistic HR images with pathological structure. Based on the proposed generator's robust generative ability, we augment the positive candidates

in the merged dataset to balance the sample ratio. On 543 real LR nodules of the early-stage subset and 1304 nodules of the entire merged dataset, we achieve two CPM scores of 0.783 and 0.915 by using improved CUMedVis with $1\times$ FRGAN-based data augmentation, respectively. In our future work, we are going to explore the latent grammar structure for sub-sentence construction. Moreover, the current architecture is too complicated, which causes higher computational and storage costs. Hence, we will consider exploring more lightweight classification architectures for the detection tasks.

CRedit authorship contribution statement

Hongbo Zhu: Conceptualization, Methodology, Software, Writing- Original Draft, Writing - Review & Editing. **Guangjie Han:** Data curation, Supervision. **Yan Peng:** Project administration. **Wenbo Zhang:** Formal analysis, Visualization. **Chuan Lin:** Software, Validation. **Hai Zhao:** Data curation, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Key R&D Program under Grant 2018YFB1702003, in part by the National Science Foundation of China (NSFC) under Grant 61806048, in part by the Open Program of Neusoft Research of Intelligent Healthcare Technology, Co. Ltd. under Grant NRIHTOP1802, and in part by the Fundamental Research Funds for the Central Universities under Grant N180716019.

Ethical approval

Ethical approval of AH-LUTCM dataset was given by the Institution Review Board (IRB) with the following approval notice number: 20171111CS (KT)-040-01. We appreciate the LUNA16 challenge organizers for providing the dataset.

References

- [1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, CA: Cancer J. Clin. 67 (1) (2017) 7–30, <http://dx.doi.org/10.3322/caac.21387>.
- [2] M.A. Grippi, J.A. Elias, J.A. Fishman, R.M. Kotloff, A.I. Pack, R.M. Senior, M.D. Siegel, *Fishman's Pulmonary Diseases and Disorders*, fifth ed., McGraw-Hill, 2015.
- [3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 105–114, <http://dx.doi.org/10.1109/CVPR.2017.19>.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Adv. Neural Inf. Process. Syst.* 3 (2014).
- [5] J.A. Ferwerda, Three varieties of realism in computer graphics, in: B.E. Rogowitz, T.N. Pappas (Eds.), *Human Vision and Electronic Imaging VIII*, vol. 5007, SPIE, 2003, pp. 290–297, <http://dx.doi.org/10.1117/12.473899>.
- [6] Q. Zhang, Y.N. Wu, S.C. Zhu, Mining and-or graphs for graph matching and object discovery, in: IEEE International Conference on Computer Vision, 2015, pp. 55–63.
- [7] X. Li, T. Wu, X. Song, H. Krim, AOGNets: Deep AND-OR grammar networks for visual recognition, 2017, CoRR [arXiv:1711.05847](https://arxiv.org/abs/1711.05847).
- [8] I.R.S. Valente, P.C. Cortez, E.C. Neto, J.M. Soares, V.H.C. de Albuquerque, J. ao Manuel R.S. Tavares, Automatic 3D pulmonary nodule detection in CT images: A survey, *Comput. Methods Programs Biomed.* 124 (2016) 91–107, <http://dx.doi.org/10.1016/j.cmpb.2015.10.006>.
- [9] K. Murphy, B.V. Ginneken, A.M.R. Schilham, B.J.D. Hoop, H.A. Gietema, M. Prokop, A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification, *Med. Image Anal.* 13 (5) (2009) 757–770.
- [10] T. Messay, H.S.K. Rogers, A new computationally efficient CAD system for pulmonary nodule detection in CT imagery, *Med. Image Anal.* 14 (3) (2010) 390–406.
- [11] J. Colin, E.M.V. Rikxoort, T. Thorsten, S. Ernst Th, P.A.D. Jong, K. Jan-Martin, O. Matthijs, H.J.D. Koning, P. Mathias, S.P. Cornelia, Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images, *Med. Image Anal.* 18 (2) (2014) 374–384.
- [12] A. El-Baz, M. Nitzken, E. Vanbogaert, G.L. Gimel'Farb, R. Falk, M.A. El-Ghar, A novel shape-based diagnostic approach for early diagnosis of lung nodules, *Chest* 140 (4) (2011) 655A, 655B–655A, 655B.
- [13] Z. Chen, H. Strange, E.R. Denton, A. Oliver, C. Boggis, R. Zwigelaar, Topological modeling and classification of mammographic microcalcification clusters, *IEEE Trans. Biomed. Eng.* 62 (4) (2015) 1203–1214.
- [14] S. Soltaninejad, M. Keshani, F. Tajeripour, Lung nodule detection by KNN classifier and active contour modelling and 3D visualization, in: The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), 2012, pp. 440–445.
- [15] A.M. Santos, A.O. de Carvalho Filho, A.C. Silva, A.C. de Paiva, R.A. Nunes, M. Gattass, Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM, *Eng. Appl. Artif. Intell.* 36 (2014) 27–39.
- [16] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R.M. Summers, A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations, *Med. Image Comput. Comput. Assist. Interv.* 17 (1) (2014) 520–527.
- [17] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sánchez, B. van Ginneken, Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1160–1169.
- [18] Q. Dou, H. Chen, L. Yu, J. Qin, P.A. Heng, Multi-level contextual 3D CNNs for false positive reduction in pulmonary nodule detection, *IEEE Trans. Biomed. Eng.* PP (99) (2017) 1–1.
- [19] J. Ding, A. Li, Z. Hu, L. Wang, Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks, CoRR [abs/1706.04303](https://arxiv.org/abs/1706.04303) (2017) URL [arXiv:1706.04303](https://arxiv.org/abs/1706.04303).
- [20] H. Tang, D.R. Kim, X. Xie, Automated pulmonary nodule detection using 3D deep convolutional neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 523–526, <http://dx.doi.org/10.1109/ISBI.2018.8363630>.
- [21] J. Jebadurai, J.D. Peter, Super-resolution of retinal images using multi-kernel SVR for IoT healthcare applications, *Future Gener. Comput. Syst.* 83 (2018) 338–346, <http://dx.doi.org/10.1016/j.future.2018.01.058>, URL <http://www.sciencedirect.com/science/article/pii/S0167739X17322136>.
- [22] L. Chen, X. Yang, G. Jeon, M. Anisetti, K. Liu, A trusted medical image super-resolution method based on feedback adaptive weighted dense network, *Artif. Intell. Med.* 106 (2020) 101857, <http://dx.doi.org/10.1016/j.artmed.2020.101857>, URL <http://www.sciencedirect.com/science/article/pii/S0933365719310073>.
- [23] S.K. Baliarsingh, S. Vipsita, K. Muhammad, S. Bakshi, Analysis of high-dimensional biomedical data using an evolutionary multi-objective emperor penguin optimizer, *Swarm Evol. Comput.* 48 (2019) 262–273, <http://dx.doi.org/10.1016/j.swevo.2019.04.010>, URL <http://www.sciencedirect.com/science/article/pii/S2210650219301580>.
- [24] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, J. Cheng, Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images, *IEEE Trans. Med. Imaging* 36 (3) (2017) 802–814.
- [25] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1137–1149, <http://dx.doi.org/10.1109/TPAMI.2016.2577031>.
- [26] X. Liu, X. Kong, L. Liu, K. Chiang, TreeGAN: Syntax-aware sequence generation with generative adversarial networks, 2018, CoRR [arXiv:1808.07582](https://arxiv.org/abs/1808.07582).
- [27] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, 2015, CoRR [arXiv:1505.04597](https://arxiv.org/abs/1505.04597).
- [28] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, H. Nakayama, Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection, in: 2019 International Conference on 3D Vision (3DV), 2019, pp. 729–737, <http://dx.doi.org/10.1109/3DV.2019.00085>.
- [29] A.A.A. Setio, A. Traverso, T. de Bel, M.S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M.E. Fantacci, B. Geurts, R. van der Gugten, P.A. Heng, B. Jansen, M.M. de Kaste, V. Kotov, J.Y.-H. Lin, J.T. Manders, A. Sónora-Mengana, J.C. García-Naranjo, E. Papavasileiou, M. Prokop, M.

Saletta, C.M. Schaefer-Prokop, E.T. Scholten, L. Scholten, M.M. Snoeren, E.L. Torres, J. Vandemeulebroucke, N. Walasek, G.C. Zuidhof, B. van Ginneken, C. Jacobs, Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge, *Med. Image Anal.* 42 (2017) 1–13, <http://dx.doi.org/10.1016/j.media.2017.06.015>, URL <http://www.sciencedirect.com/science/article/pii/S1361841517301020>.

- [30] M. Niemeijer, M. Loog, M.D. Abramoff, M.A. Viergever, M. Prokop, B. van Ginneken, On combining computer-aided detection systems, *IEEE Trans. Med. Imaging* 30 (2) (2011) 215–223, <http://dx.doi.org/10.1109/TMI.2010.2072789>.

- [31] H. Zhu, H. Zhao, C. Song, Z. Bian, Y. Bi, T. Liu, X. He, D. Yang, W. Cai, MR-Forest: A deep decision framework for false positive reduction in pulmonary nodule detection, *IEEE J. Biomed. Health Inf.* (2019) 1–1, <http://dx.doi.org/10.1109/JBHI.2019.2947506>.



Hongbo Zhu received the B.Sc., M.Eng., Ph.D. degree from Northeastern University, Shenyang, China, in 2009, 2012, and 2020 respectively. He is an assistant professor with School of Information Science and Engineering, Shenyang Ligong University. His research interests include medical image computing and deep learning.



Guangjie Han (hanguangjie@gmail.com) is currently a Distinguished Professor of the Key Laboratory for Ubiquitous Network and Service Software of Liaoning province, School of Software, Dalian University of Technology, Dalian, 116024, China and a Professor with the Department of Information and Communication System, Hohai University, Changzhou, China. He received the Ph.D. degree from Northeastern University, Shenyang, China, in 2004. In February 2008, he finished his work as a Postdoctoral Researcher with the Department of Computer Science, Chonnam National University, Gwangju, Korea. From October 2010 to October 2011, he was a Visiting Research Scholar with Osaka University, Suita, Japan. From January 2017 to February 2017, he was a Visiting Professor with City University of Hong Kong, China. His current research interests include Internet of Things, Industrial Internet, Machine Learning and Artificial Intelligence, Mobile Computing, Security and Privacy. Dr. Han has over 400 peer-reviewed journal and conference papers, in addition to 140 granted and pending patents. Currently, his H-index is 41 and i10-index is 135 in Google Citation (Google Scholar). Total citation of his papers by other people is more than 6500+ times.

Dr. Han is a Fellow of the UK Institution of Engineering and Technology (FIET). He has served on the Editorial Boards of up to 10 international journals, including the IEEE Network, IEEE Systems, IEEE/CCA JAS, IEEE ACCESS, Telecommunication Systems, etc. He has guest edited a number of special issues in IEEE Journals and Magazines, including the IEEE JSAC, IEEE Communications, IEEE Wireless Communications, IEEE Transactions on Industrial Informatics, Computer Networks, etc. Dr. Han has also served as chairs of organizing and technical committees of many international conferences. He had been awarded 2020 IEEE Systems Journal Annual Best Paper Award, 2017–2019 IEEE ACCESS Outstanding Associate Editor Award. He is a Senior Member of IEEE.



Yan Peng received the Ph.D. degrees in Pattern Recognition and Intelligent Systems from Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, in 2009. Now, she is a professor at Shanghai University as dean of Research Institute of USV Engineering, Shanghai, China. Her current research interests include modeling and control of unmanned surface vehicles, field robotics and locomotion system.



Wenbo Zhang is currently a professor of School of Information Science & Engineering, Shenyang Ligong University, China. He received his Ph.D. in Computer science at Northeastern University, China, in March 2006. He has published over 100 papers in related international conferences and journals. He has served in the editorial board of up to 10 journals, including Chinese Journal of Electronics and Journal of Astronautics. He had been awarded the ICINIS 2011 Best Paper Awards and up to 9 Science and Technology Awards including the National Science and Technology Progress Award and Youth Science and Technology Awards from China Ordnance Society. His current research interests are Internet of Things, Wireless Sensor Networks, Satellite networks, Embedded systems.



software-defined networking.

Chuan Lin is currently a Postdoctoral Researcher with the School of Software, Dalian University of Technology, Dalian, China. He received the B.S. degree in Computer Science and Technology from Liaoning University, Shenyang, China in 2011, the M.S. degree in Computer Science and Technology from Northeastern University, Shenyang, China in 2013, and the Ph.D. degree in computer architecture in 2018. He was a visiting Ph.D. student with Network Architecture and Services group, Delft University of Technology, Delft, Netherlands during 2015–2017. His research interests include UWSNs,



Hai Zhao received the B.S. degree in electrical engineering from Dalian Maritime University, Dalian, China, in 1982, and the M.S. and Ph.D. degrees in computer science from Northeastern University, Shenyang, China, in 1987 and 1995, respectively. He is currently a Professor with the School of Computer Science and Engineering, Northeastern University. His current research interests include embedded internet technology, wireless sensor networks, body area networks, data and information fusion, computing intelligence, and image processing.