Large Residual Multiple View 3D CNN for False Positive Reduction in Pulmonary Nodule Detection

5 authors, including:

Anton Dobrenkii
Innopolis University
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Ramil Kuleev
Innopolis University
6 PUBLICATIONS **1** CITATION

SEE PROFILE

Adin Ramirez Rivera
University of Campinas
24 PUBLICATIONS **307** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   LiberaCAD View project

Large Residual Multiple View 3D CNN for False Positive Reduction in Pulmonary Nodule Detection

Anton Dobrenkii, Ramil Kuleev, Adil Khan
Department of Computer Science
Innopolis University
Innopolis, Russia
Email: a.dobrenkiy, r.kuleev, a.khan@innopolis.ru

Adin Ramirez Rivera
Institute of Computing
University of Campinas
Campinas, Brazil
Email: adin@ic.unicamp.br

Asad Masood Khattak
College of Technological Innovation
Zayed University
Abu Dhabi, UAE
Email: asad.khattak@zu.ac.ae

Abstract—**Pulmonary nodules detection play a significant role in the early detection and treatment of lung cancer. False positive reduction is the one of the major parts of pulmonary nodules detection systems. In this study a novel method aimed at recognizing real pulmonary nodule among a large group of candidates was proposed. The method consists of three steps: appropriate receptive field selection, feature extraction and a strategy for high level feature fusion and classification. The dataset consists of 888 patient's chest volume low dose computer tomography (LDCT) scans, selected from publicly available LIDC-IDRI dataset. This dataset was marked by LUNA16 challenge organizers resulting in 1186 nodules. Trivial data augmentation and dropout were applied in order to avoid overfitting. Our method achieved high competition performance metric (CPM) of 0.735 and sensitivities of 78.8% and 83.9% at 1 and 4 false positives per scan, respectively. This study is also accompanied by detailed descriptions and results overview in comparison with the state of the art solutions.**

Keywords—*pulmonary nodule detection, lung cancer, false positive reduction, 3D convolution neural networks, residual learning, deep learning.*

I. INTRODUCTION

Nowadays pulmonary nodules detection still requires high carefulness from a doctor in order to perform a proper examination and analysis of volumetric thoracic Computed Tomography (CT). Nodules are actually spots inside the lung volume caused by difference in density. They vary in size from 3 mm up to 3 cm in diameter. Motes larger than 3 cm, are considered as a lung mass instead of nodules. Relatively small size of nodules and the fact that the environment around them often varies explains the complexity of the early detection of lung cancer [1]. For example Fig. I presents pairs of true pulmonary nodules and false positives with similar morphology. Usually to become visible on frontal chest X-ray, nodules need to grow up to 1 cm. However, with recent advancement in technology, screening with low-dose spiral computed tomography (LDCT) can now detect nodules of up to 3 mm in diameters. This leads to a reduction in lung cancer mortality by 20% in comparison with standard frontal chest X-ray. [2]–[4].

For the last two decades researchers have been actively studying approaches that can automatically detect pulmonary nodules in computed tomography images. In general, problem of pulmonary nodules detection can be divided into two tasks: candidates detection and further false positive reduction. In candidates detection the number of candidates are extracted
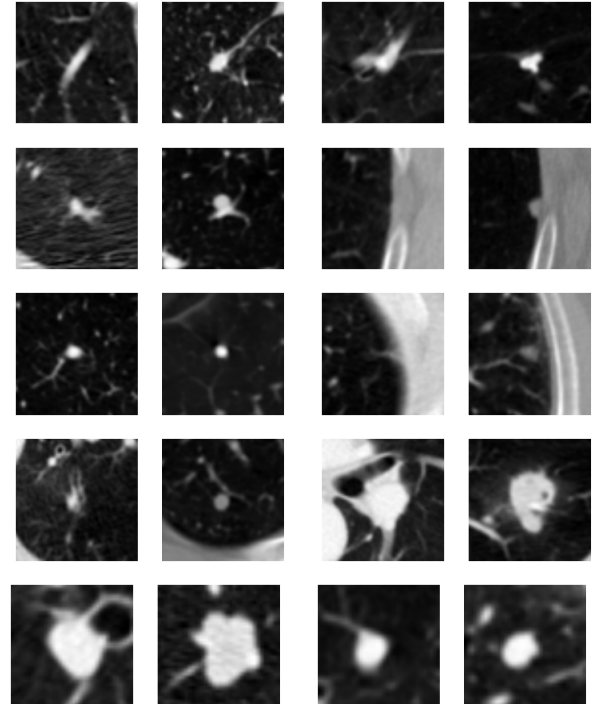

Fig. 1. Here you can see pairs of horizontal views of pictures with pulmonary noodles on the right and false positives on the left provided by LUNA16.

using different criteria e.g., intensity distribution or tissues separation via Hausfield Units thresholds. In false positive reduction discriminative methods are developed in order to lower the amount of candidates.

Major progress in dataset collection for comparative analysis of performance evaluation has been achieved due to the medical image analysis challenge LUNA16 [5] held in conjunction with ISBI 2016. It also provides an evaluation framework for automatic nodule detection. The detailed description of the dataset can be found in section II. Under the competitive ISBI LUNA16 program, a number of novel methods have been proposed for both candidates detection and false positive reduction tasks. The most notable results are listed in the LUNA16 report [5].

The candidates selection is often performed based on local intensity distribution due to the features of nodule appearance:

it may be modeled as an ellipsoid with a radial intensity decreasing. How it was done by Maxine Tan et al. [6] via merging clusters generated by the minima of divergence of normalized gradient taken over CT scan and by dot and line 3D selective enhancement filters which was also applied over CT scan [7].

In the false positive reduction track two main approaches have been use. The one that took the first place is proposed by by Dou et al. [8]. In this work three multi-level contextual 3D convolutional neural networks were employed, with an aim of handling different scale features of pulmonary nodules. Those networks were trained separately and then merged via a fusion function that was a weighted sum of predictors probability output.

Another approach was provided by A. Setio et al. [9], which works by analyzing 2D planes and despite of the 3D nature of nodule appearance, this approach showed a comparative results. The researchers developed false positive reduction method constituting a combination of committee-fusion, late-fusion and mixed-fusion of 2D CNNs acting over 2D patches that were extracted using nine views of a volumetric object.

In the LUNA16 report, it was also shown that the combination of the multiple algorithms showed better performance than the use of a single algorithm for both the false positive reduction and candidates detection tasks. However, the problem is still not fully solved, and needs further investigation. This was the motivator to implement novel false positive reduction algorithm with the method that combines the strengths of multiple views [9] with the advantages of 3D multi-scale approach in order to add it to the ensemble of algorithms.

II. METHODOLOGY

In this study we developed a novel neural network architecture with an aim of combining benefits from multiple views approach with multi-scale one. Such effect was achieved by selecting appropriate receptive fields which cover multiple views and merging strategy due to which multi-scale information was handled. To ensure robustness simple ensemble technique was employed. It takes the average from two predictions made from the same input, mirrored along profile view (due to relative symmetry between left and right lungs).

A. Convolution Neural Networks

In this study technique of spatial information extraction and aggregation such as convolutional network has been employed, and actively used over volumetric data. The convolution is a binary operator, acting on f and φ as follow:

$$(f * \varphi)(t) = \int_\Omega f(\tau)\varphi(t - \tau)d\tau. \tag{1}$$

With this definition basic convolutional neural network (CNN) with N layers and vector $l = (l_1, \quad ..., \quad l_N)$, where each i-th layer of CNN consists of l_i components, can be defined as $\mathcal{C}_l \equiv \mathcal{L}_N$, with:

$$\mathcal{L}_{top} : \mathcal{L}_N \times \mathcal{F}(\mathcal{I}, \xi) \to \mathbb{R}, \tag{2}$$
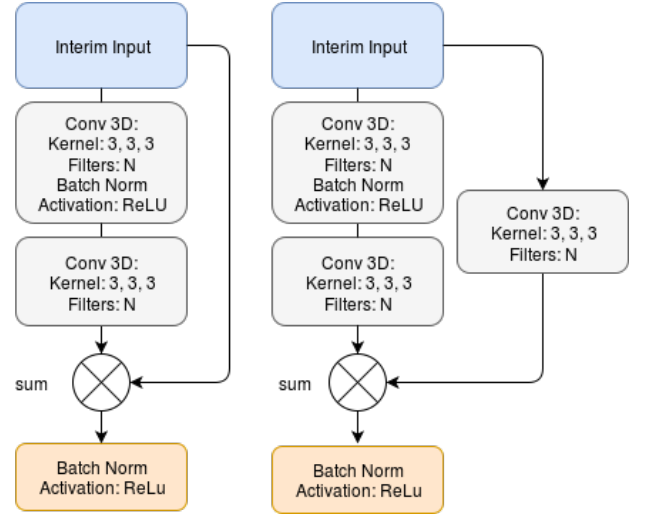

Fig. 2. Possible configurations of residual units due to limitations described above. The first architecture means that W_s is an identical tensor (in term of convolution). The second configuration includes non-identical shortcut function $\mathcal{S}$.

$$\mathcal{L}_1 \equiv \mathcal{I}, \tag{3}$$

$$\mathcal{L}_{k+1} = \{\sigma(\sum_{j=1}^{|\mathcal{L}_k|} f_i * \mathcal{L}_{k_j} + b_i)\}_{i=1}^{l_{k+1}}. \tag{4}$$

Where $\mathcal{I}$ is the given input signal or function, defined over $Span(\Omega)$ and Ω is a compact. $\mathcal{F}(\mathcal{I}, \xi)$ is a target function (in general also dependendable on some stochastic variable ξ), which is going to be approximated via $\mathcal{L}_{|N|}$. The function $\mathcal{L}_{top}$ is the discrepancy measure of approximation and often called loss function. Some non-linear element-wise function σ, along with f_i and b_i functions should be selected in order to minimize the discrepancy.

B. Receptive Fields Selection

The importance of receptive fields is caused by the distinction made in the size and shape of a nodules. Influence of nodules with the minimum size of 3mm in diameter may be fully ignored by selecting huge receptive fields. On the other hand, nodules with diameter of 3 cm can't be fully covered by small receptive fields. Due to this the decision was made to use a combination made of three intersected parallelepipeds with it's largest base parallel to the frontal, horizontal and side views respectively. Parallelepiped's largest base was selected to cover all kinds of nodules in a proper 2D view. In this way, horizontal aligned parallelepiped receptive field base's side was set to 42 units. The same operation for frontal and profile aligned parallelepipeds provide base's side 36 units. Width of this parallelepipeds was selected to be 18 units for horizontal view and 23 for frontal and profile views. Width values selection was based on volume of it's intersection which is chosen in order to cover 85% of nodules appearance. With respect to the unit space (0.9 mm, 0.7 mm, 0.7 mm). Unit space will be discussed further in Preprocessing subsection.

C. High Level Feature Extraction and Merging

Since multiple view ability for the model has been provided via multiple receptive fields, oriented along the three principal views, we needed a feature extraction method to combine information from them. For this purpose high level features [10] were extracted from each of the receptive fields and then fused together by weighted sum. In order to do this fusion, a deep neural network architecture was developed. To escape the lack of gradient vanishing in deep network, described by Xavier Glorot et al. [11] residual units have been employed. Residual units were proposed by Kaiming He et al. in deep residual learning [12] and were successfully generalized over bunch of tasks including volumetric brain MRI segmentation, by Hao Chen et al. [13] and 3D brain MRI classification, by Sergey K. et al. [14]. The main idea of residual learning is based on the approximation of the residual functions instead of the original one, and generally may be written as:

$$\mathcal{Y} - \mathcal{F}(\mathcal{X}, \theta_f) = \mathcal{F}(\mathcal{X}, \theta) \tag{5}$$

Where $\mathcal{X}$ and $\mathcal{Y}$ are sampled random variables from some joint distribution $\mathbb{P}(\mathcal{Y}, \mathcal{X})$, and θ is a set of trainable parameters. $\mathcal{F}(\cdot)$, $\mathcal{S}(\cdot)$ is the residual mapping and shortcut functions respectively, which is usually defined as stacked convolution operations alternating with non-linear activation. In our approach $\theta \equiv \{W, b, \mu_{\mathcal{B}}, \sigma_{\mathcal{B}}^2\}_i$ and $\theta_s \equiv \{W_s, b_s\}$ also residual mapping function were defined as follow:

$$\mathcal{F}(\mathcal{X}, \theta) = W_{|\theta|} * \mathcal{X}_{|\theta|}, \tag{6a}$$

$$\mathcal{X}_i = a[\mathcal{N}orm(W_i * \mathcal{X}_{i-1} + b_i)], \tag{6b}$$

$$\mathcal{S}(\mathcal{X}, \theta_s) = W_s * \mathcal{X}. \tag{6c}$$

Where a is an activation function and $\mathcal{N}orm(\cdot)$ is:

$$\mathcal{N}orm(x) = \frac{x - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}. \tag{7}$$

Equation 7 represent the mini-batch normalization technique provided by Sergey I. et al. and aimed at speed up training process of a deep networks by reducing internal covariate shift [15], in which $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}^2$ are mini-batch's mean and variance respectively. Also ϵ is a positive smoothing constant. As it was stated by Kaiming He et al. in deep residual learning the use of non-identity tensor W_s in a shortcut only makes sense if it's a paired function $\mathcal{F}: (\mathcal{X}, \theta) \mapsto \mathcal{X}'$ leads to: $dom(\mathcal{X}) \neq dom(\mathcal{X}')$ [12]. Due to this statement and with empirically fixed $|\theta| = 2$ residual units result in two possible configuration presented on Fig. 2

D. Preprocessing

In order to handle different slice thickness resampling to the unit space was performed. The unit space was selected as a tradeoff between minimization of data corruption due to resampling operation and maximization of nodule size. To satisfy that requirement new spacing was set to $(0.9\,\text{mm}, 0.7\,\text{mm}, 0.7\,\text{mm})$ as a shape of voxel. To reduce impact of ribs, scan's intensity was clipped in range from -1000 up to 400 Housfield Unit and then normalized to the range of [0, 1].

Since the dataset has a significant skewness: 1 nodule against 460 false positives, the majority class under-sampling
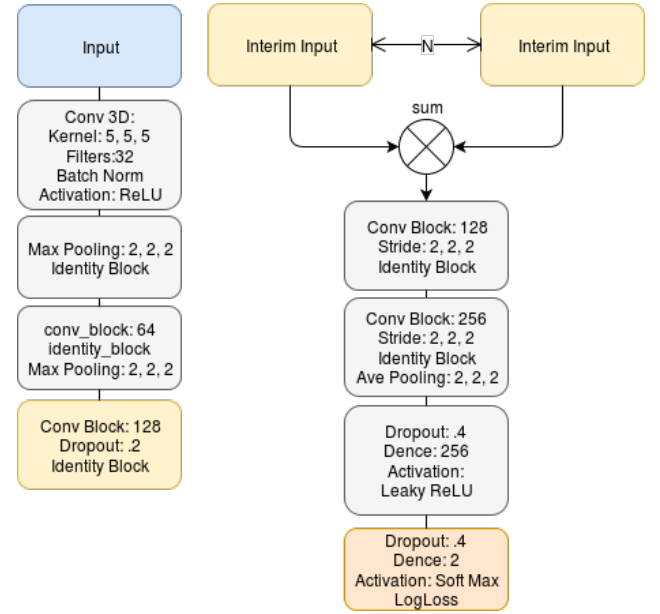

Fig. 3. The architecture on the left represents the high level features extractor, which processes each of the receptive fields. The resulting interim output is fed as the input to the architecture on the right.

technique was applied with under-sampling coefficient 80 [16]. Additionally data augmentation in form of random center shifting and rotation on π radian along horizontal plane was performed.

E. Training Process

Optimization was performed via ADAM optimizer with learning rate set to 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ [17]. Training was performed with batch size $bs = 64$ on the whole dataset, excluding the test part.

Prior to training the classifier, model's weights were set in accordance with the Xavier uniform initialization algorithm [18]. Over the training set $\mathbb{D} = \{\mathcal{X}_i, \mathcal{C}_i\}$, where $\mathcal{X}_i = \{\mathcal{I}_{ix}, \mathcal{I}_{iy}, \mathcal{I}_{iz}\}$ and $\mathcal{C}_i \in \{0, 1\}$ logarithmic loss was used and it was defined as follows:

$$\mathcal{L}(\theta, \mathbb{P}, \mathcal{B}) = -\frac{1}{|\mathcal{B}|} \sum_{\{\mathcal{X}_i, \mathcal{C}_i\} \in \mathcal{B}} \sum_{\mathcal{J} \in \{0,1\}} \mathbb{I}_{\mathcal{J}=\mathcal{C}_i} \mathbb{P}(\mathcal{J} = \mathcal{C}_i | \mathcal{X}_i, \theta). \tag{8}$$

Where $\mathcal{B} \subset \mathbb{D}$, $\mathbb{I}$ is an indicator function and $\mathbb{P}$ is, actually, the classifier which estimates class probability over input $\mathcal{X}$ and with parameters θ.

It should also be noted that the whole NoduleResNet framework is implemented using Keras [19] with Tensorflow as a back-end [20]. The computation time of the system over a scan with on average 330 candidates per scan is 2 second on a standard PC with a GPU NVIDIA Tesla K40. The training time for 1827 iterations with a batch size 32 take on average 3,465 seconds.

III. EXPERIMENTS AND RESULTS

A. Dataset

The dataset used in this work consists of 551, 065 nodule candidates and 1120 out of 1186 ground truth nodules (with

Team	CNN Type	0.125	0.25	0.5	1	2	4	8	Score (CPM)
DIAG CONVNET(arnaud.setio)	2D	0.636	0.727	0.792	0.844	0.876	0.905	0.916	0.814
iitm03(subru1603)	2D	0.394	0.491	0.570	0.660	0.732	0.795	0.851	0.642
LUNA16CAD(hirokinakano)	2D	0.113	0.165	0.265	0.465	0.596	0.695	0.785	0.440
LUNA16CAD(mattdns100689)	3D	0.640	0.698	0.750	0.804	0.847	0.874	0.897	0.787
LungNess(bim bam)	2D	0.453	0.535	0.591	0.635	0.696	0.741	0.797	0.635
UACNN(ccanoespinosa)	2D	0.655	0.745	0.807	0.849	0.880	0.907	0.925	0.824
CUMedVis(QiDou)	3D	0.677	0.737	0.815	0.848	0.879	0.907	0.922	0.827
NoduleResNet (Ours)	3D	0.517	0.602	0.720	0.788	0.822	0.839	0.856	0.735

the sensitivity of 94.4%) were publicly provided by LUNA16 [5] challenge organizers for false positive reduction task. The candidates were extracted by merging outputs of the generic nodule candidates detection algorithm ISICAD developed by Murphy et al. [21] with the SubsolidCAD algorithm aimed at detection of subsolid nodules proposed by Jacobs et al. [22] and LargeCAD, the purpose of which is to detect large nodules ($\geq$ than 10 mm in diameter). It was developed by Setio et al. [23]. Extraction was performed from 888 patient's scans selected from the publicly available dataset LIDC-IDRI [24], containing a total of 1018 CT scans. Due to the inconsistent slice spacing or high slice thickness 130 scans were omitted (CT scans with slice thickness greater than 3 mm were discarded from the final dataset). Remaining scans were divided into 10-folds with the objective to perform cross validation over them.

Ground truth for the resulting dataset was annotated by experienced thoracic radiologists in a two-phase screening process. Due to the first stage each of four radiologist marked separately any suspicious on nodule spots, and divided them into following categories: nodules with diameter $\geq$ 3 mm, nodules with diameter $>$ 3 mm, and non-nodule abnormality. In the next stage all nodules annotated only by one or two out of four radiologists, non-nodules, and nodules with diameter $>$ 3 mm were stated as irrelevant findings. Remaining 1186 nodules marked by at least three out of four radiologists were referred to as the reference standard.

B. Evaluations Metrics

Results of cross validation are then evaluated using the Free-Response Receiver Operating Characteristic (FROC [25]) and competition performance metric (CPM [26]). It computes average of the seven sensitivities measured at several false positives per scan (FPPS) thresholds, more concretely, at each FPPS $\in \{0.125, 0.25, 0.5, 1, 2, 4, 8\}$ true positive rate was computed. Mean of which forms the CPM.

C. Results

The resulted FROC calculated over zero fold provided LUNA16 dataset and is shown in Fig. 5. The sensitivity of 0.788 and 0.839 was reached at 1 and 4 false positives per scan respectively with corresponding CPM score 0.735. CPM score in comparison with different teams is provided at Table I. As it can be seen the result of this implementation is competitive in comparison with other methods.

IV. DISCUSSION

In this work a novel method aimed at false positive reduction in pulmonary nodule detection was described in detail along with results overview. Compared to the state of the art approaches our method show comparable or better performance.

The focus was made only on the false positive reduction instead of developing methods both for candidate detection and a false positive reduction. That means that this implementation is independent of how the candidate detection is done and can be used in combination with any candidate detection method. But it is also important to notice that the accuracy depends on candidate detection quality.

Due to the nature of CPM score it's sensitivities at the first four false positives per scan is highly susceptible to noise thus some ensemble techniques could make it less unstable and tend to converges to some $\mathcal{O}(FPRPS)$. Such behaviour of CPM score was shown by FROC curves in Fig. 5.

As Meindert Niemeijer et al. showed in a study [26] it is a good practice to combine different CAD systems in order to increase the performance because different CAD systems may have different strengths and weaknesses. So in the work of DIAG CONVNET(arnaud.setio), iitm03(subru1603), LUNA16CAD(hirokinakano), LungNess(bim bam), despite of the benefits provided by fusing information extracted from different views and reducing the model's complexity, such fusion strategy does not take benefit of the nature of nodule appearance as a volumetric spatial information.

At the same time, in the works of LUNA16CAD(mattdns100689) 3D space is covered via receptive fields selected to be cubes with shape (42, 42, 42) acting over input taken at different scales. But due to the proprieties of the measure concentration in high
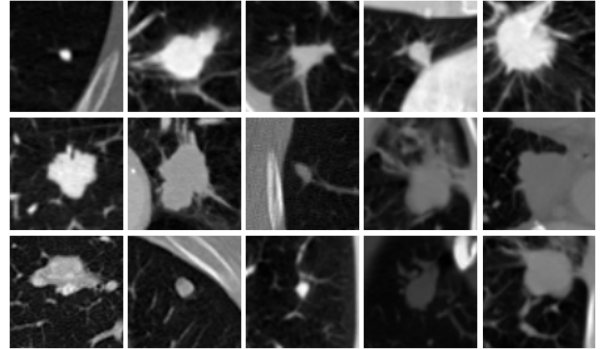

Fig. 4.   Pulmonary nodules detected by our method method with high confidence. Each patch is a representative maximum through three sequential horizontal planes of one annotated nodule. P-values for all presented candidates is ≥ 0.9
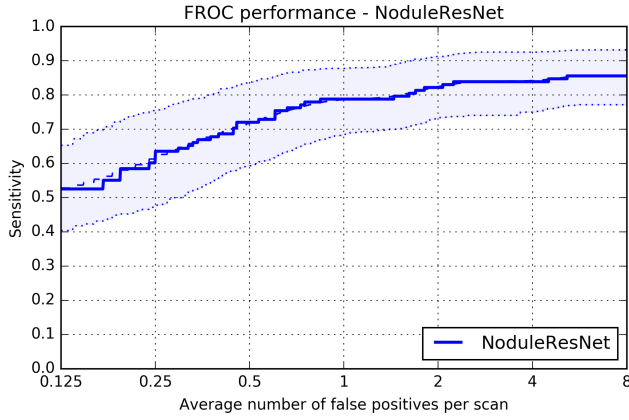
Fig. 5. FROC curve calculated over zero fold via LUNA16 framework as a result of proposed method, dashed curves stands for 95% confidence intervals, evaluated using bootstrap [27]

dimension euclidean space (even in three dimension) the major part of the cube volume results in addressing the surroundings of a nodule instead of the nodule itself, e.g. in unit space, when nodule's diameter is less than 12mm (> 80% of the whole annotated nodules) the fraction between nodule and its surroundings is about $\frac{1}{100}$. Like LUNA16CAD(mattdns100689), CUMedVis(QiDou) uses pretty similar approach, with a slight difference in that they employed specific input's size and kernel's size for each dimension (this is caused by the difference in resolution along dimensions) and they also use full ten-fold cross validation instead of the two-fold used by LUNA16CAD(mattdns100689). In contrast, our approach handles the problem of redundant contextual information by combining dimensionality reduction strategy with spatial high level features merging technique.

Fig. 6 presents the high confidence examples of true positives detection made by our system. For these examples, their morphological appearance is highly similar to the false positives shown in Fig. 4. In many LUNA16 participants' works there were a lot of cases similar to ones described above. As mentioned earlier, the solution for dealing with such challenging cases may be lies in the combination of methods.
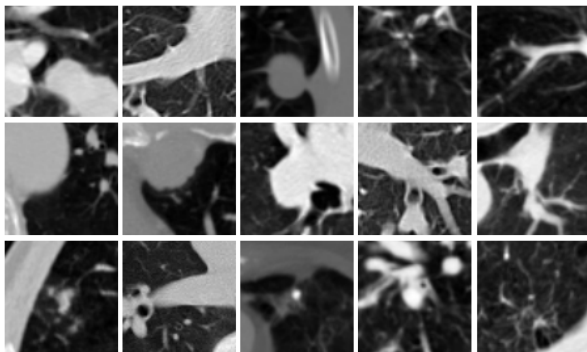
V. CONCLUSION

In this study a deep residual 3D convolutional neural network architecture aimed at feature extraction was described along with merging and dimensionality reduction strategies. The results show that proposed method is competitive and can be further proposed to use in combination with other methods to improve the detection performance and lower the risk of late cancer detection. Methods presented in this paper can be further extended to work with other kinds of 3D medical images in different tasks.

ACKNOWLEDGEMENT

Fig. 6. Non-nodule objects inside-lungs mistakenly detected by our method method with high confidence. Each patch is a representative maximum through three sequential horizontal planes. P-values for all presented candidates is ≥ 0.9.

REFERENCES

[1] Macedo Firmino, Antônio H Morais, Roberto M Mendoça, Marcel R Dantas, Helio R Hekis, and Ricardo Valentim. Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects. *Biomedical engineering online*, 13(1):41, 2014.

[2] American cancer society, 2016. cancer facts and figures 2016.

[3] 3rd annual congress of the russian breast cancer society.

[4] European consensus statement on lung cancer: Risk factors and prevention. lung cancer panel.

[5] Thomas de Bel Moira S.N. Berens Cas van den Bogaard Piergiorgio Cerello Hao Chen Qi Dou Maria Evelina Fantacci Bram Geurts Robbert van der Gugten Pheng Ann Heng Bart Jansen Michael M.J. de Kaste Valentin Kotov Jack Yu-Hung Lin Jeroen T.M.C. Manders Alexander Sora-Mengana Juan Carlos Garca-Naranjo Mathias Prokop Marco Saletta Cornelia M Schaefer-Prokop Ernst T. Scholten Luuk Scholten Miranda M. Snoeren Ernesto Lopez Torres Jef Vandemeulebroucke Nicole Walasek Guido C.A. Zuidhof Bram van Ginneken Colin Jacobs Arnaud Arindra Adiyoso Setio, Alberto Traverso. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. 2016.

[6] Maxine Tan, Rudi Deklerck, Bart Jansen, Michel Bister, and Jan Cornelis. A novel computer-aided lung nodule detection system for ct images. *Medical Physics*, 38(10):5630–5645, 2011.

[7] Qiang Li, Shusuke Sone, and Kunio Doi. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional ct scans. *Medical Physics*, 30(8):2040–2051, 2003.

[8] Q. Dou, H. Chen, L. Yu, J. Qin, and P. A. Heng. Multi-level contextual 3d cnns for false positive reduction in pulmonary nodule detection. *IEEE Transactions on Biomedical Engineering*, PP(99):1–1, 2016.

[9] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Snchez, and B. van Ginneken. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, May 2016.

[10] Y-Lan Boureau, Francis Bach, Yann LeCun, and Jean Ponce. Learning mid-level features for recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2559–2566. IEEE, 2010.

[11] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, volume 9, pages 249–256, 2010.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[13] Mikhail Belyaev Yulia Dodonova Sergey Korolev, Amir Safiullin. Residual and plain convolutional neural networks for 3d brain mri classification. 2017.

[14] Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *CoRR*, abs/1608.05895, 2016.

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[16] N Chawla, Nathalie Japkowicz, and A Kolcz. Special issue on learning from imbalanced datasets, sigkdd explorations. In *ACM SIGKDD*, 2004.

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[18] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[19] François Chollet. Keras (2015). *URL http://keras. io*.

[20] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[21] K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. *Medical Image Analysis*, 13(5):757–770, 2017/03/20 XXXX.

[22] Colin Jacobs, Eva M. van Rikxoort, Thorsten Twellmann, Ernst Th Scholten, Pim A. de Jong, Jan-Martin Kuhnigk, Matthijs Oudkerk, Harry J. de Koning, Mathias Prokop, Cornelia Schaefer-Prokop, and Bram van Ginneken. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Medical Image Analysis*, 18(2):374–384, 2017/03/20 XXXX.

[23] Arnaud A. A. Setio, Colin Jacobs, Jaap Gelderblom, and Bram van Ginneken. Automatic detection of large pulmonary solid nodules in thoracic ct images. *Medical Physics*, 42(10):5642–5653, 2015.

[24] McLennan G. Bidaut L. McNitt-Gray M.F. Meyer C.R. Reeves A.P. Zhao B. Aberle D.R. Henschke C.I. Hoffman E.A. Kazerooni E.A. MacMahon H. van Beek E.J. Yankelevitz D. Biancardi A.M. Bland P.H. Brown M.S. Engelmann R.M. Laderach G.E. Max D. Pais-R.C. Qing D.P. Roberts R.Y. Smith A.R.-Starkey A. Batra P.-Caligiuri P. Farooqi A. Gladish G.W. Jude C.M.-Munden R.F. Petkovska I. Quint L.E. Schwartz L.H. Sundaram B. Dodd L.E. Fenimore C. Gur D. Petrick N. Freymann J. Kirby J. Hughes B. Casteele A.V. Gupte S. Sallam M. Heath M.D. Kuhn M.H. Dharaiya E. Burns R. Fryd D.S. Salganicoff M. Anand V. Shreter U. Vastagh S. Croft B.Y. Clarke L.P. Armato III, S.G. Data from lidc-idri. 2015.

[25] International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. 2008.

[26] M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop, and B. van Ginneken. On combining computer-aided detection systems. *IEEE Transactions on Medical Imaging*, 30(2):215–223, Feb 2011.

[27] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.