electronics

MDPI

Article

Improving Pneumonia Classification and Lesion Detection Using Spatial Attention Superposition and Multilayer Feature Fusion

Kang Li [1,†], Fengbo Zheng [1,†], Panpan Wu [1,*], Qiuyuan Wang [2,3], Gongbo Liang [4] and Lifen Jiang [1,*]

1 College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
2 Graduate School, Beijing University of Chinese Medicine, Beijing 100029, China
3 China-Japan Friendship Hospital, Beijing 100029, China
4 Department of Computing and Cyber Security, Texas A&M University, San Antonio, TX 78224, USA
* Correspondence: pwu@tjnu.edu.cn (P.W.); wxxjlf@sina.com (L.J.); Tel.: +86-1382-031-5686 (P.W.); +86-1382-077-1820 (L.J.)
† These authors contributed equally to this work.

Abstract: Pneumonia is a severe inflammation of the lung that could cause serious complications. Chest X-rays (CXRs) are commonly used to make a diagnosis of pneumonia. In this paper, we propose a deep-learning-based method with spatial attention superposition (SAS) and multilayer feature fusion (MFF) to facilitate pneumonia diagnosis based on CXRs. Specifically, an SAS module, which takes advantage of the channel and spatial attention mechanisms, was designed to identify intrinsic imaging features of pneumonia-related lesions and their locations, and an MFF module was designed to harmonize disparate features from different channels and emphasize important information. These two modules were concatenated to extract critical image features serving as the basis for pneumonia diagnosis. We further embedded the proposed modules into a baseline neural network and developed a model called SAS-MFF-YOLO to diagnose pneumonia. To validate the effectiveness of our model, extensive experiments were conducted on two CXR datasets provided by the Radiological Society of North America (RSNA) and the AI Research Institute. SAS-MFF-YOLO achieved a precision of 88.1%, a recall of 98.2% for pneumonia classification and an AP_{50} of 99% for lesion detection on the AI Research Institute dataset. The visualization of intermediate feature maps showed that our method could facilitate uncovering pneumonia-related lesions in CXRs. Our results demonstrated that our approach could be used to enhance the performance of the overall pneumonia detection on CXR imaging.

Keywords: pneumonia detection; deep learning; lesion localization

Citation: Li, K.; Zheng, F.; Wu, P.; Wang, Q.; Liang, G.; Jiang, L. Improving Pneumonia Classification and Lesion Detection Using Spatial Attention Superposition and Multilayer Feature Fusion. *Electronics* **2022**, *11*, 3102. https://doi.org/10.3390/electronics11193102

Academic Editor: Panagiota Spyridonos

Received: 23 August 2022
Accepted: 26 September 2022
Published: 28 September 2022

1. Introduction

Pneumonia is a lung inflammation usually caused by bacterial or viral infections. It could cause serious complications including lung abscesses, bacteremia, respiratory failure, etc. [1,2] Early diagnosis and treatment of pneumonia are critical to prevent potential complications. A chest X-ray (CXR) is the most commonly used technique for pneumonia diagnosis due to its high availability, lower cost, and lower radiation exposure compared with computed tomography (CT) [3]. With the rapid development of computer science, various machine-learning-based methods have been proposed for automatic pneumonia detection [3–11]. However, further studies are still needed in some aspects before these intelligent diagnosis systems can be used in practice. For instance, given a deep-learning-based diagnosis model, we need to confirm that pneumonia-related lesions can be detected and serve as the basis for pneumonia detection so that the decisions made are reliable [12].

In typical deep-learning-based CXR image diagnosis models, decisions are made based on the feature maps (also known as channels in neural network layers) that are generated

from the input CXR image or output from prior layers by applying filters or feature detectors. Feature maps capture different kinds of features in a CXR image. The performance of a classifier highly depends on whether the feature maps could capture useful characteristics for the classification task. To make a diagnosis of pneumonia, the characteristics we should leverage are the lesions and their locations in the CXR image.

Although there is a range of research efforts focusing on improving the overall performance (e.g., the accuracy of clinical decisions) of CXR diagnosis models [13–17], generating feature maps that can embody the lesions and their locations remains challenging and underinvestigated. In this paper, we propose a deep-learning-based method with spatial attention superposition (SAS) and multilayer feature fusion (MFF) to facilitate pneumonia diagnosis based on CXRs. The SAS module, which leverages visual attention mechanisms, aims at highlighting important intrinsic characteristics (i.e., lesions and their locations) that are related to the task of pneumonia classification. Furthermore, a new multilayer feature fusion module (MFF) is designed to merge different features contained in different channels in which process important features for pneumonia detection are maintained while trivial features are faded to improve the robustness of a model. To validate the effectiveness of our model, we applied our SAS-MFF-YOLO model to perform pneumonia classification and lesion detection on two CXR datasets provided by the Radiological Society of North America (RSNA) and the AI Research Institute. SAS-MFF-YOLO achieved a precision of 18.3%, a recall of 58.9% and an AP_{50} of 31.0% on the dataset from RSNA and a precision of 88.1%, a recall of 98.2%, an AP_{50} of 99.0% and an mAP of 67.9% on the dataset from the AI Research Institute, which showed significant improvements in the performance for both pneumonia classification and lesion detection.

Our main contributions in this work can be summarized as follows: We propose an SAS module that leverages visual attention mechanisms to highlight important intrinsic characteristics (i.e., lesions and their locations) that are related to the task of pneumonia classification. In addition, we develop a new multilayer feature fusion module (MFF) to merge different features contained in different channels in which process important features for pneumonia detection are maintained while trivial features are faded to improve the robustness of the model. The two proposed modules can be easily embedded into existing baseline classification neural networks to enhance their performance in pneumonia detection.

2. Related Work

Deep learning, which plays vital roles in all kinds of computer vision tasks, has been widely used in the field of medical imaging diagnosis. Rajpurkar et al. proposed a 121-layer convolutional network work (CNN) called CheXNet that could classify 14 lung diseases [13]. However, this work only predicted the probabilities of the disease. The location information of the lesions was not provided. Sirazitdinov et al. [14] conducted comparative experiments on the CXR dataset provided by the RSNA using several baseline deep learning models and investigated the problem of unbalanced categories. Jaiswal et al. [15] leveraged the Mask R-CNN network framework to identify and locate the lesion area. However, the training process was complicated and required an enormous amount of computation, preventing it from being used in practice. Yao et al. [16] explored the usefulness of statistical dependence between labels for making more accurate predictions. However, their method was not able to distinguish similar characteristics of different chest diseases in CXRs, and the accuracy of their method only reached 71.3%. Wang et al. [17] built a multilabeled CXR dataset Chest X-ray8 and explored various deep convolutional neural networks to make a diagnosis on eight lung diseases, but the performance of their classifiers was quite limited.

3. Methods

In computer vision, the attention mechanism is widely adopted to make deep learning models focus on important parts of the input images that are valuable for a specific task (e.g., classification or object detection). It can reduce complicated tasks into more manageable areas of attention to understand and process sequentially. In this work, we designed

a spatial attention superposition (SAS) module that leveraged the idea of both channel attention [18,19] and spatial attention [20,21] to reveal regions of interest (i.e., the appearance of lesions) and their locations in CXR imaging. Furthermore, a multilayer feature fusion module (MFF) was developed to preserve important features resulting from SAS and ensure the model was not distracted by unimportant features. These two modules together ensured that useful features for pneumonia detection could be extracted and focused on, and thus could improve the performance of a pneumonia classifier or a lesion detector.

3.1. Spatial Attention Superposition Module

The SAS module considered both channel-wise and space-wise characteristics of image features that are important for pneumonia detection. Figure 1 shows the overall structure of the SAS module. Given an input image or a feature map x_i, spatial attention was first calculated for the entire channel. The attention score a_i was computed by averaging and maximizing the channel dimension and through a CBL layer (i.e., constructed by a convolutional block, batch normalization and leaky Relu) and a *Sigmoid* function to assist the generation of the following two attentions.

$$a_i = Sigmoid(CBL([\text{mean}(x_i), \max(x_i)])) \quad (1)$$

where i represents the position index, x_i stands for the vector value of all channels at position i, mean(x_i) represents the average value, max(x_i) denotes the maximum value and [ ] is the concatenation operation.

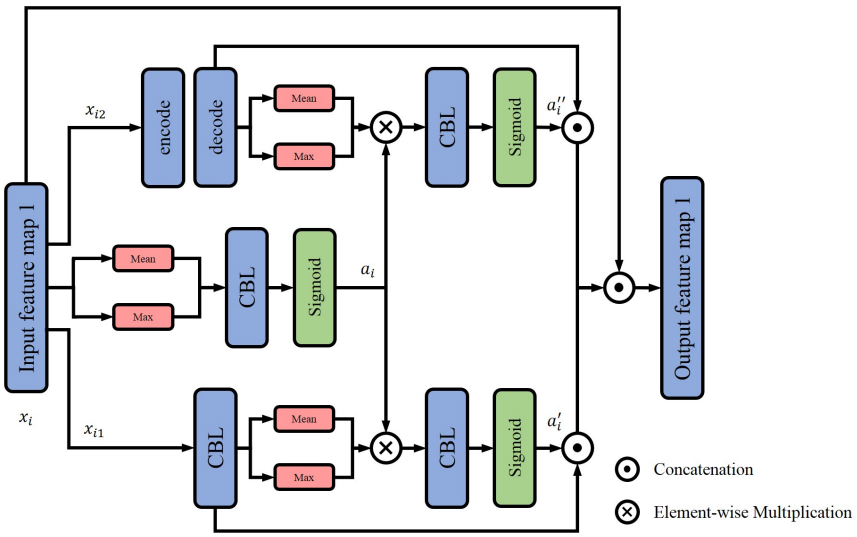


Figure 1. Structure of SAS module.

Meanwhile, the input feature x_i was divided into two parts, x_{i1} and x_{i2}, to further calculate two kinds of attention values—one focused on channelwise features, the other focused on target recognition. The first part, x_{i1}, passed through a CBL layer to unify the overall channel dimension and obtain the attention value a_i' through a spatial attention calculation module. The other part, x_{i2}, which aimed at improving the receptive field and semantics of features, passed through an encoding–decoding layer that benefited the target recognition. Then, the spatial attention a_i'' of features with a wide receptive field and rich semantics were calculated.

At last, the attention values of the two parts a_i' and a_i'' that were derived based on the initial attention a_i, were, respectively, multiplied with the original feature maps and fused. Taking the advantages of both channel attention and spatial attention, the output feature map of the SAS module was expected to contain and highlight the area of interest (i.e., lesions and their locations if they exist) for pneumonia detection.

3.2. Multilayer Feature Fusion Module

Our SAS module could identify features in different channels or feature maps that needed to be focused on while detecting pneumonia. Due to the heterogeneity of different channels, these features usually depict different characteristics of the original CXR imaging. Some features are important for pneumonia diagnosis, but some are trivial. In the best case, in the learning phase, a model will just focus on important features and ignore those trivial ones so that the model can be accurate (e.g., utilizing features related to pneumonia diagnosis) and conserve its robustness (e.g., tolerating variance of trivial features). However, traditional feature fusion methods usually mix important features with trivial ones, which affects the model performance. In this paper, we designed an MFF model to consolidate the assorted information. While doing feature fusion, our MFF module keeps discriminating/important information and neutralizes those trivial features. This way, the model hardly adjusts itself to accommodate those trivial features and focuses on important features.

The structure of our MFF module is shown in Figure 2. The input was the feature map resulting from the SAS model. We first used a one-dimensional convolution to implement a local cross-channel interaction strategy without dimensionality reduction to decrease the complexity [22]. Specifically, we performed a global average pooling (*GAP*) on the input and a layer with a size of $1 \times 1 \times C$ was generated, where C was the size of the channel dimension. Then, we applied a one-dimensional convolution to obtain the weight ω of each channel. The calculation formula of the weight can be expressed as

$$\omega = Sigmoid(C1D_k(GAP(x))) \tag{2}$$

where $GAP(x)$ represents the global average pooling, $C1D_k$ refers to a one-dimensional convolution with kernel size k and the kernel size k is nonlinearly proportional to the channel dimension C.

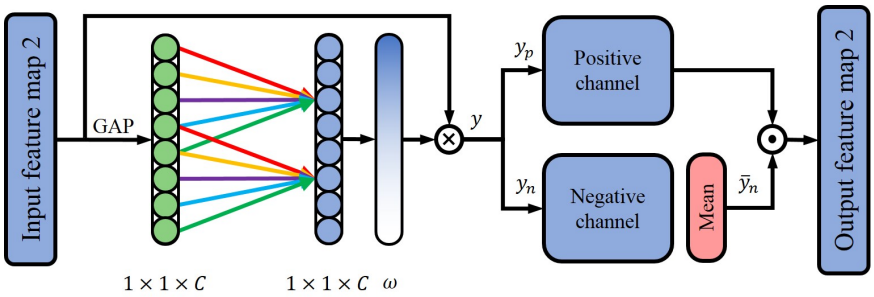


Figure 2. Structure of MFF module.

Different from normal feature fusion, the result channels or feature maps were then divided into two parts in a ratio of λ after sorting based on attention values (i.e., importance and contribution to the final classification decision)—the first λc channels were chosen as positive channels y_p and the rest as negative channels y_n.

Therefore, channels with more discriminating and dominant features (i.e., playing more important roles in pneumonia diagnosis) became the active channels, while channels with tiny hidden information became negative ones. For the negative channels, we calculated the average value $\overline{y}_n$ to substitute the original y_n. At last, we concatenated the channels together to form a new feature map which was then forwarded to a classifier or an object detector. In this case, we applied cheap means to complement tiny hidden details without losing intrinsic features so that our MFF model could summarize important information for pneumonia lesion detection and promote the robustness of a model.

3.3. SAS-MFF-YOLO Model

To perform pneumonia diagnosis, we embedded the two proposed modules into a baseline neural network YOLOv5 [23] and developed a model called SAS-MFF-YOLO.

It mainly consisted of three parts, a backbone layer to extract initial features from input images, a neck layer to generate feature pyramids so that important features with different scales could be identified and a prediction layer for classification and object detection. Among them, the focus was utilized as the beginning of the network, CSP1 and CSP2 were constructed on the basis of CSPNet and SPP was employed to expand the receptive field of the feature layer. The proposed SAS and MFF modules served as the feature extraction and feature fusion modules in the neck part, respectively. The general idea was that the SAS highlighted the intrinsic and discriminating information of the lesion, while the MFF merged the enhanced and complementary features of different channels processed differently. Figure 3 shows the overall structure of the developed SAS-MFF-YOLO model.

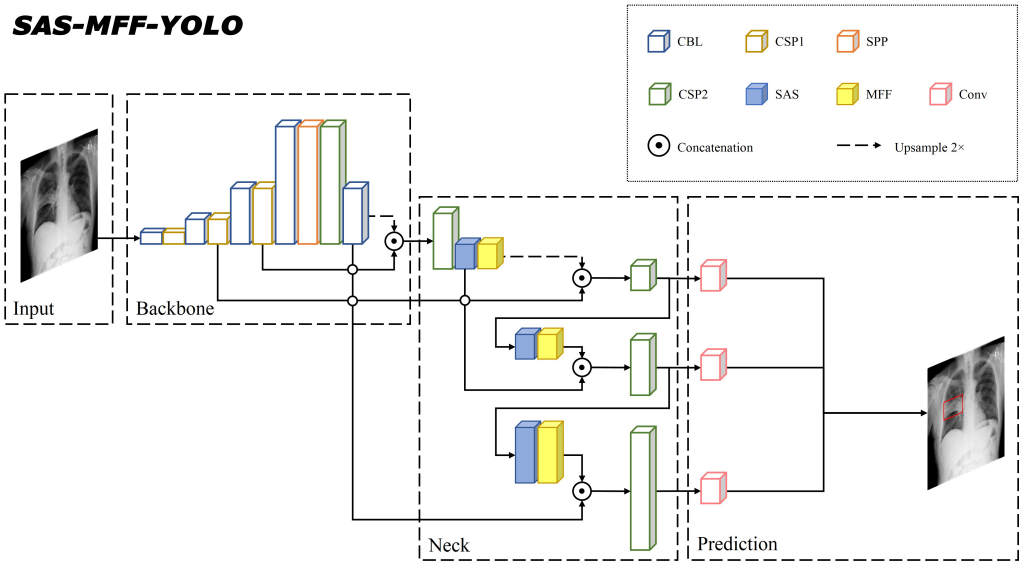


Figure 3. The overall structure of the embedded model SAS-MFF-YOLO. The blue blocks and yellow blocks represent the proposed SAS and MFF modules, respectively.

4. Experiments

4.1. Dataset

In this work, we performed experiments on two CXR image datasets to validate the effectiveness of our method. One was publicly provided by the Radiological Society of North America (RSNA) [24] and contains 30,028 CXR images as training data and 3000 as testing data. There are three classes in the dataset: Normal, Lung Opacity and No Lung Opacity/Not Normal. An image without any evidence of pneumonia is considered "Normal". "Lung Opacity" indicates the finding of pneumonia-related lesions. If a CXR image does not contain pneumonia-related lesions but has other abnormalities, it is labeled as "No Lung Opacity/Not Normal". The distribution of these three classes is shown in Table 1. Besides the labels, the dataset also provides bounding boxes that specify the lesions (including the coordinate of the upper left corner, width and height of the box) which can be used as the ground truth for object detection.

The other dataset used in this work was provided by the AI Research Institute [25] and consists of 20,013 training images and 6671 testing ones.

Table 1. Distribution of the RSNA CXR image dataset.

Class	Target	Images
Normal	No	8851
Lung Opacity	Yes	9555
No Lung Opacity/Not Normal	No	11,821

4.2. Experiment Setup

We developed the SAS-MFF-YOLO model using the PyTorch framework. To train the model, we adopted a stochastic gradient descent (SGD) optimizer with a learning rate of 0.1 and a weight decay of 5×10^{-4}. The overall classification loss, objectness loss and anchor box regression loss were combined jointly as the total loss. For our MFF module, λ (see Section 3.2) was set to 1/2. The batch size was set to 128.

We evaluated the performance of our models in two common tasks in pneumonia diagnosis—classification (i.e., pneumonia diagnosis) and object detection (i.e., marking bounding boxes for lesions). Regarding the experiments, we first performed an ablation study to verify the effectiveness of the proposed SAS and MFF modules. Specifically, a comparison of the baseline model YOLOv5 and three embedded models including SAS-YOLO, MFF-YOLO and SAS-MFF-YOLO was conducted. To further validate the effectiveness of our method, we compared SAS-MFF-YOLO with models that are generally adopted for object detection tasks, including YOLOv3, YOLOv5 and RetinaNet.

For pneumonia classification, we adopted two commonly used evaluation metrics—precision and recall, to assess different models. The formulas for precision and recall are:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

For object detection, if the intersection over union (IoU) of a predicted bounding box and the ground truth (i.e., the bounding box marked by radiologists) was greater than the set threshold T, it was considered a true positive (TP) case. Otherwise, it was a false positive (FP) case. If the model correctly stated that there was not any lesion (i.e., output bounding boxes) in the CXR image, it counted for a true negative (TN) case. For the lesions (i.e., bounding boxes) that were not detected by the models, they were considered false negative (FN) cases. Therefore, precision was the ability of the model to identify only the relevant objects, and recall was the ability of a model to find all the relevant cases (all ground truth bounding boxes).

To evaluate the performance of the object detection with different models, taking the precision as the vertical axis and the recall as the horizontal axis, the P–R curve was obtained, and the area under the curve was the average precision (AP) value. Specifically, AP_{50} and mAP were adopted. The former represented the AP value when the threshold T was set to 0.5 while the latter was computed by averaging the results with the threshold T varying in the range of 0.50∼0.95 with a step size of 0.05.

5. Results

5.1. Ablation Study

We leveraged datasets from the RSNA and AI Research Institute and performed an ablation study to verify the effectiveness of the SAS and MFF modules. The performance of different models applied to the two datasets is shown in Tables 2 and 3. Overall, employing SAS in conjunction with MFF greatly improved the performance in both pneumonia classification and lesion detection. Regarding the dataset from RSNA, our SAS-MFF-YOLO model achieved a precision of 18.3%, a recall of 58.9%, an AP_{50} of 31.0% and a mAP of 9.7%, improving the baseline model by 4.5%, 3.2%, 3.2% and 1.5% in terms of precision, recall, AP_{50} and mAP, respectively. For the dataset from the AI Research Institute, our SAS-MFF-YOLO model achieved a precision of 88.1%, a recall of 98.2%, an AP_{50} of 99.0% and a mAP of 67.9%, also indicating a significant improvement.

Table 2. The performance of different models on the dataset provided by RSNA. Precision and recall are for pneumonia classification, while AP_{50} and mAP are for the task of lesion object detection.

Model	Precision (%)	Recall (%)	AP_{50} (%)	mAP (%)
YOLOv5	13.8	55.7	27.8	8.2
SAS-YOLO	14.0	55.8	28.9	8.3
MFF-YOLO	14.1	57.1	27.7	8.4
SAS-MFF-YOLO	**18.3**	**58.9**	**31.0**	**9.7**

Table 3. The performance of different models on the dataset provided by AI Research Institute. Precision and recall are for pneumonia classification, while AP_{50} and mAP are for the task of lesion object detection.

Model	Precision (%)	Recall (%)	AP_{50} (%)	mAP (%)
YOLOv5	77.6	98.5	97.6	61.2
SAS-YOLO	81.9	98.6	98.2	61.6
MFF-YOLO	86.0	**99.3**	82.5	57.1
SAS-MFF-YOLO	**88.1**	98.2	**99.0**	**67.9**

These results proved that utilizing an attention stacking mechanism (i.e., channel and spatial attention) and leveraging multilayer features could promote pneumonia diagnosis. After the SAS module was embedded, the performance of the model, especially the AP_{50}, was enhanced. This is mainly because the SAS module could highlight important features for pneumonia identification (i.e., pneumonia-related lesions) to stimulate object detection. It could also be observed that the MFF module played an important role in boosting the recall. Since the discriminating and important features in the active channels were maintained and focused, while negative channels with redundant patterns were neutralized in MFF, few trivial features were taken into consideration by the model while performing classification/object detection (i.e., trivial features hardly affected the judgment of the model), the robustness of the model was improved. Therefore, both SAS and MFF modules played important roles in promoting performance.

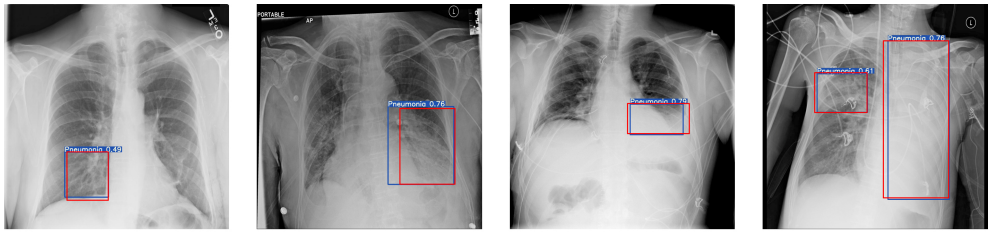
5.2. Comparison with Other Object Detection Models

We also compared the proposed SAS-MFF-YOLO model with other state-of-art object detection models to further validate the effectiveness of our method. The results are shown in Table 4. We can find that our method achieved the best performance in the task of lesion object detection.

Table 4. The performance of different models in pneumonia-related lesion detection on the RSNA and the AI Research Institute datasets. The evaluation metric shown is the AP_{50}.

Dataset	YOLOv3 (%)	YOLOv5 (%)	RetinaNet (%)	SAS-MFF-YOLO (%)
RSNA	24.3	27.8	25.5	**31.0**
AI Research Institute	86.8	97.6	88.2	**99.0**

Several lesion-labeled CXR images are shown in Figure 4. The ground truth is marked by the red bounding boxes while our prediction is marked by the blue ones. The numbers above the predicted bounding box are the confidence scores that reflect how confident the model is regarding its judgment (i.e., a box contains pneumonia-related lesion). It can be seen that our model could effectively recognize pneumonia-related lesions from CXR imaging.



(a) left lower (b) right (c) right middle (d) both side

Figure 4. Four samples of pneumonia-related lesion detection. The red bounding boxes display the ground truth while the blue ones show the prediction results.

6. Discussion

Comparing the performance of different models demonstrated that our SAS and MFF modules could identify important intrinsic features in the CXR images and stay focused on them to facilitate pneumonia diagnosis. To further illustrate the advantages of our method, the intermediate feature maps were visualized, which can also provide insight into the internal representations for the features on which a model relies to make a decision.

6.1. Intermediate Feature Map Visualization

A sample of visualization results is shown in Figure 5. The feature maps were generated by our SAS-MFF-YOLO model. The heatmaps represent the feature maps in which light colors (e.g., yellow) indicate potentially important area for pneumonia detection. By assessing the intermediate feature maps, we can find that initially, our model extracted the overall characteristics (e.g., texture, outline) of the lung. With the progress advancing, features related to pneumonia diagnosis were extracted and their locations were highlighted, which showed the changes in the focus of the model.

6.2. Feature Maps Analysis

In this work, we also compared the feature maps learned by our SAS-MFF-YOLO model and the baseline model YOLOv5 to provide a deeper insight into what caused the performance differences in pneumonia classification and lesion detection.

Figure 6 shows several selected feature maps for a CXR image generated by SAS-MFF-YOLO and YOLOv5. It can be observed from the features maps that even though the features maps from these two models showed similar patterns, pneumonia-related lesions revealed by SAS-MFF-YOLO were more accurate in terms of the lesion area and its location. Furthermore, we could find that feature maps generated by SAS-MFF-YOLO concentrated more on the lesion area (i.e., trivial information was faded), while the feature maps of YOLOv5 depicted more features but some were clearly not related to pneumonia diagnosis. Therefore, YOLOv5 was likely to be distracted by those trivial features. As a result, our SAS-MFF-YOLO model found the lesion correctly with a relatively high confidence, but YOLOv5 made an incorrect detection of the lesion with a lower confidence score which showed its uncertainty regarding the decision.

Another example is shown in Figure 7. In this case, even though YOLOv5 correctly detected the lesion as SAS-MFF-YOLO did, its confidence score was much lower since important and trivial information received equal-weight attention in the decision-making. These comparison results proved that our SAS-MFF-YOLO model obtained better performance in highlighting the areas that were related to pneumonia diagnosis, which again validated the effectiveness of the proposed SAS and MFF modules.

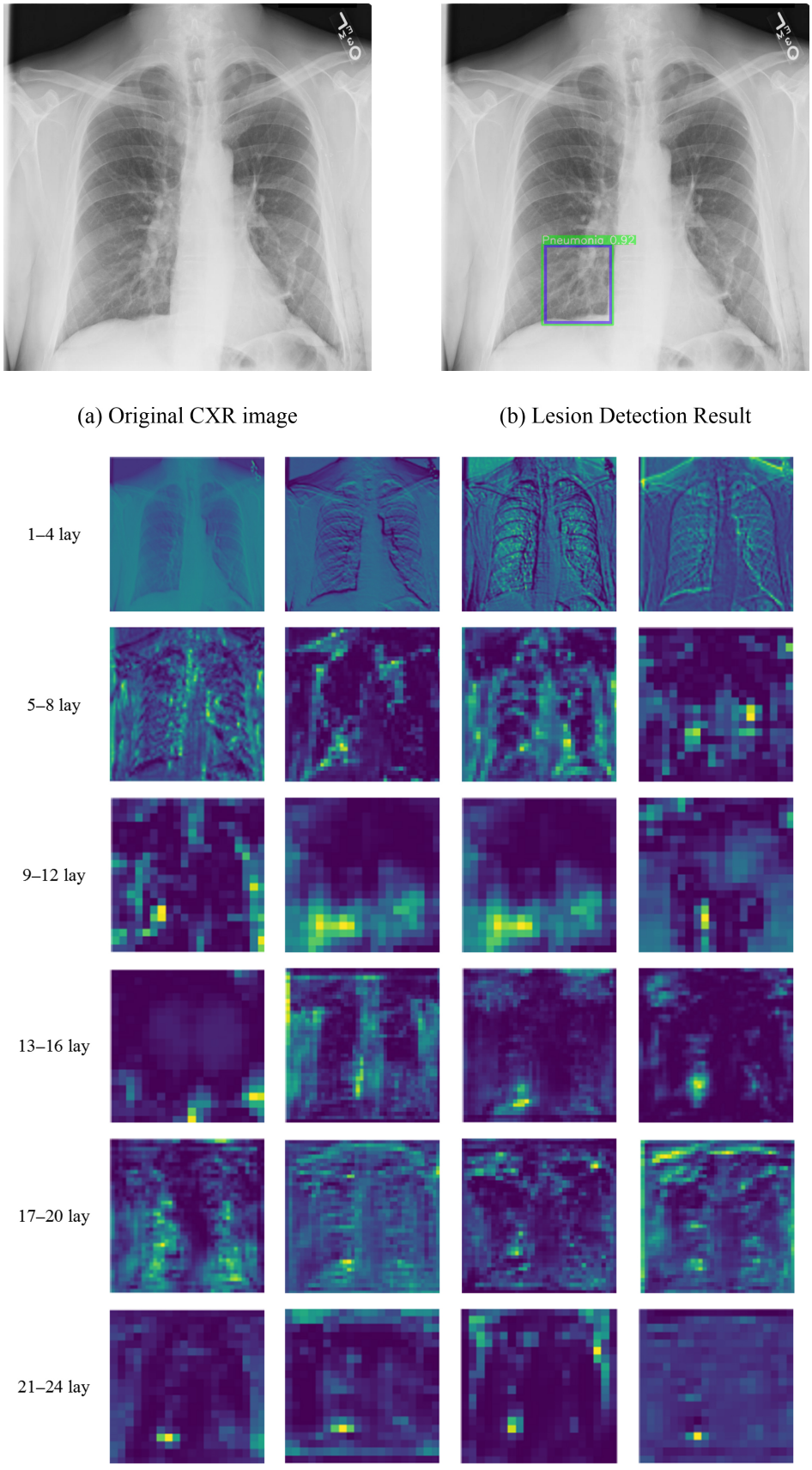


Figure 5. Visualization of intermediate feature map produced by our SAS-MFF-YOLO model. (**a**,**b**) show the original CXR image and pneumonia-related lesion detection result. The heatmaps show the feature maps in which the yellow color indicates potentially important areas for pneumonia detection.

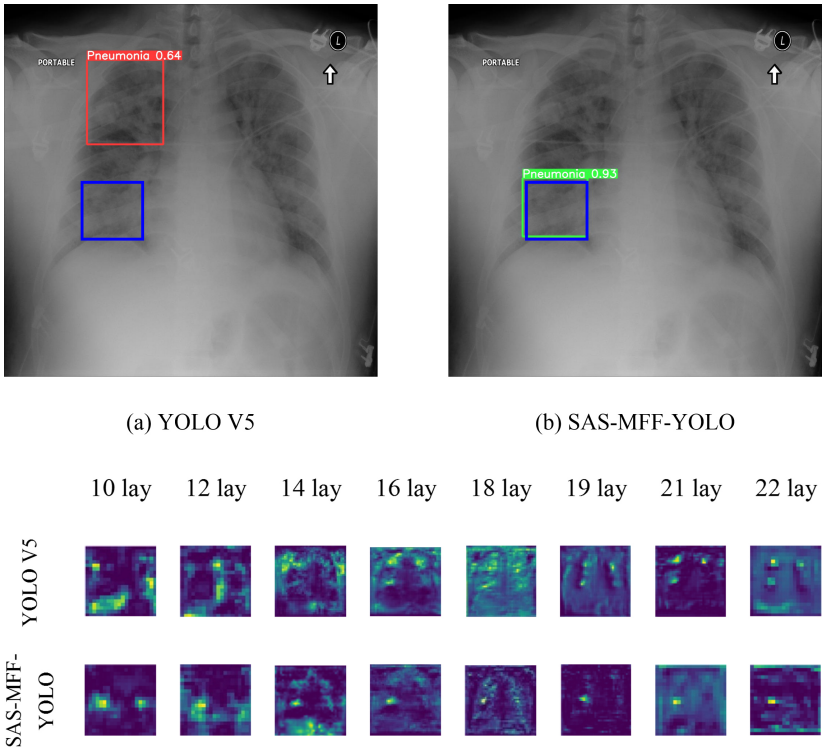


Figure 6. An example of feature maps visualization. (**a**,**b**) are the results of lesion detection produced by YOLOv5 and SAS-MFF-YOLO, respectively. The blue bounding boxes are the ground truth, while the red and green ones were labeled by the two models. The heat maps show the selected feature maps.

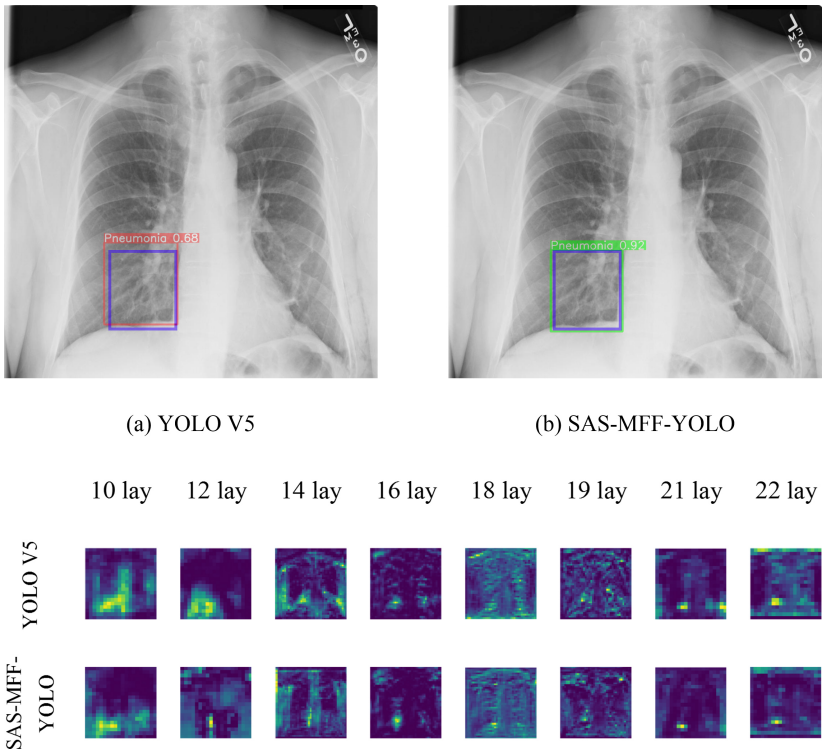


Figure 7. Another example of feature maps visualization. (**a**,**b**) are the results of lesion detection produced by YOLOv5 and SAS-MFF-YOLO, respectively. The blue bounding boxes are the ground truth, while the red and green ones were labeled by the two models. The heat maps show the selected feature maps.

6.3. Comparison with Related Work

Most of the previous studies have used artificial intelligence (AI) and deep learning (DL) techniques to detect the presence or absence of pneumonia (binary classification problem) [26,27]. However, few studies have used deep learning, especially convolutional neural networks (CNNs), to classify and locate pneumonia based on their features.

The CheXNet [13] proposed by Rajpurkar et al. could only predict disease probability. Our model took the advantages of channel attention and spatial attention to identify intrinsic imaging features of pneumonia-related lesions and their locations, based on the prediction of disease probability.

Although most researchers [8,14,15] used two-stage object detectors with higher accuracy, our SAS-MFF-YOLO model achieved better performance in terms of efficiency and accuracy on the same dataset. This stemmed from the ability of our proposed MFF module to coordinate different features from different channels and emphasize important information, which could reduce the amount of calculation and improve accuracy.

Previous research [14,16,19] has focused on using transfer learning and pretrained models. However, the pneumonia detection of CXR images is different from the traditional target detection, which needs targeted improvement. In this paper, we modified an existing model to detect pneumonia. Specifically, we proposed the SAS-MFF-YOLO model, which could easily be embedded into existing baseline classification neural networks to enhance their performance in pneumonia detection.

6.4. Broader Impact

Artificial intelligence diagnosis has the advantages of a high efficiency and low cost. Artificial intelligence pneumonia detection can improve the working efficiency of doctors and relieve the pressure of hospitals. At the same time, there are relatively few experienced doctors and high-end medical hardware equipment in remote areas. The application of artificial intelligence pneumonia detection can alleviate the current situation where medical workers only rely on their own clinical experience to diagnose patients' diseases, and it can help solve the uneven distribution of medical resources.

6.5. Limitation and Future Work

In this study, we aimed to develop an effective method to facilitate pneumonia diagnosis. However, even though deep-learning-based diagnosis approaches have already achieved expert-like performance, they are usually not trusted by physicians. The main reason is the low interpretability of decisions neural networks make. For most intelligent diagnosis systems, either no interpretation of the decision is provided, or the interpretation provided, e.g., the heatmap adopted in this work which shows the feature map that serves as the basis of a decision, is not straightforward and clear for the physicians to understand. In the future, we plan to explore deep learning approaches that associate the highlighted area in the feature map (i.e., lesions) with its practical meaning (e.g., clinical terms used by radiologists). This way, the physicians could easily understand why a certain decision is being made, which will greatly improve the interpretability and reliability of deep-learning-based diagnosis methods.

7. Conclusions

In this article, we propose a deep neural network called SAS-MFF-YOLO with spatial attention superposition (SAS) and multilayer feature fusion (MFF) to facilitate pneumonia diagnosis based on CXRs. Leveraging both attention stacking mechanism and a new strategy of feature fusion, SAS-MFF-YOLO was able to extract and focus on useful features for pneumonia diagnosis. Experiments on datasets from the RSNA and the AI Research Institute proved that our method could enhance a model's performance in pneumonia classification and lesion detection. Comparing feature maps generated by our method and the baseline model further illustrated the advantages of our SAS and MFF modules.

Author Contributions: L.J., F.Z. and P.W. conceptualized this study; F.Z. and K.L. designed the model; K.L. implemented the model; P.W. and G.L. contributed to the improvement of the model; F.Z. designed the experiment; Q.W. reviewed and evaluated the results; F.Z., K.L. and L.J. analyzed the evaluation results; F.Z. and K.L. wrote the manuscript. All authors reviewed the manuscript and contributed to revisions. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (grant no. 61902282), Tianjin Municipal Education Commission Project for Scientific Research Plan (grant no. 2018KJ155) and Doctoral Foundation of Tianjin Normal University (grant no. 043135202-XB1707).

Data Availability Statement: The trained model and related experiment results are available at: https://github.com/CS-likang/SAS-MFF-YOLO.git (accessed on 22 July 2022).

Acknowledgments: The authors thank the anonymous reviewers for their valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rolston, K.V. The spectrum of pulmonary infections in cancer patients. *Curr. Opin. Oncol.* **2001**, *13*, 218–223. [CrossRef] [PubMed]
2. Mizgerd, J.P. Inflammation and pneumonia: Why are some more susceptible than others? *Clin. Chest Med.* **2018**, *39*, 669–676. [CrossRef] [PubMed]
3. Qin, C.; Yao, D.; Shi, Y.; Song, Z. Computer-aided detection in chest radiography based on artificial intelligence: A survey. *Biomed. Eng. Online* **2018**, *17*, 1–23. [CrossRef]
4. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **2020**, *121*, 103792. [CrossRef]
5. Yee, S.L.K.; Raymond, W.J.K. Pneumonia diagnosis using chest X-ray images and machine learning. In Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology, Tokyo, Japan, 15–18 September 2020; pp. 101–105.
6. Sousa, R.T.; Marques, O.; Soares, F.A.A.; Sene Jr, I.I.; de Oliveira, L.L.; Spoto, E.S. Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. *Procedia Comput. Sci.* **2013**, *18*, 2579–2582. [CrossRef]
7. Imran, B.; Hambali, H.; Bakti, L.D. Implementation of Machine Learning Model for Pneumonia Classification Based on X-ray Images. *J. Mantik* **2021**, *5*, 2101–2107.
8. Gabruseva, T.; Poplavskiy, D.; Kalinin, A. Deep learning for automatic pneumonia detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 350–351.
9. Toğaçar, M.; Ergen, B.; Cömert, Z.; Özyurt, F. A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *IRBM* **2020**, *41*, 212–222. [CrossRef]
10. Sourab, S.Y.; Kabir, M.A. A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. *Sens. Int.* **2022**, *3*, 100167. [CrossRef]
11. Sun, L.; Song, F.; Shi, N.; Liu, F.; Li, S.; Li, P.; Zhang, W.; Jiang, X.; Zhang, Y.; Sun, L.; et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *J. Clin. Virol.* **2020**, *128*, 104431. [CrossRef]
12. Franquet, T. Imaging of community-acquired pneumonia. *J. Thorac. Imaging* **2018**, *33*, 282–294. [CrossRef] [PubMed]
13. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.
14. Sirazitdinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; Ibragimov, B. Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database. *Comput. Electr. Eng.* **2019**, *78*, 388–399. [CrossRef]
15. Jaiswal, A.K.; Tiwari, P.; Kumar, S.; Gupta, D.; Khanna, A.; Rodrigues, J.J. Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement* **2019**, *145*, 511–518. [CrossRef]
16. Yao, L.; Poblenz, E.; Dagunts, D.; Covington, B.; Bernard, D.; Lyman, K. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv* **2017**, arXiv:1710.10501.
17. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
18. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
19. Cha, S.M.; Lee, S.S.; Ko, B. Attention-Based transfer learning for efficient pneumonia detection in chest X-ray images. *Appl. Sci.* **2021**, *11*, 1242. [CrossRef]
20. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
21. Li, J.; Wang, Y.; Wang, S.; Wang, J.; Liu, J.; Jin, Q.; Sun, L. Multiscale attention guided network for COVID-19 diagnosis using chest X-ray images. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1336–1346. [CrossRef] [PubMed]

22. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
23. Patel, S.C. Survey on Different Object Detection and Segmentation Methods. *Int. J. Innov. Sci. Res. Technol.* **2021**, *6*, 608–611.
24. Radiological Society of North America. RSNA Pneumonia Detection Challenge. Available online: www.kaggle.com/c/rsna-pneumonia-detection-challenge (accessed on 25 September 2022).
25. AI YanXiShe. Identification of X-ray Focus of Pneumonia. Available online: https://god.yanxishe.com/23 (accessed on 25 September 2022).
26. Masad, I.S.; Alqudah, A.; Alqudah, A.M.; Almashaqbeh, S. A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest X-ray images. *Int. J. Electr. Comput. Eng. (IJECE)* **2021**, *11*, 5530–5540. [CrossRef]
27. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; De Albuquerque, V.H.C. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl. Sci.* **2020**, *10*, 559. [CrossRef]