Research article

Region fine-grained attention network for accurate bone age assessment

Yamei Deng[1], **Ting Song**[1,*], **Xu Wang**[2], **Yonglu Chen**[1] and **Jianwei Huang**[1]

[1] Department of Radiology, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou 510150, China

[2] School of Automation, Guangdong University of Technology, Guangzhou 510006, China

* **Correspondence:** Email: flair@gzhmu.edu.cn.

Abstract: Bone age assessment plays a vital role in monitoring the growth and development of adolescents. However, it is still challenging to obtain precise bone age from hand radiography due to these problems: 1) Hand bone varies greatly and is always masked by the background; 2) the hand bone radiographs with successive ages offer high similarity. To solve such issues, a region fine-grained attention network (RFGA-Net) was proposed for bone age assessment, where the region aware attention (RAA) module was developed to distinguish the skeletal regions from the background by modeling global spatial dependency; then the fine-grained feature attention (FFA) module was devised to identify similar bone radiographs by recognizing critical fine-grained feature regions. The experimental results demonstrate that the proposed RFGA-Net shows the best performance on the Radiological Society of North America (RSNA) pediatric bone dataset, achieving the mean absolute error (MAE) of 3.34 and the root mean square error (RMSE) of 4.02, respectively.

Keywords: bone age assessment; region aware; fine-grained feature; attention network

1. Introduction

Worldwide, there are 6.1 million teenagers with developmental abnormalities annually, leading to great trouble for numerous families [1]. In this case, growth development monitoring will provide appropriate preventive measures and treatment plans for patients, further effectively preventing the deterioration [2]. Bone age assessment, one of the key means of monitoring physical development, can provide strong references for doctors to diagnose the endocrine system and other developmental-related diseases [3, 4]. Therefore, it is necessary to perform the bone age assessment to monitor adolescent development.

Currently, many methods are proposed for bone age assessment, including manual diagnosis and
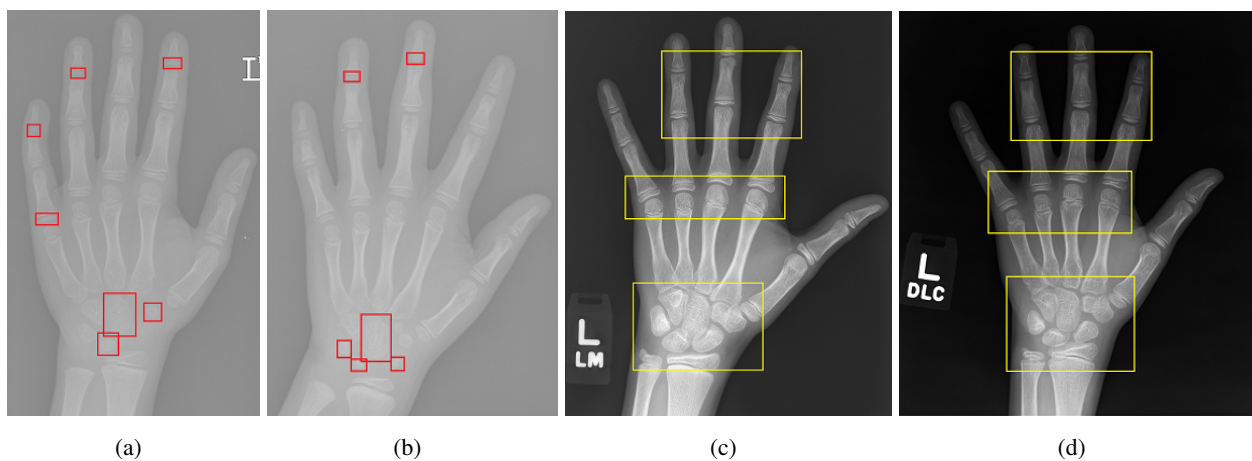
Figure 1. Examples of hand radiographs. (a) and (b) show the feature bones with significant changes. (c) and (d) describe the hand radiographs with successive bone age, where the characteristic regions are highly similar.

machine learning-based approaches. In manual diagnosis, the experienced doctors evaluate the bone age based on the reference standards, e.g., Greulich-Pyle (GP) [5], Tanner-Whitehouse (TW2) [6] and Radius-China (RUS-CHN) [7]. However, these methods have large errors and are time-consuming, leading to an overloaded workload for radiologists. For machine learning methods, thanks to the powerful feature extraction capabilities of deep learning techniques, they have already achieved or even surpassed the performance of manual diagnosis methods and are more efficient than radiologists [8–12]. For example, Wang et al. [14] developed a coarse and fine feature joint learning network to perform bone age assessment by capturing long-term dependencies and fusing attention features. However, it is not possible to eliminate the interference caused by the background, resulting in poor performance. Nguyen et al. [13] proposed a convolutional neural network (CNN) for bone age assessment using transfer learning from hand radiographs, but it is still far from available in clinical practice. In addition, Liu et al. [15] proposed a self-supervised attention network for bone age assessment using specific skeletal regions, e.g., the regions of interest for epiphysis and carpal, while ignoring other areas, such as the metacarpal bone which contains bone age information. Moreover, Ren et al. proposed a regression CNN to automatically assess the pediatric bone age using coarse attention and fine attention maps from hand radiograph [16]. However, similar hand radiographs may lead to poor separability, making it difficult for the networks to distinguish them. Furthermore, an attention-guided network is designed to automatically localize the discriminative regions for bone age assessment [17], and it performs well on images with significant bone age differences, but shows lower performance on images with successive bone ages.

In a word, it is still challenging to perform bone age assessment because of these problems. First, the skeletal regions vary greatly in shape and are prone to background interference [18], resulting in some effective feature information being submerged, as shown in Figure 1 (a),(b). Second, hand radiographs with successive bone age are often similar with poor separability [19, 20] [see Figure 1 (c) (150 months) and (d) (132 months)], leading to larger prediction errors and performance degradation of the networks.

To address such problems, a region fine-grained attention network (RFGA-Net) is proposed for
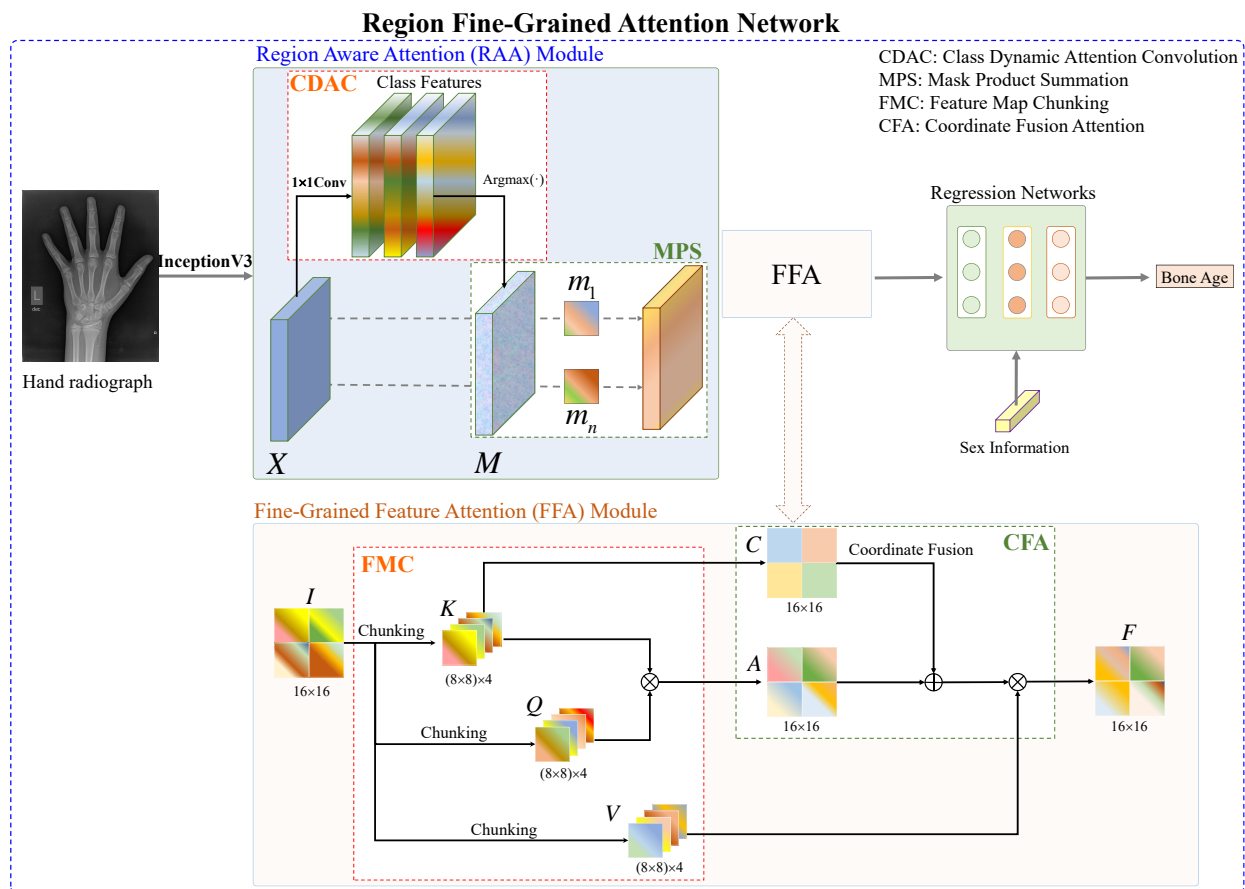
Figure 2. The architecture of the RFGA-Net. First, the RAA module is devised to distinguish the skeletal regions from the background, then the FFA module is developed to discriminate similar bone radiographs. Finally, the regression networks are employed to estimate the bone age by integrating the feature maps and sex information.

bone age assessment, where the region aware attention (RAA) module is developed to distinguish the skeletal regions from the background, and the fine-grained feature attention (FFA) module is devised to identify similar bone radiographs. The main contributions of this paper are described as follows:

1) The RAA module, which can model global spatial dependency, is developed to distinguish the specific skeletal regions from the background.

2) The FFA module, which can recognize critical fine-grained feature regions, is devised to identify similar bone radiographs.

3) The experimental results on the benchmark dataset show the superiority of the proposed RFGA-Net.

The rest of this paper is organized as follows: The proposed RFGA-Net is described in Section 2, while Section 3 gives the experiments and results and the conclusions are shown in Section 4.
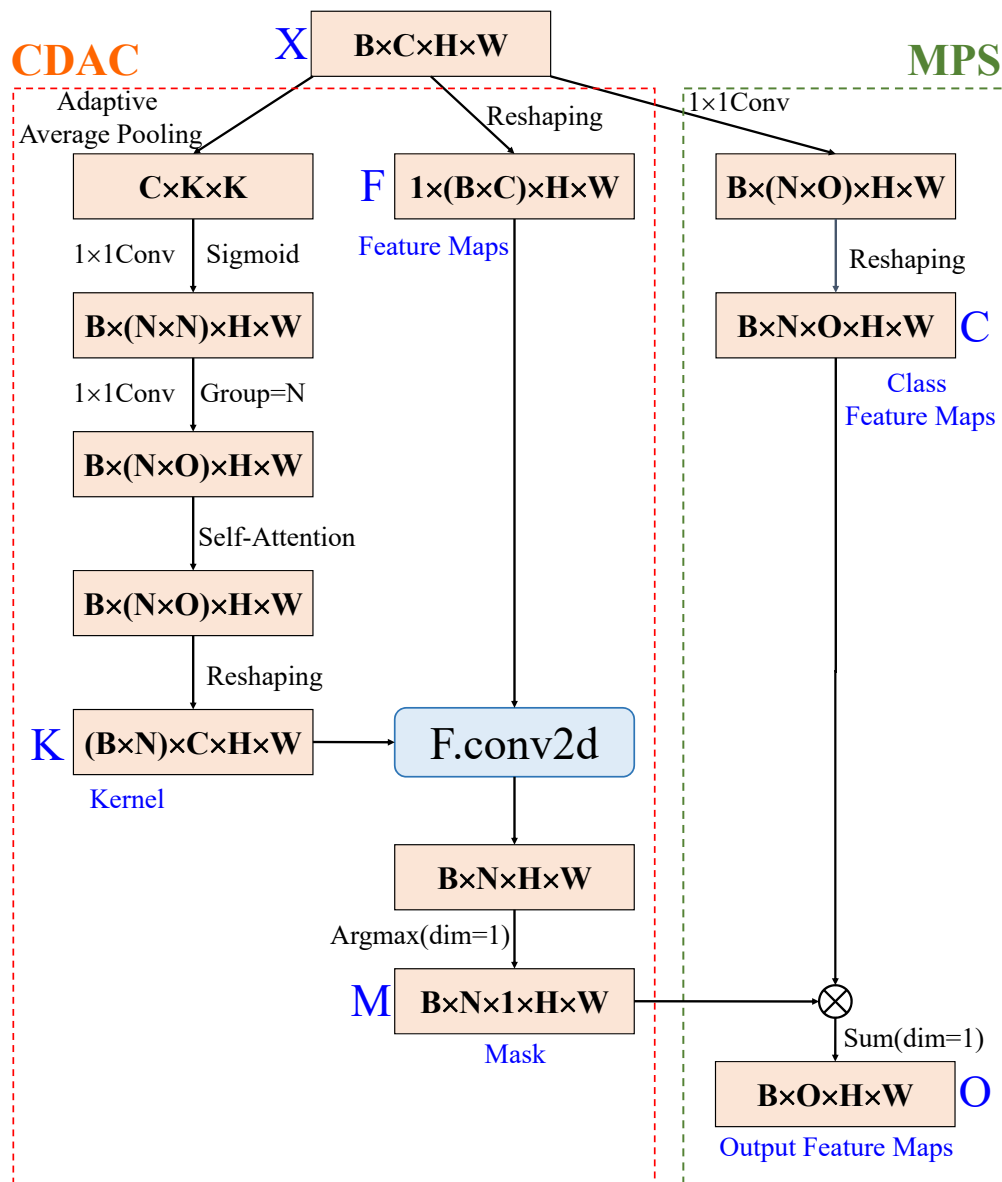
Figure 3. The architecture of the RAA module for global spatial dependency modeling.

2. RFGA-Net for bone age assessment

In this section, an RFGA-Net is proposed for bone age assessment, as shown in Figure 2, and it contains two key modules: The RAA module and the FFA module, which are described as follows.

2.1. Region aware attention module for skeletal region discrimination

In bone age diagnosis, doctors often determine bone age according to the specific skeletal regions, such as the size and shape of the metacarpal bones [21]. However, the large variations and diverse shapes of bones often cause visual fatigue [22]. In addition, some small bones are easily over-masked by the background, resulting in misdiagnosis for deep neural networks [23]. Correspondingly, the dynamic region convolution can adaptively register target feature regions and achieve target tracking

of different sizes [24]. Inspired by it, the RAA module was developed to distinguish the specific skeletal regions from the entire bone region by modeling global spatial dependency, where the class dynamic attention convolution was designed to track the different sized hand bones through adaptive feature extraction, as shown in Figure 2, defined as

$$RAA(X) = MPS[CDAC(X)], \tag{2.1}$$

where CDAC denotes the class dynamic attention convolution, and MPS represents the mask product summation, and the detailed operation is shown in Figure 3, which is described as follows:

- Similar to dynamic region-aware convolution for semantic segmentation [24], the input feature maps $X \in \mathbf{R}^{B \times C \times H \times W}$ are first fed into the CDAC to generate the class mask $M \in \mathbf{R}^{B \times N \times 1 \times H \times W}$, where the left branch is developed to generate the convolution kernel $K \in \mathbf{R}^{(B \times N) \times C \times H \times W}$ and the middle one is designed to obtain the class mask $M \in \mathbf{R}^{B \times N \times 1 \times H \times W}$.

  Specifically, in the left branch, the adaptive average pooling, 1×1 convolution layer and self-attention are introduced to generate the convolution kernel $K \in \mathbf{R}^{(B \times N) \times C \times H \times W}$, then the reshaped feature maps $F \in \mathbf{R}^{1 \times (B \times C) \times H \times W}$ are convoluted with the kernel to obtain the class mask $M \in \mathbf{R}^{B \times N \times 1 \times H \times W}$ after the argmax function, as shown in the middle one of Figure 3.

- The input feature maps $X \in \mathbf{R}^{B \times C \times H \times W}$ will be processed by the MPS to produce the output feature maps $O \in \mathbf{R}^{B \times O \times H \times W}$.

  Specifically, in the right branch, the input feature maps $X \in \mathbf{R}^{B \times C \times H \times W}$ are processed by the 1×1 convolution and then reshaped to obtain the class feature maps $C \in \mathbf{R}^{B \times N \times O \times H \times W}$. Finally, the class feature maps C will be multiplied by the mask to obtain the feature maps $O \in \mathbf{R}^{B \times O \times H \times W}$ after the summation at class dimension.

2.2. FFA module for similar bone radiograph distinguish

In clinical practice, the size and shape of hand bone areas are utilized to perform bone age assessment [25], but a large workload often leads to low efficiency [26]. Moreover, similar hand radiographs are always wrongly diagnosed by radiologists [27], resulting in significant errors between the predicted bone age and the actual one. Correspondingly, fine-grained feature perception can effectively mine dissimilation information from similar features for image recognition [28, 29]. Motivated by it, the FFA module is devised to distinguish similar bone radiographs by learning the fine-grained features, as shown in Figure 2, defined as

$$FFA(X) = CFA[FMC(X)], \tag{2.2}$$

where FMC and CFA denote the feature map chunking and coordinate fusion attention, respectively, which are shown in detail as follows:

- First, the input feature maps I are divided into identical small feature maps, i.e., FMC (as shown in Figure 2, the feature maps of 16×16 are reshaped into four small ones of 4×4), where the fine-grained features will be transmitted into the self-attention for learning.

- The chunking group feature maps are fed into the linear layer to generate the corresponding group queries (Q), keys (K) and values (V), respectively. Further, the coordinate coefficient position

relationship (C) generated from the keys (K) by the linear layer is utilized to model contextual dependencies and avoid spatial information loss caused by chunking.

- Next, the multiplied results of Q and K will be added with the coordinate information C to generate the new group feature maps.

- Finally, the values (V) will multiply with the new group feature map C to obtain the output feature maps.

In [28], the destruction and construction are proposed to focus on fine-grained details and learn robust and discriminative features, where the destruction is developed to obliterate some redundant regions to focus on other relevant parts, and the construction is designed to capture and understand the important features. In addition, it can be seen from [29] that the RefineMask is proposed to capture more detailed and precise information about the object boundaries and shapes, where the attention mechanism is used to selectively refine the mask predictions for each instance. Differentiated from them, the proposed FFA module is devised to distinguish similar bone radiographs by designing FMC and CFA, where the FMC is performed before attention learning, meaning the FFA can pay more attention to local features to mine the fine-grained features and the coordinate position relationship is utilized to model contextual dependencies and avoid spatial information loss.

3. Experiments and results analysis

3.1. Evaluation metric

In the experiments, we use the mean absolute error (MAE) and root mean square error (RMSE) as evaluation metrics to evaluate the performance of bone age assessment [16], defined as

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |\widehat{P}(n) - P(n)|, \tag{3.1}$$

and

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} [\widehat{P}(n) - P(n)]^2}, \tag{3.2}$$

where $P(n)$ and $\widehat{P}(n)$ represent the actual bone age and predicted one, respectively.

3.2. Dataset

To evaluate the effectiveness of the proposed RFGA-Net, the Radiological Society of North America (RSNA) pediatric bone dataset [30] is used as the benchmark, where it consists of the training set with 12,611 images, validation set with 1425 images and test set with 200 images. Furthermore, the RSNA dataset includes images with the ages from zero to 228 months, while the corresponding ground truth is provided for the public with the sex information. In the experiments, the same data partitioning is used for model testing, i.e., the RSNA dataset is divided into the training set, validation set, and testing set with the number of 12,611, 1425 and 200 images, respectively.

Table 1. Results of different methods for bone age assessment on the RSNA dataset.

Methods	MAE (month)	RMSE (month)
Spampinato et al. [33]	9.48	–
Deng et al. [34]	7.34	9.75
Bui et al. [35]	7.08	9.12
Wibisono et al. [36]	6.97	9.34
Zhou et al. [37]	–	6.00
Ren et al. [16]	5.20	–
Chen et al. [17]	4.40	4.78
Li et al. [38]	4.09	–
Liu et al. [15]	3.99	–
RFGA-Net	**3.34**	**4.02**

3.3. Implementation details

The proposed RFGA-Net is constructed by Pytorch 1.6.0 [31] on a server with two NVIDIA GeForce GTX 3090 GPUs. During the experiments, the stochastic gradient descent (SGD) optimizer [32] with an initial learning rate of $4e^4$ is used to optimize the proposed RFGA-Net, where the epoch, batch size and momentum are set as 100, 36, and 0.9, respectively. In addition, the L1Loss (absolute value loss) is regarded as the loss function while the learning rate is optimized by the StepLR (step learning rate) scheduler. Furthermore, all hand radiographs are resized to 576×576, then they are normalized by subtracting the mean of [0.485, 0.456, 0.406] and dividing the standard deviation of [0.229, 0.224 0.225], respectively.

3.4. Comparisons to other SOTA methods

To verify the superiority of the proposed RFGA-Net, nine SOTA (state-of-the-art) methods, i.e., Spampinato et al. [33], Deng et al. [34], Bui et al. [35], Wibisono et al. [36], Zhou et al. [37], Ren et al. [16], Chen et al. [17], Li et al. [38] and Liu et al. [15] are used for bone age assessment, and the results are recorded in Table 1, where the methods of Spampinato et al. [33] and Bui et al. [35] are tested on the digital hand atlas database*, while the model proposed by Zhou et al. [37] is evaluated on the tailored private hand radiograph dataset, and these approaches, i.e., Deng et al. [34], Wibisono et al. [36], Ren et al. [16], Chen et al. [17], Li et al. [38] and Liu et al. [15], are validated on the RSNA dataset.

It can be observed that the RFGA-Net shows the best performance on the RSNA dataset, achieving the MAE of 3.34 months and RMSE of 4.03 months, respectively, demonstrating the superiority of RFGA-Net for bone age assessment. In addition, the proposed RFGA-Net performs better than Ren et al. [16] and Liu et al. [15], indicating that the proposed region-aware and fine-grained feature attention can assist the self-attention mechanism to focus on the key bone regions, further improving the performance of the network. Moreover, compared to other region-specific guided methods, such as Spampinato et al. [33], Deng et al. [34], Bui et al. [35] and Wibisono et al. [36], the proposed
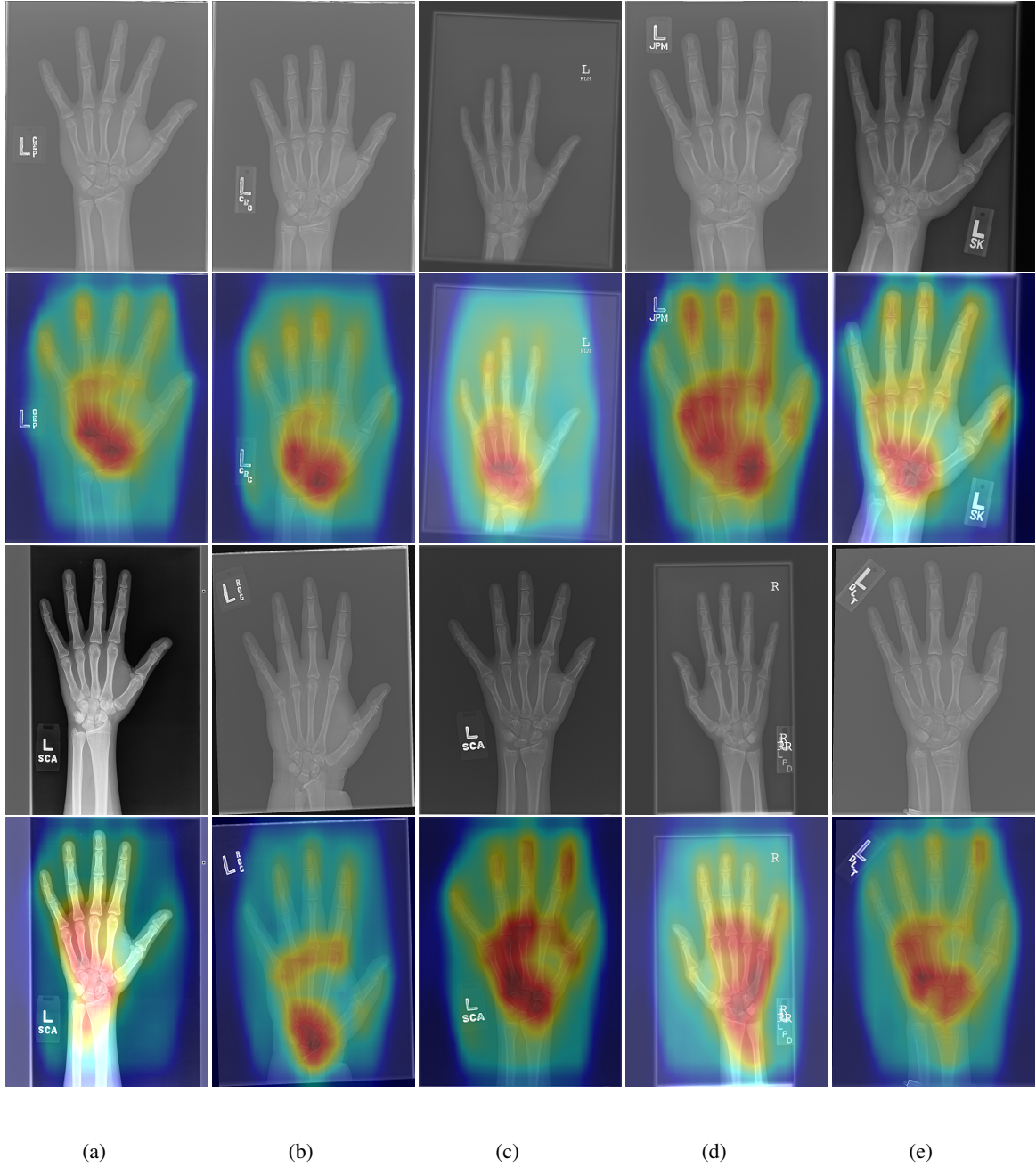
---

*http://www.ipilab.org/BAAweb/

Figure 4. The class activation map of the proposed RFGA-Net for bone age assessment, where the region with red color stands for high attention.
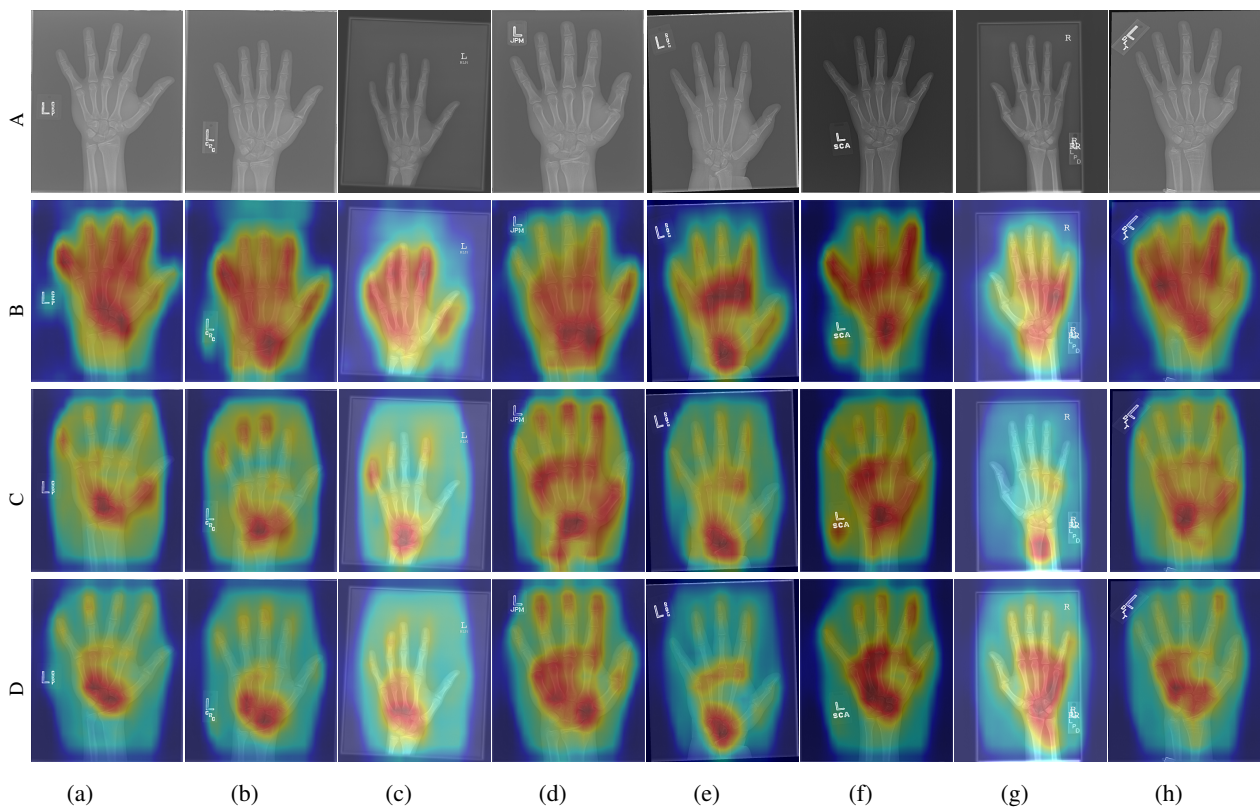
Figure 5. Class activation maps of the ablation experiment for the proposed module, where the red color denotes the high attention: A. raw images, B. baseline, C. baseline+RAA, D. baseline+RAA+FFA.

RFGA-Net also exhibits better performance because it can adaptively learn information from all bone feature regions.

In addition, the visualized results of the class activation map are given in Figure 4, where the first and third rows represent the raw hand radiography, and the second and last rows correspond to the class activation maps. From it, we can note that the proposed RFGA-Net mainly focuses on the regions of the epiphysis, metacarpal-phalanges, and carpal bones, which is consistent with the diagnosis of radiologists, demonstrating the effectiveness of the proposed RFGA-Net.

3.5. Ablation studies of the proposed modules in RFGA-Net

In this section, a large number of ablation studies are conducted to verify the effect of the proposed modules in RFGA-Net, including the RAA module and the FFA module, which are described as follows.

3.5.1. Impact of RAA module

As described in Section 2, the RAA module is devised to distinguish the specific skeletal regions from the entire hand bone. To validate the effect of the RAA module, several experiments are carried out on the RSNA dataset and the results are shown in Table 2, where No.1 is the baseline performance

Table 2. Ablation studies of the proposed modules for bone age assessment, where the baseline represents the InceptionV3 [39] integrating sex information.

No.	Baseline	RAA	FFA	MAE	RMSE
1	✓			5.34	6.85
2	✓	✓		4.19	5.61
3	✓		✓	4.49	5.83
4	✓	✓^{2}	✓^{1}	3.87	4.26
5	✓	✓^{1}	✓^{2}	3.34	4.02

Table 3. Ablation studies of the sex information for bone age assessment.

No.	Mode	MAE	RMSE
1	Baseline	5.76	7.19
2	**Baseline+Sex**	**5.34**	**6.85**
3	Baseline+RAA	4.83	6.18
4	**Baseline+RAA+Sex**	**4.19**	**5.61**
5	Baseline+FAA	4.94	6.54
6	**Baseline+FFA+Sex**	**4.49**	**5.83**
7	Baseline+RAA+FFA	3.88	4.29
8	**Baseline+RAA+FFA+Sex**	**3.34**	**4.02**

without the RRA and FFA module, and the baseline represents the backbone network of InceptionV3 [39]. From it, we can see that the MAE and RMSE can be reduced to 4.19 and 5.61, respectively, after inserting the RRA module into the baseline, showing that the RAA module is beneficial for bone age assessment. In addition, comparing No.3 with No.4 and No.5, we can note that the performance for bone age assessment can be further improved, validating the effectiveness of the proposed RAA module.

Moreover, several class activation maps for testing results are shown in Figure 5, where the first row (A) denotes the raw images, while the second one (B) and the third one (C) represent the baseline and baseline+RAA, respectively. We can observe from it that inserting the RAA module can further assist the network in focusing the bone regions in the hand radiograph, while the baseline network is affected by complex background interference (the entire hand bone region). These excellent results demonstrate that the proposed RAA module can distinguish the skeletal regions from the background, and further improve the performance of bone age assessment.

3.5.2. Impact of FFA module

Similarly, as illustrated in Section 2, the FFA module is employed to identify similar bone radiographs. To explore the effect of the FFA module, ablation experiments are carried out on the RSNA dataset and the ablation results are given in Table 2, where the proposed FFA module is embedded into the baseline+RAA to show its superiority for bone age assessment. From Table 2,

comparing the second row and the third one, we can note that the MAE and RMSE can be further reduced to 3.34 and 4.02, respectively, indicating that the proposed FFA module is helpful for estimating bone age based on the hand radiography. Similarly, it can be noted from No.2, No.4 and No.5 that adding the FAA module can further improve the performance of the network. For example, comparing No.2 with No.5, the MAE and RMSE can be reduced to 3.34 and 4.02, respectively, showing the superiority of the FAA module. Moreover, it can be seen that adding two modules shows better performance than the single module. For example, we can observe from No.2, No.3 and No.4 that the combination of RAA and FAA can achieve reductions of 0.32 and 0.62 on the MAE, respectively, compared to the single RAA and FAA modules, verifying the effectiveness of the RAA and FAA modules for bone age assessment.

Furthermore, several class activation maps are shown in Figure 5, where the last row denotes the results of baseline+RAA+FFA. It can be found that inserting the FFA module can help the network to locate the fine-grained feature regions, such as carpal and epiphysis regions, which are always ignored by radiologists. In contrast, the baseline+RAA performs feature extraction on the rough region of hand bones. This differentiated attention map proves that the proposed FFA module can distinguish similar bone radiographs by capturing fine-grained features.

3.5.3. Impact of order of RAA and FFA module

The order of the proposed RAA and FFA modules is also crucial for bone age assessment. To explore the effect of the order, the order of RAA before FFA and the opposite one is tested on the RSNA dataset. The results are listed in No.4 and No.5 of Table 2, where No.4 represents the order of the FFA module before the RAA module, and the opposite one is shown in No.5. From the results, we can find that the performance of No.5 is better than that of No.4. Thus, we can conclude that the order of RAA before FFA performs better than the opposite one, and it is used in the proposed RFGA-Net for bone age assessment.

3.5.4. Impact of sex information

Here, a series of experiments are performed to explore the impact of the sex information, where the sex information is added to the ablation experiments of the proposed modules. The predicted results for bone age are shown in Table 3. It can be seen from it that compared with the basic module, incorporating sex information can improve the performance of the network. For example, from No.3 and No.4, the MAE and RMSE can be reduced to 4.19 and 5.61 from 4.83 and 6.18, respectively after adding the sex information to Baseline+RAA. These substantial performance improvements demonstrate that adding sex information into the network can further advance the capability of bone age assessment. Therefore, the sex information is used as input for the proposed RFGA-Net for bone age assessment.

4. Conclusions

In this paper, an RFGA-Net was proposed for bone age assessment, where the RAA module was designed to distinguish the skeletal regions from the background and the FFA module was developed to identify similar bone radiographs. Extensive experiments conducted on the RSNA dataset showed

the superiority of the proposed RFGA-Net, further providing accurate references for radiologists to perform bone age assessment.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

Conflicts of interest

No potential conflict of interest.

References

1. M. L. Danielson, R. H. Bitsko, R. M. Ghandour, J. R. Holbrook, M. D. Kogan, S. J. Blumberg, Prevalence of parent-reported ADHD diagnosis and associated treatment among US children and adolescents, 2016, *J. Clin. Child Adolesc. Psychol.*, **47** (2018), 199–212. https://doi.org/10.1080/15374416.2017.1417860

2. S. E. Barlow, Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: Summary report, *Pediatrics*, **120** (2007), 164–192. https://doi.org/10.1542/peds.2007-2329C

3. S. L Truesdell, M. M Saunders, Bone remodeling platforms: Understanding the need for multicellular lab-on-a-chip systems and predictive agent-based models, *Math. Biosci. Eng.*, **17** (2020), 1233–1252. https://doi.org/10.3934/mbe.2020063

4. P. Hao, S. Chokuwa, X. Xie, F. Wu, J. Wu, C. Bai, Skeletal bone age assessments for young children based on regression convolutional neural networks, *Math. Biosci. Eng.*, **16** (2019), 6454–6466. https://doi.org/10.3934/mbe.2019323

5. K. C. Lee, K. H. Lee, C. H. Kang, K. S. Ahn, L. Y. Chung, J. J. Lee, Clinical validation of a deep learning-based hybrid (Greulich-Pyle and modified Tanner-Whitehouse) method for bone age assessment, *Korean J. Radiol.*, **22** (2021), 2017–2025. https://doi.org/10.3348/kjr.2020.1468

6. P. Lv, C. Zhang, Tanner–Whitehouse skeletal maturity score derived from ultrasound images to evaluate bone age, *Eur. Radiol.*, **2022** (2022), 1–8. https://doi.org/10.1007/s00330-022-09285-2

7. S. Zhang, L. Liu, The skeletal development standards of hand and wrist for Chinese ChildrenłChina 05 I. TW3-C RUS, TW3-C Carpal, and RUS-CHN methods, *Chin. J. Sports Med.*, **2023** (2023), 6–13.

8. Z. Ling, S. Yang, F. Gou, Z. Dai, J. Wu, Intelligent assistant diagnosis system of osteosarcoma MRI image based on transformer and convolution in developing countries, *IEEE J. Biomed. Health Inf.*, **26** (2022), 5563–5574. https://doi.org/10.1109/JBHI.2022.3196043

9. Y. Deng, X. Wang, Y. Liao, ASA-Net: Adaptive sparse attention network for robust electric load forecasting, *IEEE Internet Things J.*, (2023), 1–13. https://doi.org/10.1109/JIOT.2023.3300695

10. Q. H. Nguyen, R. Muthuraman, L. Singh, G. Sen, G. Sen, B. P. Nguyen, et al., Diabetic retinopathy detection using deep learning, in *Proceedings of the International Conference on Machine Learning and Soft Computing*, (2020), 103–107. https://doi.org/10.1145/3380688.3380709

11. Q. H. Nguyen, B. P. Nguyen, S. Dao, B. Unnikrishnan, R. Dhingra, S. R. Ravichandran, et al., Deep learning models for tuberculosis detection from chest X-ray images, in *Proceedings of the International Conference on Telecommunications (ICT)*, (2019), 381–385. https://doi.org/10.1109/ICT.2019.8798798

12. H. N. Pham, R. J. Tan, Y. T. Cai, S. Mustafa, N. C. Yeo, H. J. Lim, Automated grading in diabetic retinopathy using image processing and modified efficientnet, in *Proceedings of the International Conference on Computational Collective Intelligence (ICCCI)*, (2020), 505–515.

13. Q. H. Nguyen, B. P. Nguyen, M. T. Nguyen, M. C. Chua, T. T. Do, N. Nghiem, Bone age assessment and sex determination using transfer learning, *Expert Syst. Appl.*, **200** (2022), 1–11. https://doi.org/10.1016/j.eswa.2022.116926

14. X. Wang, W. Fan, M. Hu, Y. Wang, F. Ren, CFJLNet: Coarse and fine feature joint learning network for bone age assessment, *IEEE Trans. Instrum. Measure.*, **71** (2022), 1–11. https://doi.org/10.1109/TIM.2022.3193711

15. C. Liu, H. Xie, Y. Zhang, Self-supervised attention mechanism for pediatric bone age assessment with efficient weak annotation, *IEEE Trans. Med. Imaging*, **40** (2020), 2685–2697. https://doi.org/10.1109/TMI.2020.3046672

16. X. Ren, T. Li, X. Yang, S. Wang, S. Ahmad, L. Xiang, et al., Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph, *IEEE J. Biomed. Health Inf.*, **23** (2018), 2030–2038. https://doi.org/10.1109/JBHI.2018.2876916

17. C. Chen, Z. Chen, X. Jin, L. Li, W. Speier, C. W. Arnold, Attention-guided discriminative region localization and label distribution learning for bone age assessment, *IEEE J. Biomed. Health Inf.*, **26** (2021), 1208–1218. https://doi.org/10.1109/JBHI.2021.3095128

18. N. Li, B. Cheng, J. Zhang, A cascade model with prior knowledge for bone age assessment, *Appl. Sci.*, **12** (2022), 1–18. https://doi.org/10.55708/js0108002

19. S. Li, B. Liu, S. Li, X. Zhu, Y. Yan, D. Zhang, A deep learning-based computer-aided diagnosis method of X-ray images for bone age assessment, *Complex Intell. Syst.*, **8** (2021), 1929–1939. https://doi.org/10.1007/s40747-021-00376-z

20. A. A. Kasani, H. Sajedi, Hand bone age estimation using divide and conquer strategy and lightweight convolutional neural networks, *Eng. Appl. Artif. Intell.*, **120** (2023), 1–12. https://doi.org/10.1016/j.engappai.2023.105935

21. F. Cavallo, A. Mohn, F. Chiarelli, C. Giannini, Evaluation of bone age in children: A mini-review, *Front. Pediatr.*, **9** (2021), 1–11. https://doi.org/10.3389/fped.2021.580314

22. H. N. Tuan, N. D. Hai, N. T. Thinh, Shape prediction of nasal bones by digital 2D-photogrammetry of the nose based on convolution and back-propagation neural network, *Comput. Math. Methods Med.*, **2022** (2022), 1–18. https://doi.org/10.1155/2022/5938493

23. B. Dalisson, B. Charbonnier, A. Aoude, M. Gilardino, E. Harvey, N. Makhoul, Skeletal regeneration for segmental bone loss: Vascularised grafts, analogues and surrogates, *Acta Biomater.*, **136** (2021), 37–55. https://doi.org/10.1016/j.actbio.2021.09.053

24. J. Chen, X. Wang, Z. Guo, X. Zhang, J. Sun, Dynamic region-aware convolution, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 8064–8073.

25. L. Wang, Y. Mao, J. Xu, J. Wu, K. Wu, K. Mao, A ROI extraction method for wrist imaging applied in smart bone-age assessment system, *IEEE J. Biomed. Health Inf.*, (2023), 1–11. https://doi.org/10.1109/JBHI.2023.3284060

26. R. Alexander, S. Waite, M. Bruno, E. Krupinski, L. Berlin, Mandating limits on workload, duty, and speed in radiology, *Radiology*, **304** (2022), 274–282. https://doi.org/10.1148/radiol.212631

27. P. Agarwal, A. Jagati, S. Rathod, K. Kalra, S. Patel, Clinical features of mycetoma and the appropriate treatment options, *Res. Rep. Trop. Med.*, (2021), 173–179. https://doi.org/10.2147/RRTM.S282266

28. Y. Chen, Y. Bai, W. Zhang, T. Mei, Destruction and construction learning for fine-grained image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 5157–5166.

29. G. Zhang, X. Lu, J. Tan, J. Li, Refinemask: Towards high-quality instance segmentation with fine-grained features, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 6861–6869.

30. S. S. Halabi, L. M. Prevedello, J. Kalpathy-Cramer, A. B. Mamonov, A. Bilbily, M. Ciceroet, et al., The RSNA pediatric bone age machine learning challenge, *Radiology*, **290** (2019), 498–503. https://doi.org/10.1148/radiol.2018180736

31. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, Pytorch: An imperative style, high-performance deep learning library, in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2019, 8026–8037.

32. L. Bottou, F. E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, *Siam Review*, **60** (2018), 223–311. https://doi.org/10.1137/16M1080173

33. C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, R. Leonardi, Deep learning for automated skeletal bone age assessment in X-ray images, *Med. Image Anal.*, **36** (2017), 41–51. https://doi.org/10.1016/j.media.2016.10.010

34. Y. Deng, Y. Chen, Q. He, X. Wang, Y. Liao, J. Liu, et al., Bone age assessment from articular surface and epiphysis using deep neural networks, *Math. Biosci. Eng.*, **20** (2023), 13133–13148. https://doi.org/10.3934/mbe.2023585

35. T. D. Bui, J. Lee, J. Shin, Incorporated region detection and classification using deep convolutional networks for bone age assessment, *Artif. Intell. Med.*, **97** (2019), 1–8. https://doi.org/10.1016/j.artmed.2019.04.005

36. A. Wibisono, P. Mursanto, Multi Region-Based Feature Connected Layer (RB-FCL) of deep learning models for bone age assessment, *J. Big Data*, **7** (2020), 1–17. https://doi.org/10.1186/s40537-019-0278-0

37. X. Zhou, E. Wang, Q. Lin, G. Lin, W. Wu, K. Huang, Diagnostic performance of convolutional neural network-based Tanner-Whitehouse 3 bone age assessment system, *Quant. Imaging Med. Surg.*, **10** (2020), 657–667. https://doi.org/10.21037/qims.2020.02.20

38. X. Li, Y. Jiang, Y. Liu, J. Zhang, S. Yin, H. Luo, RAGCN: Region aggregation graph convolutional network for bone age assessment from X-ray images, *IEEE Trans. Instrum. Measure.*, **71** (2022), 1–12. https://doi.org/10.1109/TIM.2022.3190025

39. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2818–2826.