LowRank-CAM: A Computationally Efficient and Interpretable Framework for Medical Image Analysis (Student Abstract)

Gokaramaiah Thota, Nagaraju K, Sathya Babu Korra

Department of Computer Science and Engineering,
Indian Institute of Information Technology Design and Manufacturing Kurnool,
Jagannathagattu Hill, Kurnool-518 008, Andhra Pradesh, India.
322CS0001@iiitk.ac.in, knagaraju@iiitk.ac.in, ksb@iiitk.ac.in

Abstract

Deep learning has advanced medical imaging, but limited interpretability hinders clinical adoption. Class activation maps (CAM) provide visual explanations, yet methods such as Score-CAM are computationally expensive, requiring a forward pass for each activation map and limiting real-time applicability despite their high fidelity. To overcome this limitation, LowRank-CAM is proposed, which aggregates activation maps into a global matrix and applies singular value decomposition (SVD) to extract dominant spatial modes. The resulting top-r low-rank attention masks, with $r \ll K$ (r denotes the low-rank dimension and K is the total number of activation maps) replace per-channel perturbations and require only r forward passes through the classifier head. The resulting top-r low-rank attention masks, with $r \ll K$, replace per-channel perturbations and require only r forward passes through the classifier head. This low-rank formulation substantially reduces complexity while preserving class-discriminatory importance. Experiments on Inception-v3 musculoskeletal radiographs (MURA) demonstrate that LowRank-CAM achieves a $\mathbf{4.73}\times$ speedup over Score-CAM while maintaining comparable visual clarity and diagnostic relevance.

Introduction

Integration of deep learning into clinical decision making highlights the need for interpretability and computational efficiency (Sibilano et al. 2023; Thota et al. 2025). However, the non-linearity and high dimensionality of convolutional networks hinder reliable localization of class-discriminative regions. Existing explanation techniques face a trade-off: gradient-based methods are efficient but produce coarse maps, whereas perturbation-based methods yield sharper visual explanation at high computational cost (Thota, Karinagappa, and Korra 2024). For K activation maps, Score-CAM (Wang et al. 2020) requires $(K+1)$ forward passes, with complexity $\mathcal{O}((K+1)T_f + KC)$, where T_f is the forward pass cost and C is the mask pre-processing cost. This high computational demand limits its use in time-sensitive or resource-constrained settings.

LowRank-CAM addresses this issue by using SVD to exploit the low-rank structure of activation maps. A small set of dominant spatial modes is extracted to form low-rank attention masks $r \ll K$, replacing per-channel perturbations and reducing inference overhead while preserving class-discriminative detail. This enables fast and memory-efficient visual explanation generation, reducing the gap between interpretability and deployability in diagnostic applications. CAM methods have evolved to balance fidelity, robustness, and efficiency. Score-CAM (Wang et al. 2020) generates high-fidelity maps by perturbing each activation map and measuring class confidence. gScore-CAM (Chen et al. 2022) improves efficiency by selecting only the most prominent channels, while Score-CAM++ (Chen and Zhong 2022) improves robustness by focusing on positive relevance maps. Group-CAM (Zhang et al. 2021) aggregates channels into representative groups, FIMF-Score-CAM (Li et al. 2023) reduces redundancy through feature-space filtering, and Recipro-CAM (Sharma and Lee 2024) constructs visual explanation directly from feature-level masks without perturbations. Complementary approaches such as Eigen-CAM (Muhammad and Yeasin 2020), which projects activations onto dominant modes via SVD (Thota, Nagaraju, and Korra 2024).

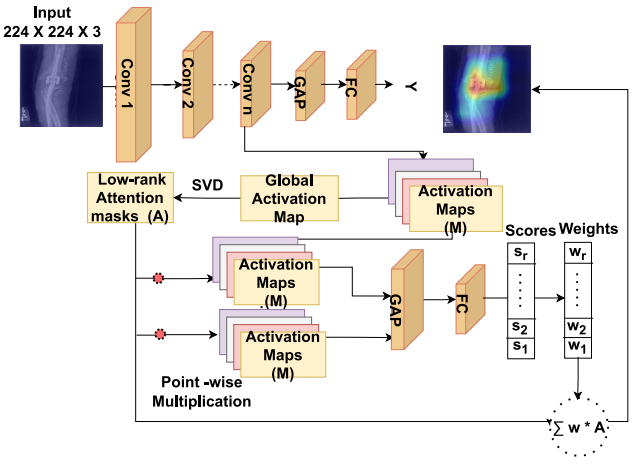

Figure 1: LowRank-CAM system architecture.

SL.	Method	AD↓	AIn↑	CH↑	CX↓	ADCC↑	Time(s)↓
1	CAM	6.89	31.25	87.05	43.23	75.30	**0.446**
2	Grad-CAM	6.51	43.75	82.57	43.13	74.27	0.554
3	Grad-CAM++	4.97	43.75	82.85	33.44	79.75	1.512
4	XGrad-CAM	5.51	43.75	82.57	43.13	74.48	0.732
5	Ablation-CAM	1.78	43.75	**89.44**	67.47	57.58	6.491
6	Eigen-CAM	1.75	43.75	86.91	28.38	84.16	0.442
7	Score-CAM	1.75	43.75	79.45	27.47	82.07	6.591
8	Score-CAM++	1.89	43.75	80.33	26.85	82.62	4.244
9	FIMF-Score-CAM	**0.95**	50.00	91.25	45.74	75.98	5.208
10	gScore-CAM	2.84	43.75	89.62	40.57	78.38	1.712
11	Group-CAM	6.25	50.00	77.02	41.27	73.75	1.526
12	Recipro-CAM	5.91	**62.50**	85.97	55.91	66.75	1.512
13	LowRank-CAM	1.60	43.75	89.01	**24.30**	**86.69**	1.392

Table 1: Comparison of CAM methods on MURA elbow with Inception-v3. Metrics include average drop (AD), average increase (AIn), coherency (CH), complexity (CX), ADCC, and computation time (s) per image.

Methodology

LowRank-CAM constructs a global activation matrix, applies SVD to extract dominant spatial patterns, and combines them with class-specific scores to produce cleaner and more efficient attention maps using only $r \ll K$ forward passes. The overall workflow is illustrated in Figure 1.

Step 1. Extract Activations From the final convolutional layer, extract the activation maps $\mathbf{M} \in \mathbb{R}^{L\times D\times K}$, where K is the number of channels. These activations contain spatial and channel responses that indicate important regions.

Step 2. Global Activation Matrix Flatten the spatial dimensions and stack all $\mathbf{M}$ to form $M^g \in \mathbb{R}^{(L\cdot D)\times K}$. This representation unifies spatial responses across channels, enabling global analysis.

Step 3. SVD Decomposition Apply SVD $M^g = U\Sigma V^\top$ and retain the top r components (σ_i, u_i, v_i). The channel filters (U), spatial patterns ($V^\top$), and their strengths (Σ).

Step 4. Low-rank Attention Mask Construction For each retained rank-r, the column $u_i \in \mathbb{R}^K$ of U and the row $v_i^\top \in \mathbb{R}^{L\cdot D}$ of $V^\top$. Form $\tilde{A}_i = \sigma_i u_i v_i^\top$, reshape to $(L \times D \times K)$, sum across channels, apply ReLU, and normalize: $A_i^{(sp)}(x,y) = \text{Norm}\left(\sum_{k=1}^K \tilde{\mathcal{A}}_i(x,y,k)\right)$. This produces clean low-rank attention masks that highlight consistent activation regions.

Step 5. Class Relevance Scoring Each low-rank attention mask is applied to the activation maps by dot product operation, followed by global average pooling (GAP) and fully connected (FC) layers to compute $w_i^c = f^{cls}(\mathbf{M}\odot A_i^{(sp)})[c]$, indicating the contribution of the mask to the class c.

Step 6. Final CAM Generation Combine masks using class-specific weights: $L_{\text{LowRank-CAM}}(c) = \sum_{i=1}^r w_i^c A_i^{(sp)}$. **Key Advantage:** LowRank-CAM reduces the cost from $\mathcal{O}((K+1)T_f)$ (Score-CAM) to $\mathcal{O}(KP^2+rT_h)$, where T_h is the inexpensive FC pass and SVD decomposition $O(KP^2)$, producing faster and coherent class-specific explanations.

Experiments and Results

The MURA dataset of musculoskeletal radiographs (MURA dataset) was used to evaluate LowRank-CAM. The evaluation of CAM-based visual explanations is assessed using four primary metrics: average drop (AD), average increase (AIn), coherence (CH), and complexity (CX), combined into a unified measure called average drop-complexity-coherency (ADCC) (Poppi et al. 2021). The LowRank-CAM rank was tuned using the Insertion and Deletion AUC on ten wrist samples, varying ranks from 5 to 45. The higher AUC for inserting and the lower AUC for deletion (that is, higher $1-$Deletion AUC) indicate better interpretability. As shown in Figure 2, the insert AUC peaks at rank 10 (0.8374), while the $1-$Deletion AUC is highest at rank 5 (0.4569). Balancing both metrics, rank 10 was chosen as optimal. Evaluation of CAM in the MURA dataset (elbow study type) Using Inception-v3 As presented in Table 1, LowRank-CAM achieved an ADCC of 86.69, outperforming Score-CAM (82.07) and Grad-CAM (74.27). It also exhibited low complexity (CX: 24.30) and reduced inference time to 1.392 seconds, providing a speedup of 4.73 $\times$ over Score-CAM (6.591 seconds).

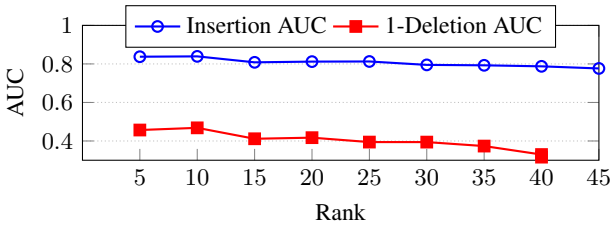

Figure 2: Insertion AUC and $1-$ Deletion AUC vs. rank.

Conclusion and Future Work

LowRank-CAM employs SVD to extract dominant spatial patterns from convolutional activations, reducing the complexity to $\mathcal{O}(KP^2+rT_h)$, where P is the spatial dimension (length $\times$ width) of the activation maps, r is the reduced rank and T_h the inexpensive classification head pass. This yields a substantial speed-up over Score-CAM while maintaining visual fidelity, with MURA evaluations showing superior coherency and ADCC. Future directions include adaptive rank selection, temporal consistency, and user-centered clinical studies.

Acknowledgments

The authors acknowledge the support of the Visvesvaraya PhD Scheme (Phase II), Ministry of Electronics and Information Technology (MeitY), Government of India, for providing a research fellowship to the first author.

References

Chen, P.; Wei, Y.; Shen, X.; Lin, Z.; and Zhang, J. 2022. gScoreCAM: Visual Explanation for CLIP via Gradient-Weighted Score-CAM. In *ACCV*.

Chen, Y.; and Zhong, G. 2022. Score-CAM++: Class discriminative localization with feature map selection. In *Journal of Physics: Conference Series*, volume 2278, 012018. IOP Publishing.

Li, M.; Xu, L.; Wang, Z.; and Zhang, Y. 2023. FIMF Score-CAM: Fast Score-CAM via Local Multi-Feature Integration. *IET Image Processing*, 17(3): 761–772.

Muhammad, M. B.; and Yeasin, M. 2020. Eigen-cam: Class activation map using principal components. In *IJCNN*, 1–7.

MURA dataset. 2018. MURA dataset. https://stanfordmlgroup.github.io/competitions/mura/. Accessed on 8-April-2023.

Poppi, S.; Cornia, M.; Baraldi, L.; and Cucchiara, R. 2021. Revisiting the evaluation of class activation mapping for explainability: A novel metric and experimental analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2299–2304.

Sharma, A.; and Lee, S. 2024. ReciproCAM: Lightweight Gradient-free CAM for Real-Time XAI. *arXiv:2401.12345*.

Sibilano, E.; Brunetti, A.; Buongiorno, D.; Lassi, M.; Grippo, A.; Bessi, V.; Micera, S.; Mazzoni, A.; and Bevilacqua, V. 2023. An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG. *Journal of Neural Engineering*, 20(1): 016048.

Thota, G.; Karinagappa, N.; and Korra, S. B. 2024. SVD-Grad-CAM: Singular Value Decomposition filtered Gradient Weighted Class Activation Map. In *ICPR*, 90–105. Springer.

Thota, G.; Nagaraju, K.; Babu, K. S.; and Pulabaigari, V. 2025. A Robust and Efficient Approach using Aggregated-Flexi Net for Interpretable Musculoskeletal Radiograph Classification. *Pattern Recognition*, 112703.

Thota, G.; Nagaraju, K.; and Korra, S. B. 2024. Quantitative assessment of class activation maps: an empirical study on musculoskeletal disorders. In *International Conference on Computer Vision and Image Processing*, 338–352. Springer.

Wang, H.; Wang, Z.; Du, M.; Yang, F.; Zhang, Z.; Ding, S.; Mardziel, P.; and Hu, X. 2020. Score-CAM: Score-weighted visual explanations for convolutional neural networks. In *CVF*, 24–25.

Zhang, J.; Chen, Y.; Lin, Z.; Shen, X.; Cao, Y.; and Zhang, J. 2021. Group-CAM: Group Score-Weighted Visual Explanations for Deep Convolutional Networks. *arXiv preprint arXiv:2103.13859*.