

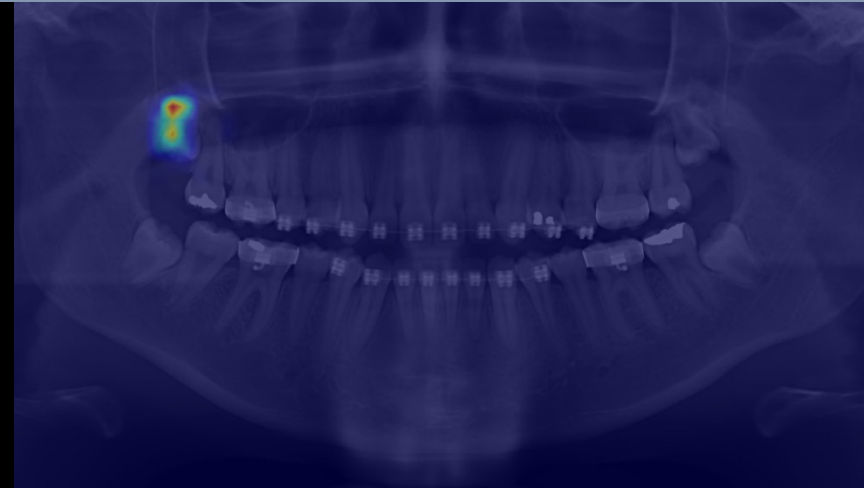
Leveraging CAM Algorithms for Explaining Medical Semantic Segmentation

Tillmann Rheude^{1*}, Andreas Wirtz², Stefan Wesarg², Arjan Kuijper^{1,2}

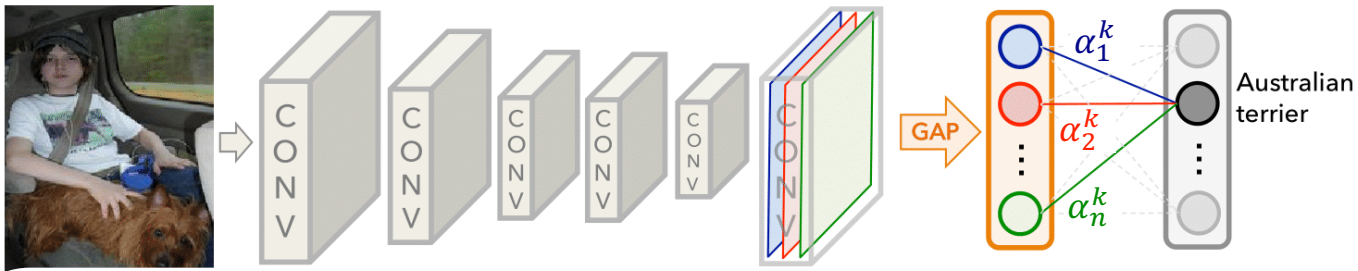
¹ TU Darmstadt, Darmstadt, Germany

² Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany

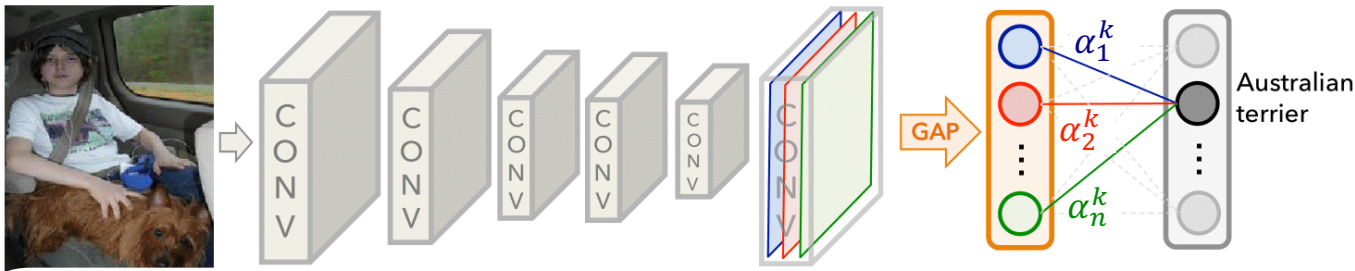
* meanwhile at Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany



CAM – Motivation with Classification Tasks

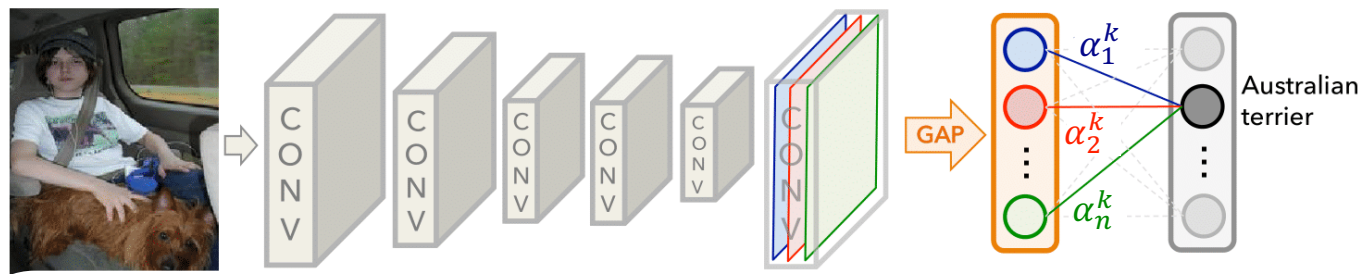


CAM – Motivation with Classification Tasks



Where did the
CNN look?

CAM – Motivation with Classification Tasks

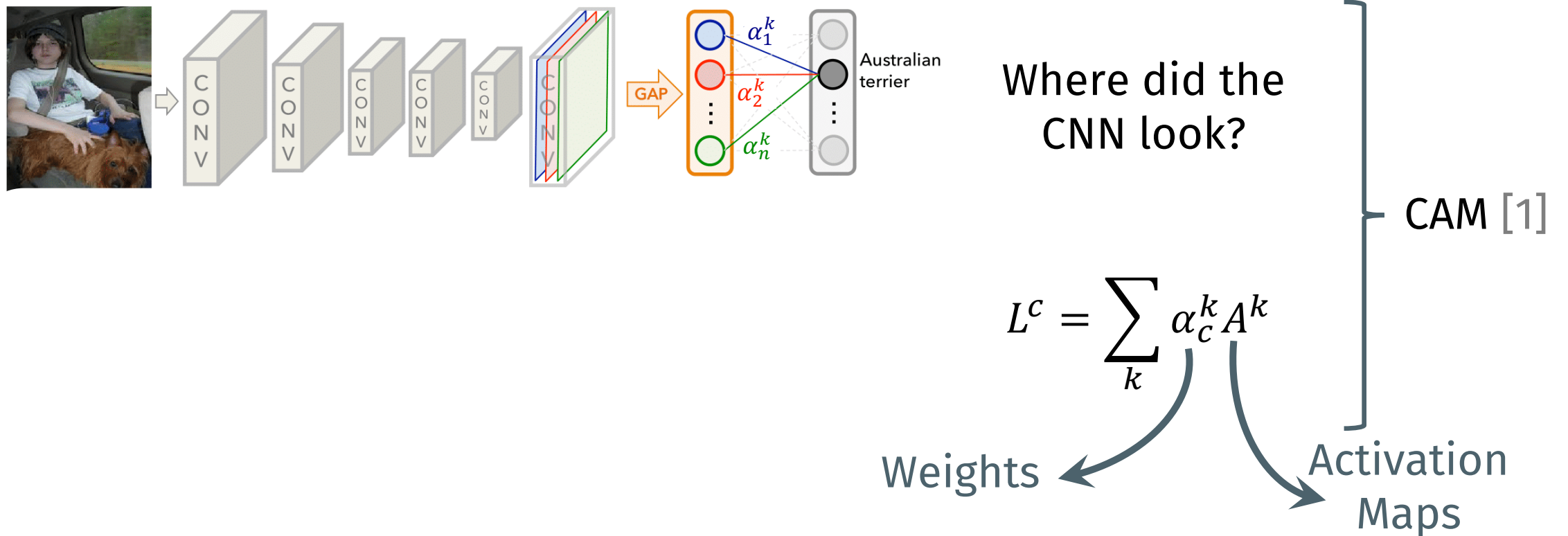


Where did the
CNN look?

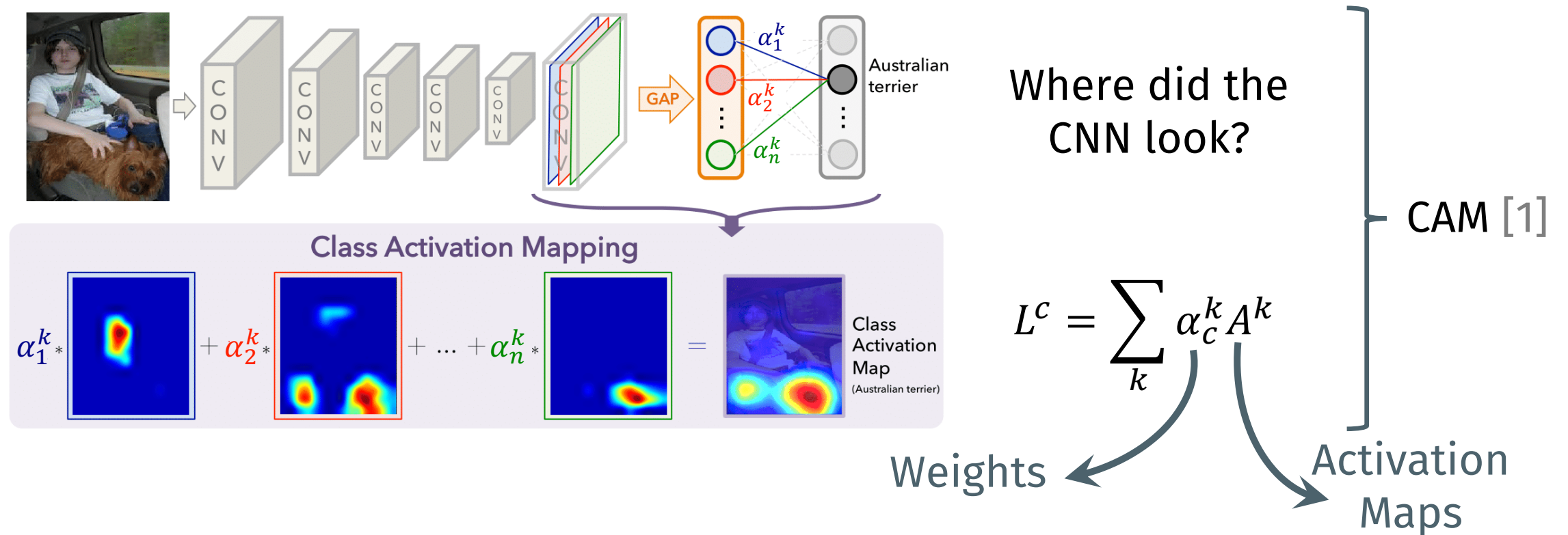
$$L^c = \sum_k \alpha_c^k A^k$$

CAM [1]

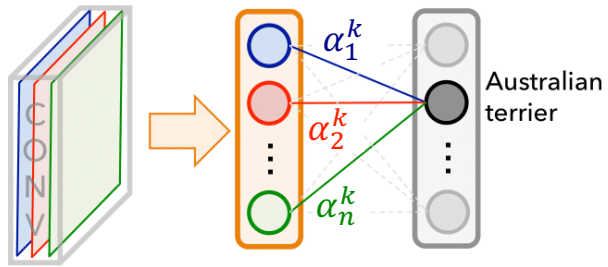
CAM – Motivation with Classification Tasks



CAM – Motivation with Classification Tasks



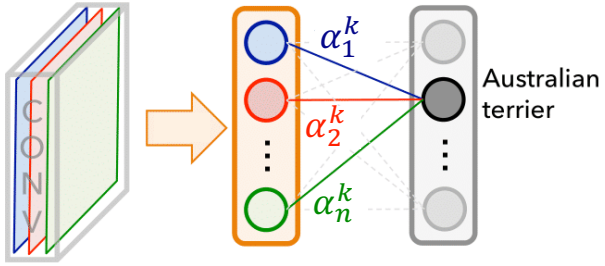
HiRes / Grad CAM – Extensions to CAM



$\alpha_c^k = \text{weights}$

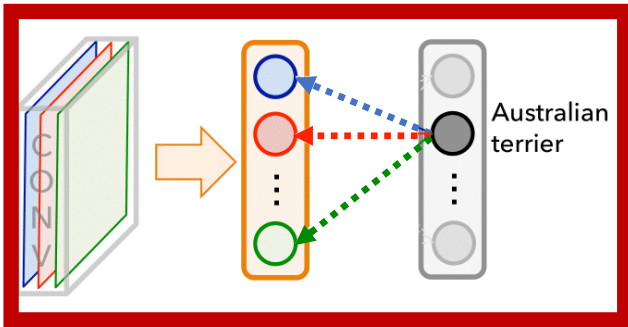
$$L_{CAM}^c = \sum_k \alpha_c^k A^k \quad \left. \vphantom{\sum_k} \right\} \text{CAM [1]}$$

HiRes / Grad CAM – Extensions to CAM



$\alpha_c^k = \text{weights}$

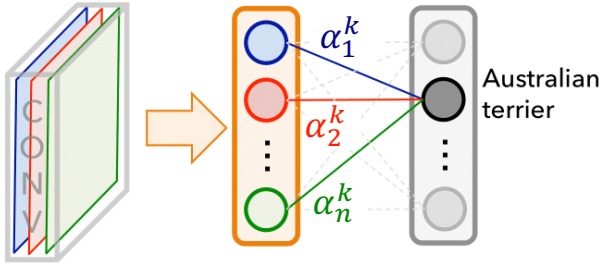
$$L_{CAM}^c = \sum_k \alpha_c^k A^k \quad \left. \vphantom{\sum_k} \right\} \text{CAM [1]}$$



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right) \quad \left. \vphantom{\sum_k} \right\} \text{Grad CAM [3]}$$

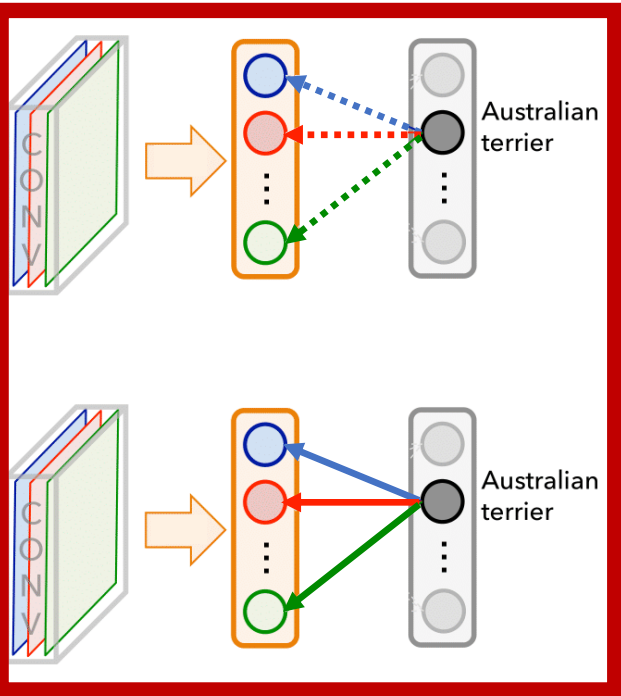
HiRes / Grad CAM – Extensions to CAM



$$\alpha_c^k = \text{weights}$$

$$L_{CAM}^c = \sum_k \alpha_c^k A^k$$

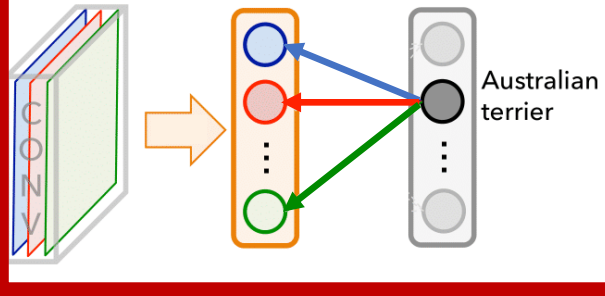
CAM [1]



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

Grad CAM [3]

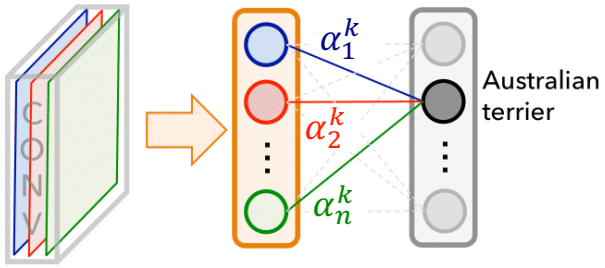


$$\alpha_c^k = \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

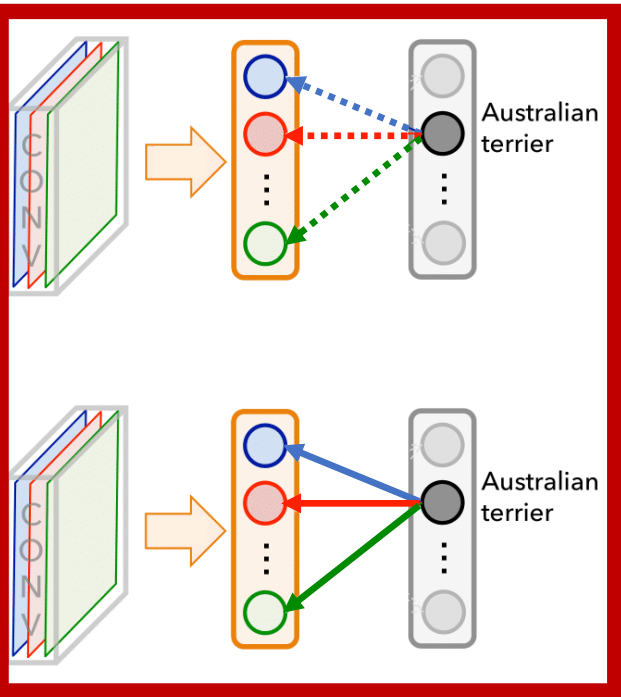
HiRes CAM [4]

HiRes / Grad CAM – Extensions to CAM



$$\alpha_c^k = \text{weights}$$

$$L_{CAM}^c = \sum_k \alpha_c^k A^k \quad \left. \vphantom{L_{CAM}^c} \right\} \text{CAM [1]}$$



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right) \quad \left. \vphantom{L^c} \right\} \text{Grad CAM [3]}$$

Fix inaccuracies

$$\alpha_c^k = \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right) \quad \left. \vphantom{L^c} \right\} \text{HiRes CAM [4]}$$

Overview – CAMs for Classification Tasks

Classification

CAM [1]

Grad CAM [3]

Grad CAM++ [9]

XGrad CAM [10]

...

Eigen CAM [11]

Layer CAM [12]

HiRes CAM [4]

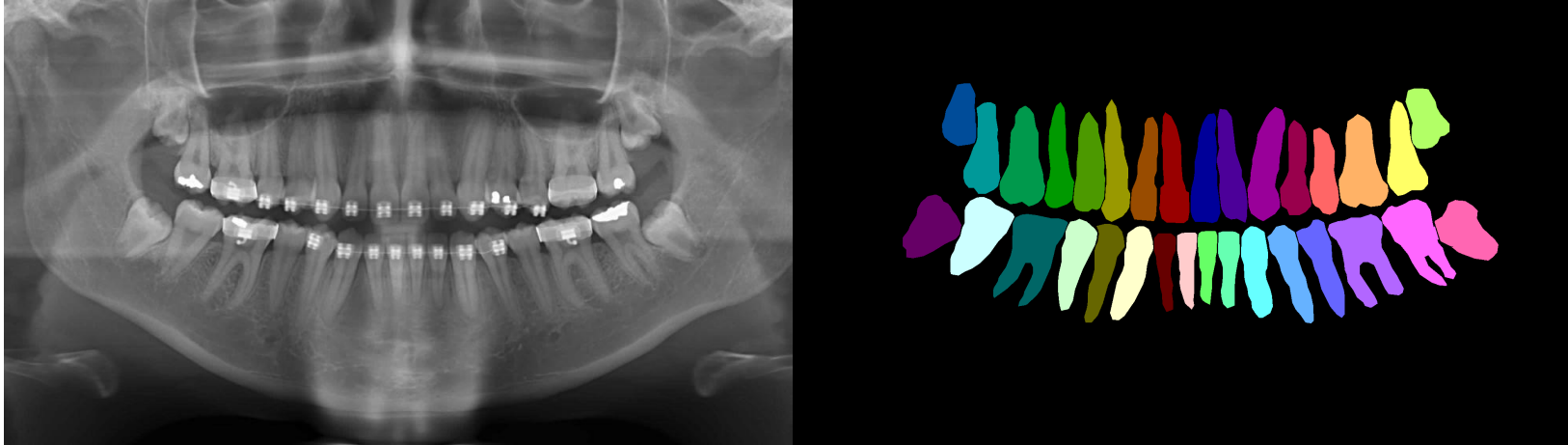
Overview – CAMs for Classification and Segmentation Tasks

Classification	Segmentation
CAM [1]	<i>n/a</i>
Grad CAM [3]	Seg-Grad CAM [2]
Grad CAM++ [9]	<i>n/a</i>
XGrad CAM [10]	<i>n/a</i>
...	...
Eigen CAM [11]	<i>n/a</i>
Layer CAM [12]	<i>n/a</i>
HiRes CAM [4]	<i>n/a</i>

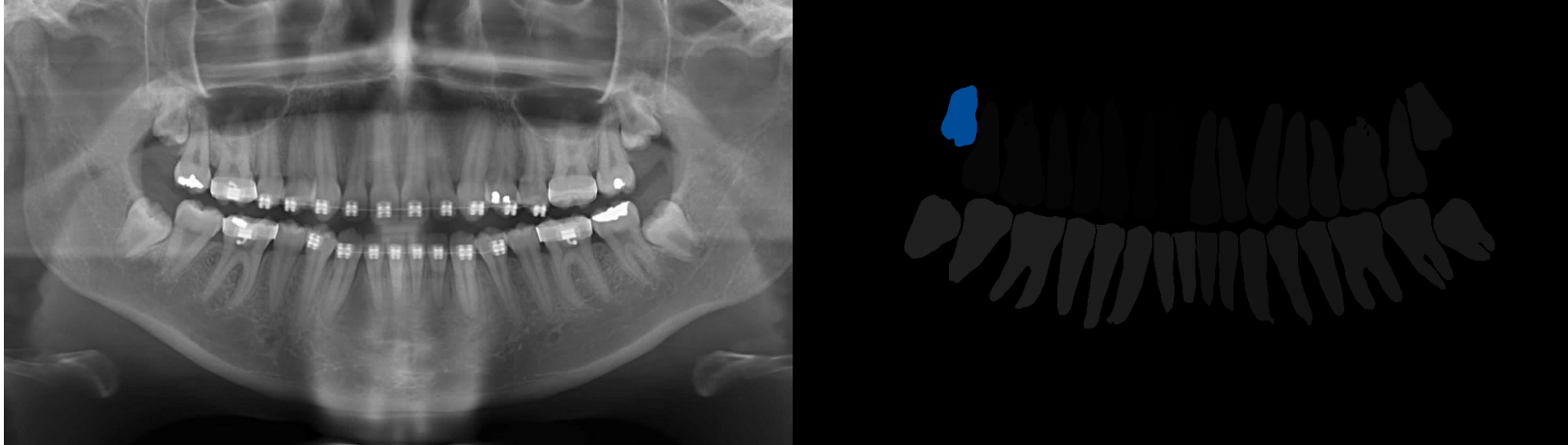
Overview – CAMs for Classification and Segmentation Tasks

Classification	Segmentation
CAM [1]	<i>n/a</i>
Grad CAM [3]	Seg-Grad CAM [2]
Grad CAM++ [9]	<i>n/a</i>
XGrad CAM [10]	<i>n/a</i>
...	...
Eigen CAM [11]	<i>n/a</i>
Layer CAM [12]	<i>n/a</i>
HiRes CAM [4]	<i>n/a</i>

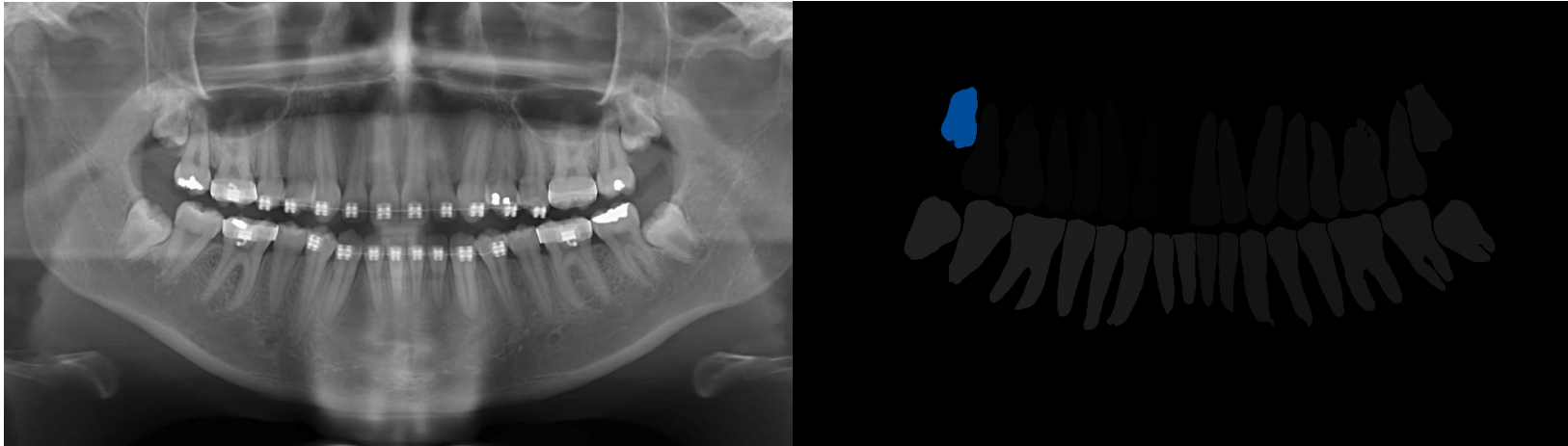
Seg-Grad CAM – Transfer to Segmentation Tasks



Seg-Grad CAM – Transfer to Segmentation Tasks

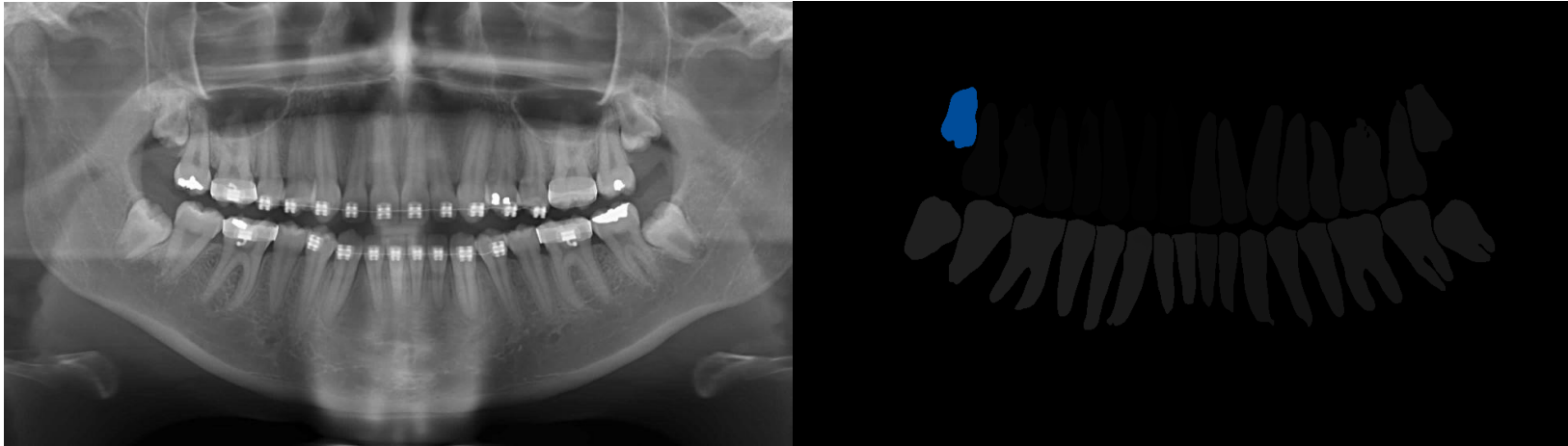


Seg-Grad CAM – Transfer to Segmentation Tasks



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

Seg-Grad CAM – Transfer to Segmentation Tasks



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

Output for a certain class

Seg-Grad CAM – Transfer to Segmentation Tasks

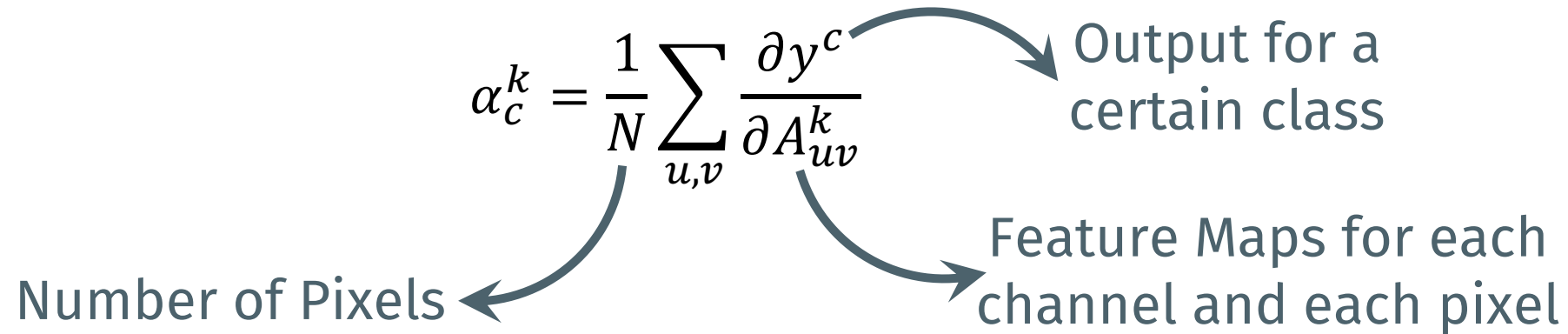
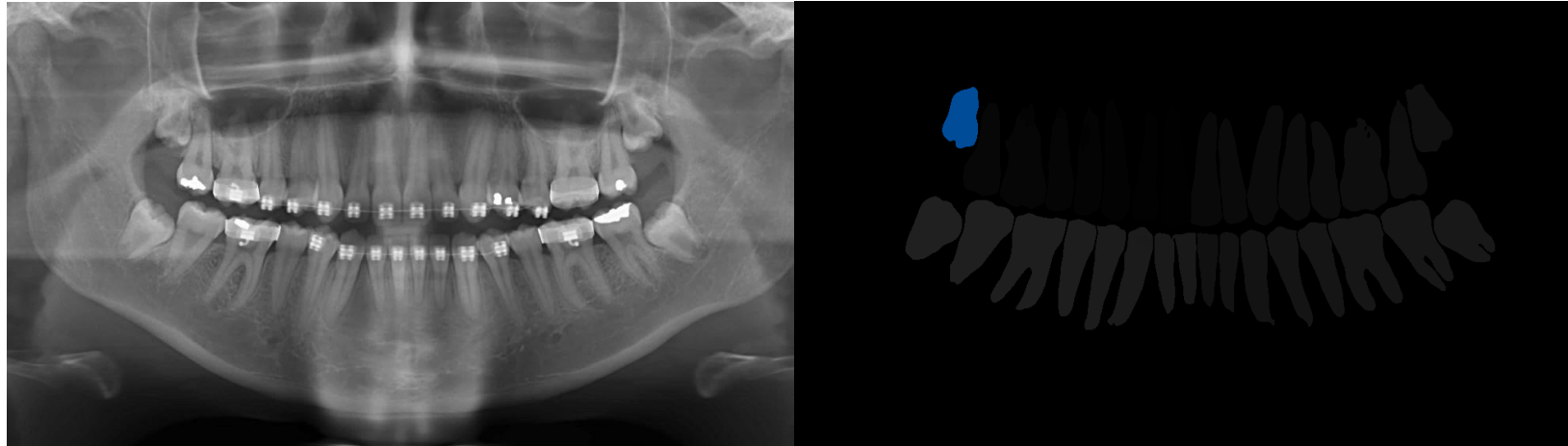


$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

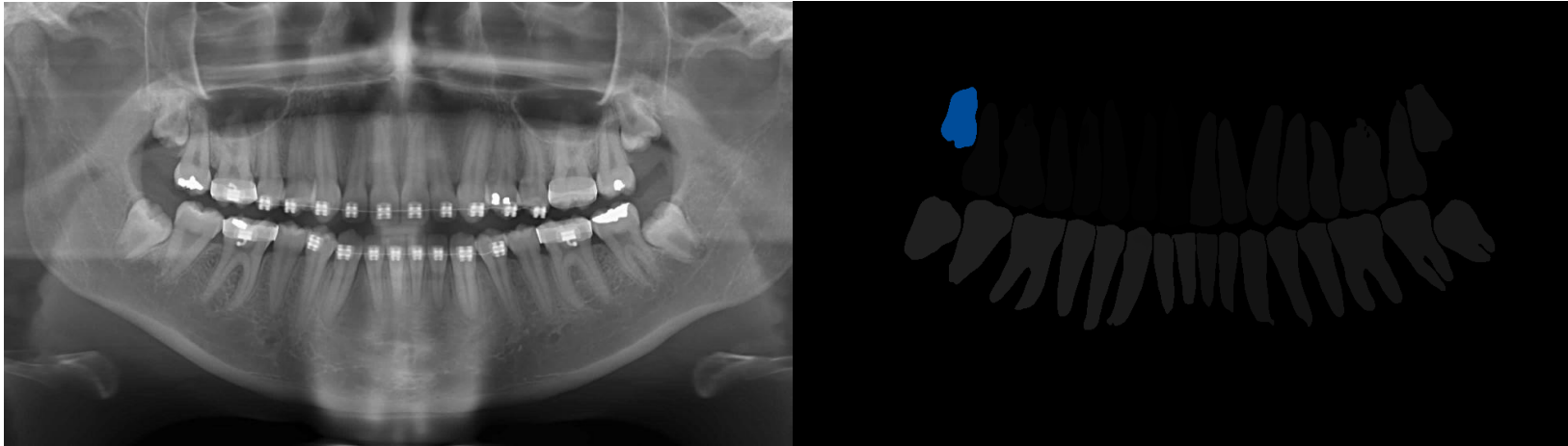
Output for a certain class

Feature Maps for each channel and each pixel

Seg-Grad CAM – Transfer to Segmentation Tasks



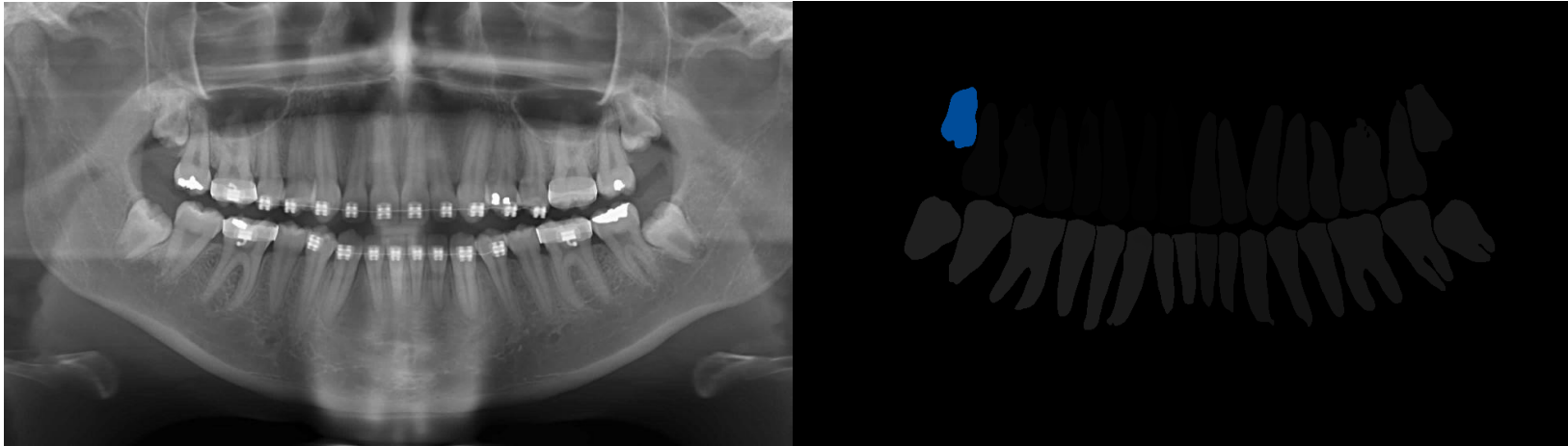
Seg-Grad CAM – Transfer to Segmentation Tasks



$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$y^{c,new} = \sum_{i,j \in \mathcal{M}} y_{i,j}^c$$

Seg-Grad CAM – Transfer to Segmentation Tasks

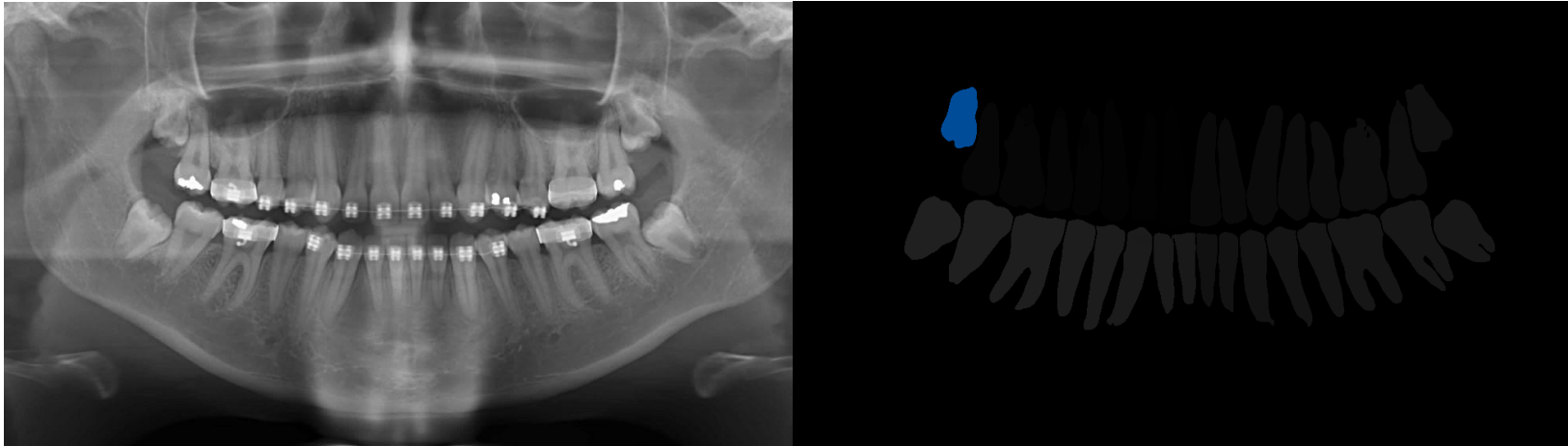


$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$y^{c,new} = \sum_{i,j \in \mathcal{M}} y_{i,j}^c$$

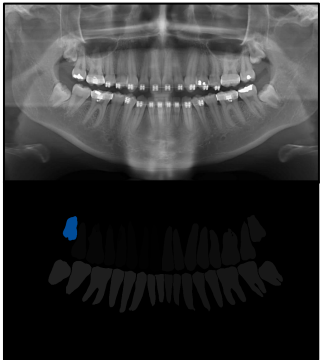
Set of Pixel Indices

Seg-Grad CAM – Transfer to Segmentation Tasks



$$\left. \begin{aligned} \alpha_c^k &= \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k} \\ y^{c,new} &= \sum_{i,j \in \mathcal{M}} y_{i,j}^c \end{aligned} \right\} \alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A_{uv}^k}$$

Seg-Grad CAM – Transfer to Segmentation Tasks



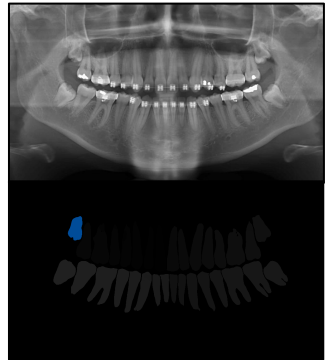
Input and
Pixelset \mathcal{M}

$$\begin{bmatrix} 1.25 \\ 0 \end{bmatrix}$$

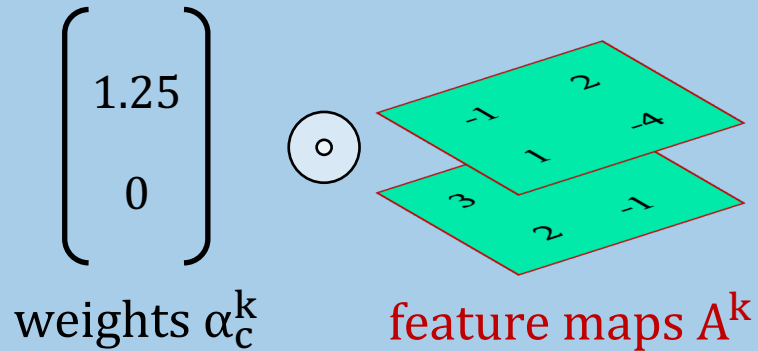
weights α_c^k

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

Seg-Grad CAM – Transfer to Segmentation Tasks

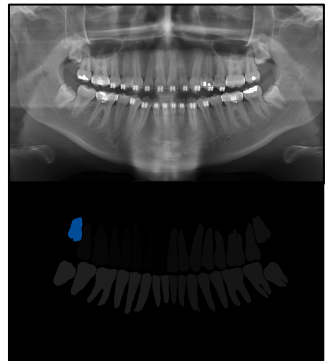


Input and
Pixelset \mathcal{M}

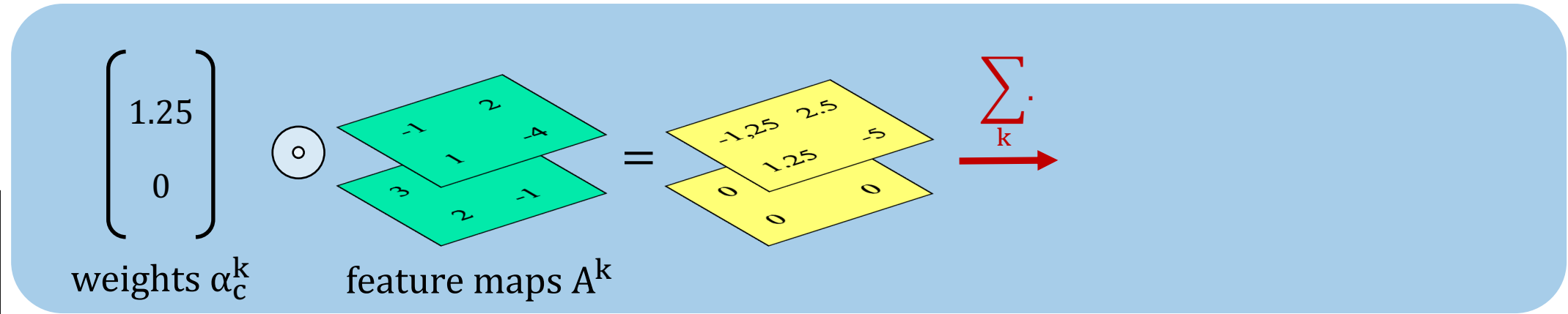


$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

Seg-Grad CAM – Transfer to Segmentation Tasks

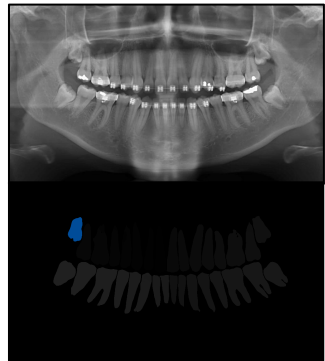


Input and
Pixelset \mathcal{M}

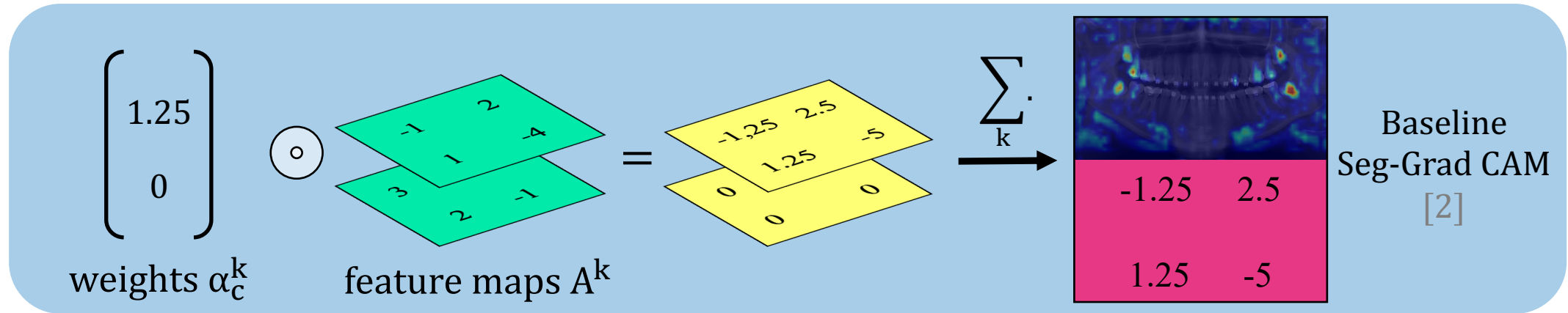


$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

Seg-Grad CAM – Transfer to Segmentation Tasks



Input and Pixelset \mathcal{M}



$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

* ReLU not visualized to ensure a better comparability later on

Overview – CAMs for Classification and Segmentation Tasks

Classification	Segmentation
CAM [1]	<i>n/a</i>
Grad CAM [3]	Seg-Grad CAM [2]
Grad CAM++ [9]	<i>n/a</i>
XGrad CAM [10]	<i>n/a</i>
...	...
Eigen CAM [11]	<i>n/a</i>
Layer CAM [12]	<i>n/a</i>
HiRes CAM [4]	<i>n/a</i>

Overview – CAMs for Classification and Segmentation Tasks

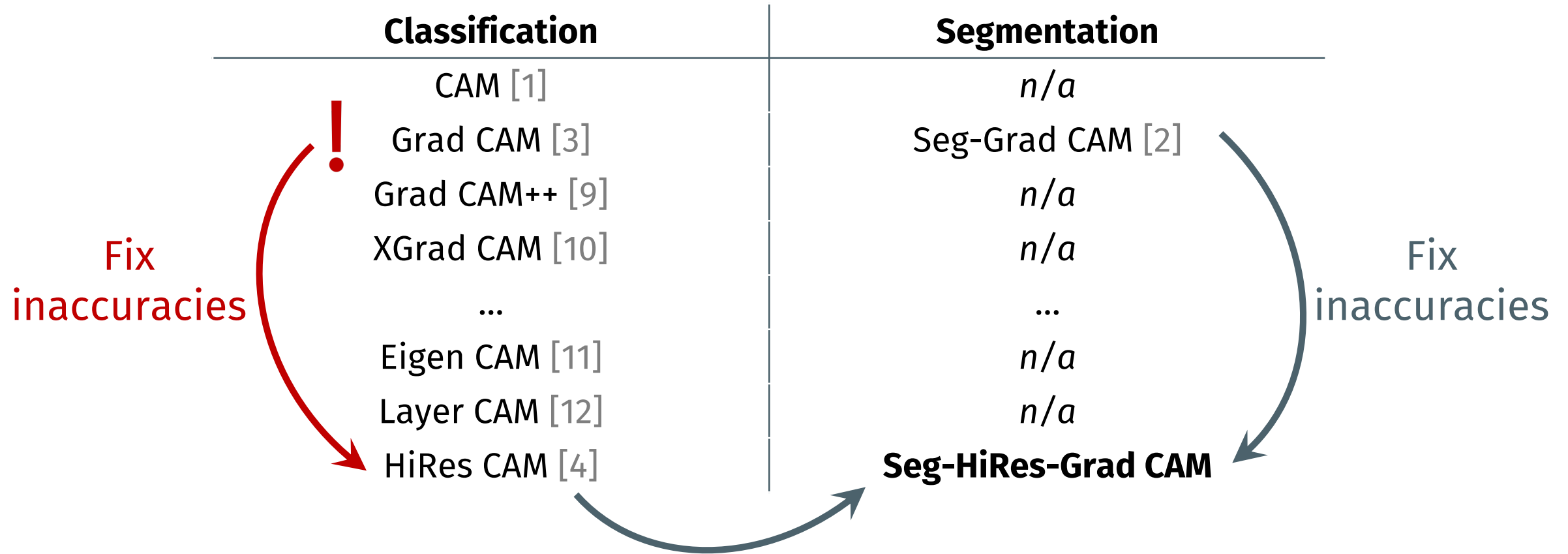
Classification	Segmentation
CAM [1]	<i>n/a</i>
! Grad CAM [3]	Seg-Grad CAM [2]
Grad CAM++ [9]	<i>n/a</i>
XGrad CAM [10]	<i>n/a</i>
...	...
Eigen CAM [11]	<i>n/a</i>
Layer CAM [12]	<i>n/a</i>
HiRes CAM [4]	<i>n/a</i>

Overview – CAMs for Classification and Segmentation Tasks

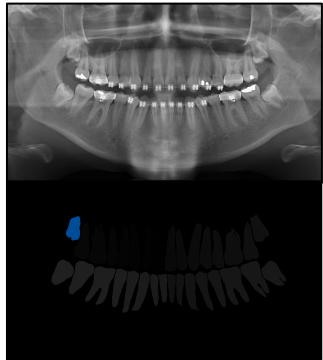
	Classification	Segmentation
	CAM [1]	<i>n/a</i>
Fix inaccuracies	! Grad CAM [3]	Seg-Grad CAM [2]
	Grad CAM++ [9]	<i>n/a</i>
	XGrad CAM [10]	<i>n/a</i>

	Eigen CAM [11]	<i>n/a</i>
	Layer CAM [12]	<i>n/a</i>
	HiRes CAM [4]	<i>n/a</i>

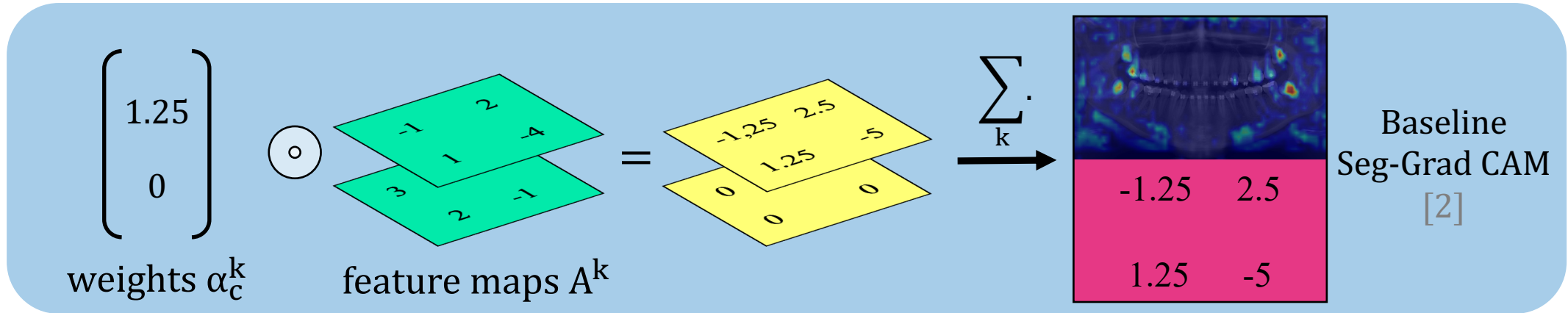
Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



Input and Pixelset \mathcal{M}



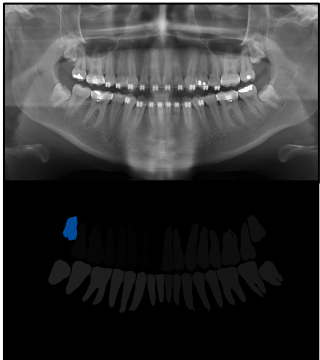
$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

$$y^{c,new} = \sum_{i,j \in \mathcal{M}} y_{i,j}^c$$

$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A_{uv}^k}$$

* ReLU not visualized to ensure a better comparability later on

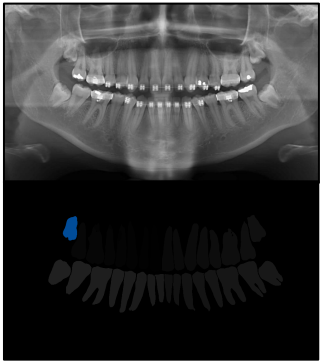
Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



Input and
Pixelset \mathcal{M}

$$\left. \begin{aligned} \alpha_c^k &= \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k} \\ y^{c,new} &= \sum_{i,j \in \mathcal{M}} y_{i,j}^c \end{aligned} \right\} \alpha_c^k = \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A^k}$$

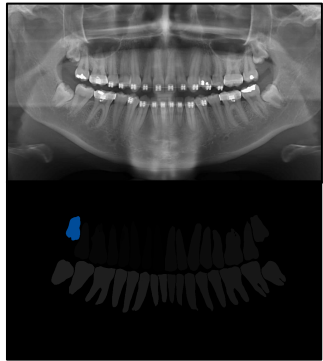
Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



Input and
Pixelset \mathcal{M}

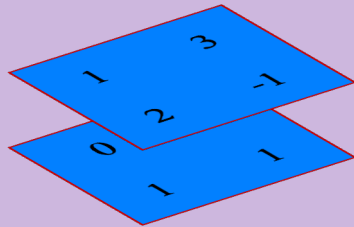
$$\left. \begin{aligned} \alpha_c^k &= \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k} \\ y^{c,new} &= \sum_{i,j \in \mathcal{M}} y_{i,j}^c \end{aligned} \right\} \alpha_c^k = \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A^k}$$

Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



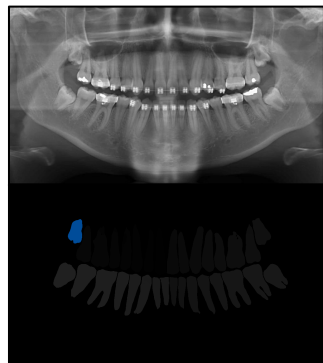
Input and Pixelset \mathcal{M}

$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$
$$y^{c,new} = \sum_{i,j \in \mathcal{M}} y_{i,j}^c$$
$$\alpha_c^k = \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A^k}$$



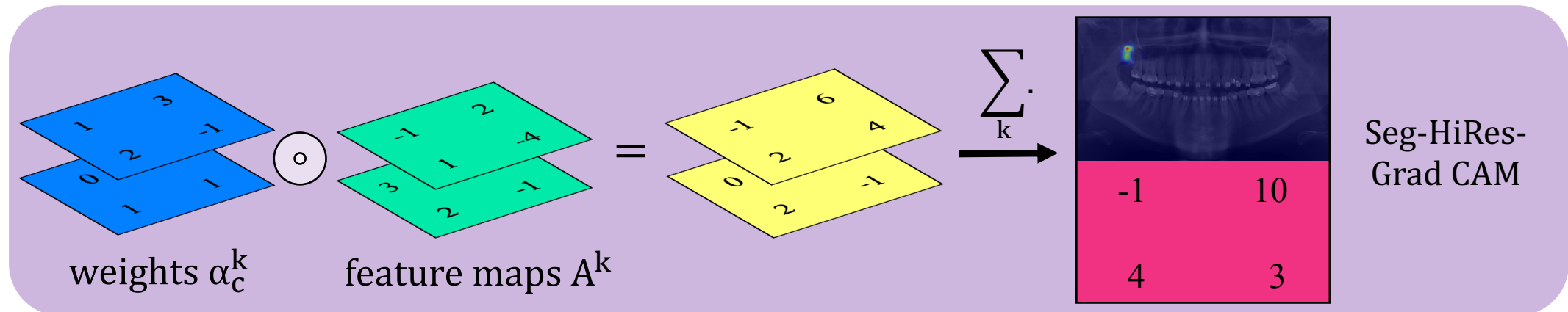
weights α_c^k

Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



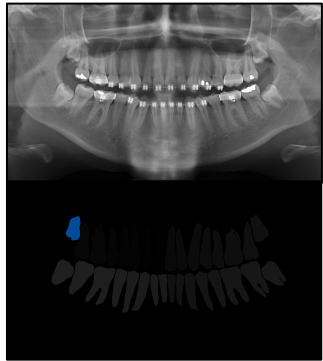
Input and Pixelset \mathcal{M}

$$L^c = \text{ReLU} \left(\sum_k \alpha_c^k A^k \right)$$

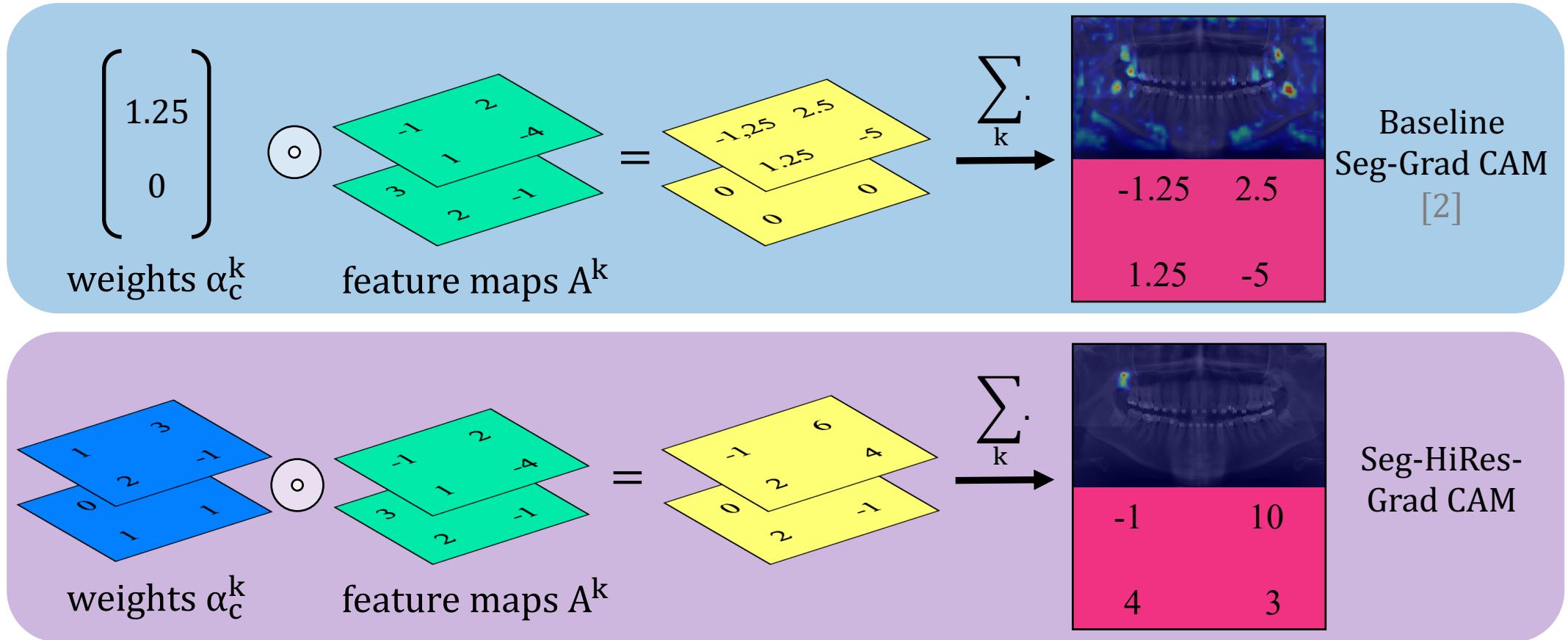


* ReLU not visualized to ensure a better comparability later on

Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



Input and Pixelset \mathcal{M}



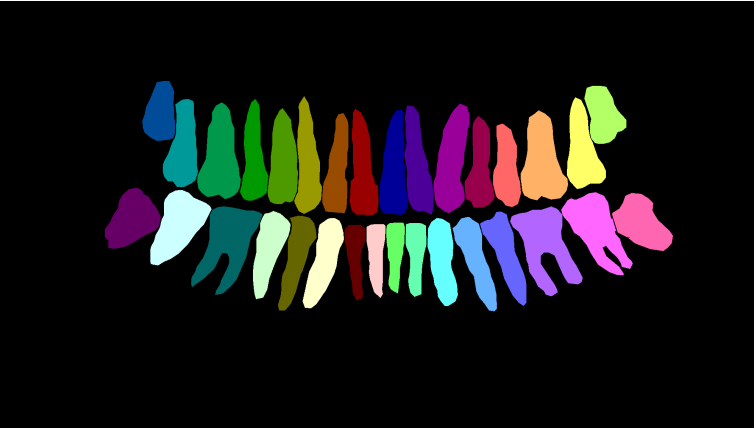
* ReLU not visualized to ensure a better comparability

Examples with OPG [5]

Input Image



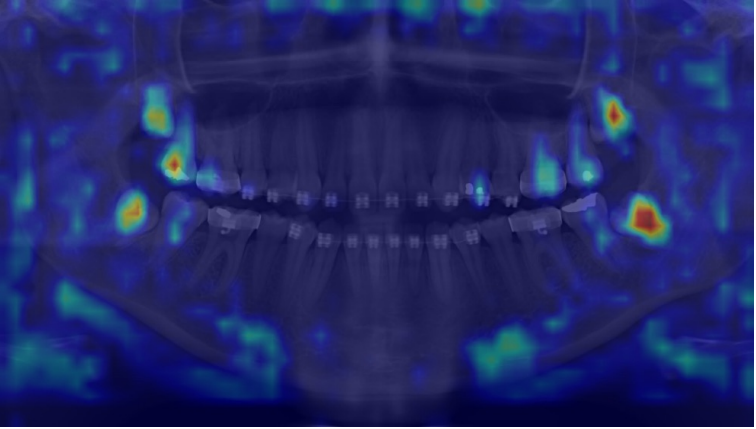
Ground Truth



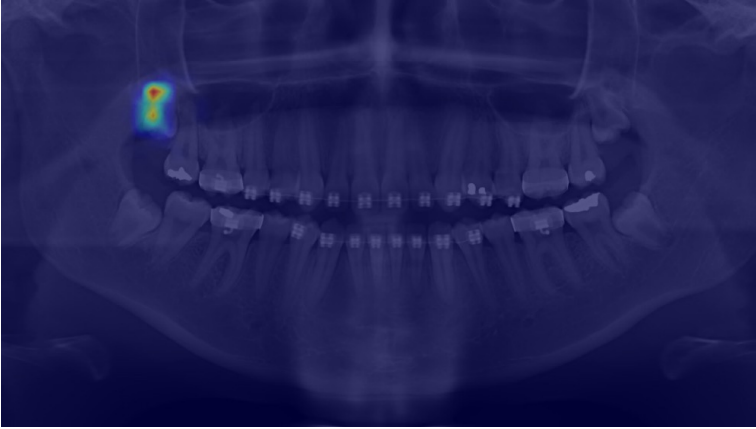
Prediction with U-Net [8]



Segmentation to be explained



Seg-Grad CAM [2]



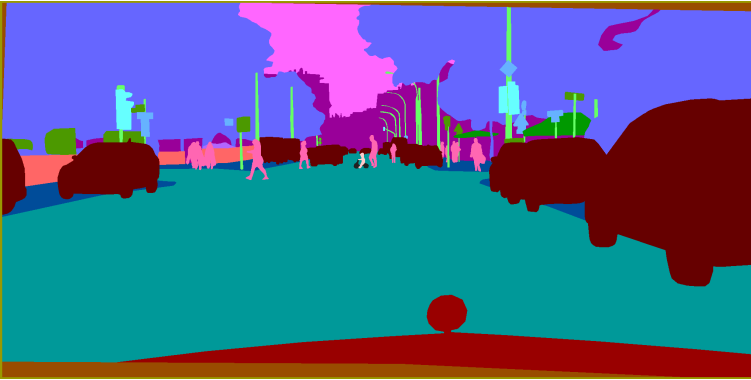
Seg-HiRes-Grad CAM

Examples with Cityscapes [6]

Input Image



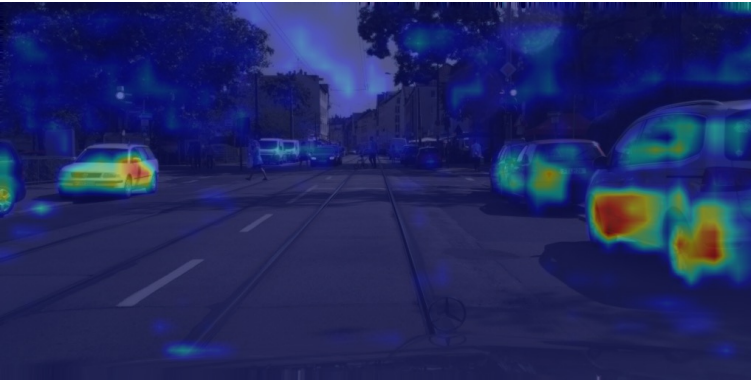
Ground Truth



Prediction with U-Net [8]



Segmentation to be explained



Seg-Grad CAM [2]



Seg-HiRes-Grad CAM

Discussion & Conclusion

Discussion & Conclusion

- Examples which are not explainable at all
- Dimensions of feature map must be large enough

Discussion & Conclusion

- Examples which are not explainable at all
- Dimensions of feature map must be large enough
- Future Work could include ...
 - ... different CNNs
 - ... 3D segmentation tasks
 - ... transfer to further classification-based CAM methods



Discussion & Conclusion

- Examples which are not explainable at all
- Dimensions of feature map must be large enough
- Future Work could include ...
 - ... different CNNs
 - ... 3D segmentation tasks
 - ... transfer to further classification-based CAM methods
- Seg-HiRes-Grad CAM can lift the SOTA of classification-based CAM methods to segmentation tasks

Thank you for your attention!

Feel free to ask questions & to stop by at the poster sessions!

 Tillmann_Rhe
 tillmann.rheude@bih-charite.de

Leveraging CAM Algorithms for Explaining Medical Semantic Segmentation

Tillmann Rheude¹, Andreas Wirtz², Arjan Kuijper^{1,2}, and Stefan Wesarg²

¹TU Darmstadt, Darmstadt, Germany
²Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany

Figure 1: Fraunhofer illustration of our proposed method Seg-HiRes-Grad CAM compared to the safety segmentation-based CAM baseline Seg-Grad CAM [3] for a medical image and the predicted post-act. The distinction lies in the computation of weights, which remain averaged in Seg-HiRes-Grad CAM. This elementary modification results in alterations to the final CAM values leading to more explainable results especially for medical images.

Abstract
One way of interpreting a Convolutional Neural Network (CNN) is by using class activation maps (CAMs) that represent heatmaps indicating the importance of image areas for the prediction of the CNN. For classification tasks, a variety of CAM algorithms exist. But for segmentation tasks, only one CAM algorithm exists. We propose a transfer between existing classification- and (the) segmentation-based method(s) for more detailed, explainable, and consistent results which show salient pixels in semantic segmentation tasks. The resulting Seg-HiRes-Grad CAM is an extension of the segmentation-based Seg-Grad CAM with the transfer to the classification-based HiRes CAM. Our method improves the existing segmentation-based method by adjusting it to the classification-based HiRes CAM. Especially for medical image segmentation, this transfer solves existing explainability disadvantages.

Introduction
Many well-known CAM algorithms (Table 1) were proposed in the past – dominated by classification-based work. The safety transfer to segmentation was proposed by Seg-Grad CAM [3], which is based on Grad CAM [4], so that certain inaccuracies, e.g., the visualization of irrelevant regions, occur. Therefore, we propose a combination of the segmentation-based Seg-Grad CAM and classification-based HiRes CAM to solve these inaccuracies: Seg-HiRes-Grad CAM.

Classification Task	Seg-Grad CAM [3]	Seg-HiRes-Grad CAM
CAM [2]	n/a	n/a
Grad CAM [4]	n/a	n/a
HiRes CAM [5]	n/a	n/a

Table 1: Comparison of existing classification- and segmentation-based CAM algorithms delineating the contribution of our work.

Method
The classification-based CAM is described mathematically as follows:

$$L_{CAM} = \sum_{i=1}^N w_i^k A^k$$

where w_i^k are weights per class and A^k are activation maps.

To improve visual results, Grad CAM uses ReLU and the respective gradients instead of only the weights of the last feed-forward network:

$$L_{Grad-CAM} = ReLU(L_{CAM}) \cdot \frac{1}{N} \sum_{i=1}^N \frac{\partial y^i}{\partial A^k}$$

But the mean calculation for the weights introduces inaccuracies. These are solved by HiRes CAM so that the weights are calculated as follows instead:

$$w_i^k = \frac{\partial y^i}{\partial A^k}$$

where y^i is the predicted mask.


By modifying the numerator of eq. (2), Grad CAM is transferred to segmentation tasks [3]:




$$y^{i,NEW} = \sum_{j=1}^N x_j^i$$

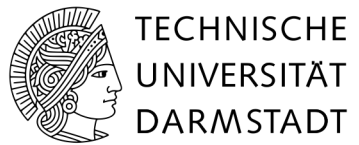
where x_j^i is the set of pixels in the predicted mask.

Results
Comparing our proposed Seg-HiRes-Grad CAM with the Seg-Grad CAM baseline, the resulting heatmaps differ regarding their accuracy and explainability. We can show this for computer vision benchmarks datasets, such as Chysoptes [7] (Fig. 2), and even more drastically for medical images (Fig. 3, Fig. 3). Future work could investigate transferring further classification-based CAMs to segmentation tasks.

References
[1] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[2] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[3] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[4] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[5] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[6] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.
[7] S. Zhou, L. Yang, J. Han, "Visual Question Answering on Medical Images for Learning the Link between Visual and Textual Information", *CVPR*, 2019.

Code available on 

 tillmann.rheude@bih-charite.de
 https://github.com/TillmannRheude/SegHiResGrad_CAM
 Tillmann_Rhe



References I

- [1] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, 2016, pp. 2921–2929.
- [2] K. Vinogradova, A. Dibrov, and G. Myers, “Towards Interpretable Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Student Abstract),” in Proceedings of the AAAI Conference on Artificial Intelligence, Apr. 2020, pp. 13943–13944.
- [3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, IEEE Computer Society, 2017, pp. 618–626.
- [4] R. L. Draelos and L. Carin, “Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks,” ArXiv, Nov. 2021.
- [5] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, “Deep Instance Segmentation of Teeth in Panoramic X-Ray Images,” in 31st SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2018, Paraná, Brazil, October 29 - Nov. 1, 2018, IEEE Computer Society, 2018, pp. 400–407.
- [6] M. Cordts et al., “The Cityscapes Dataset for Semantic Urban Scene Understanding,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, 2016, pp. 3213–3223.
- [7] N. Heller et al., “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 Challenge,” Medical Image Analysis, p. 101821, 2020.

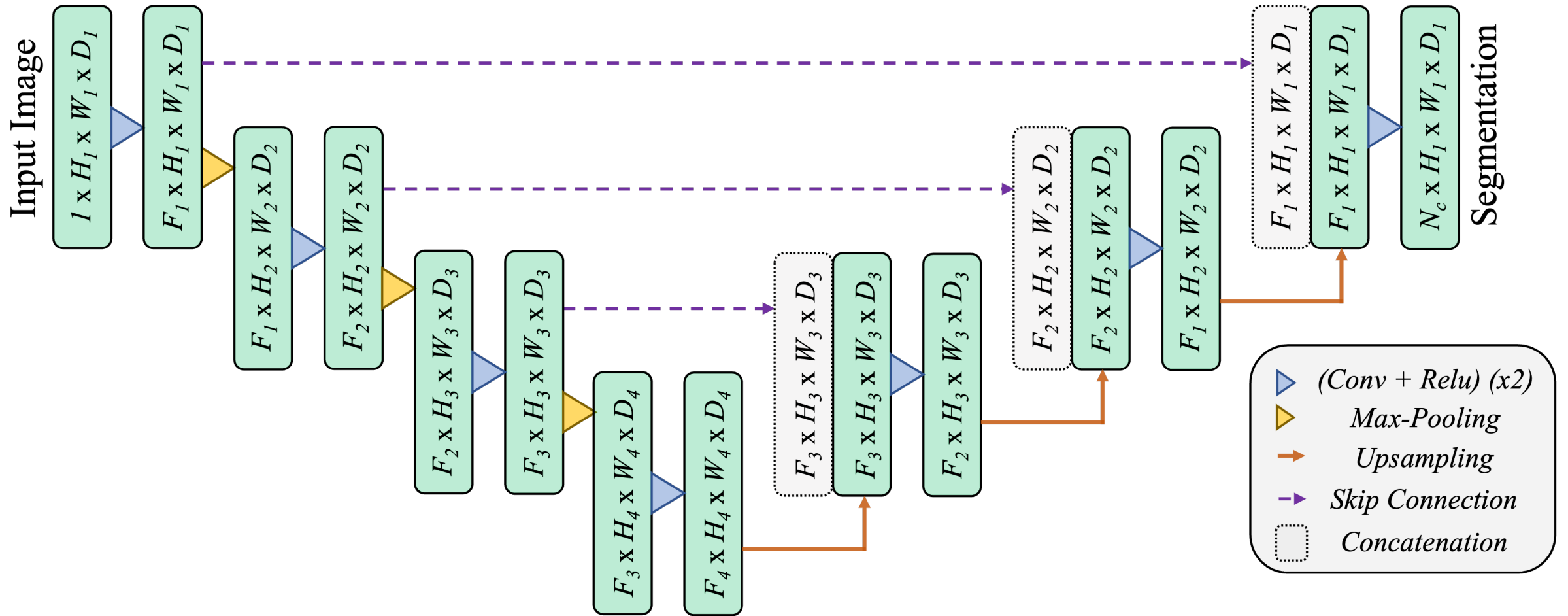
References II

- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, N. Navab, J. Hornegger, W. M. W. III, and A. F. Frangi, Eds., in *Lecture Notes in Computer Science*, vol. 9351. Springer, 2015, pp. 234–241.
- [9] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 839–847. doi: 10.1109/WACV.2018.00097.
- [10] R. Fu, Q. Hu, X. Dong, Y. Guo, Y. Gao, and B. Li, “Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs,” in *31st British Machine Vision Conference 2020, BMVC 2020, Virtual Event, UK, September 7-10, 2020*, BMVA Press, 2020.
- [11] M. B. Muhammad and M. Yeasin, “Eigen-CAM: Class Activation Map using Principal Components,” in *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, IEEE, 2020, pp. 1–7.
- [12] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, “LayerCAM: Exploring Hierarchical Class Activation Maps for Localization,” *IEEE Trans. Image Process.*, vol. 30, pp. 5875–5888, 2021.

Backup Slides



U-Net

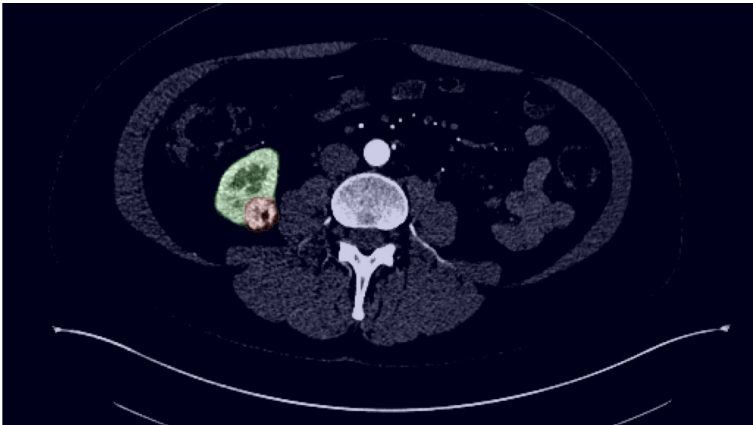


Examples with KITS23 [7]

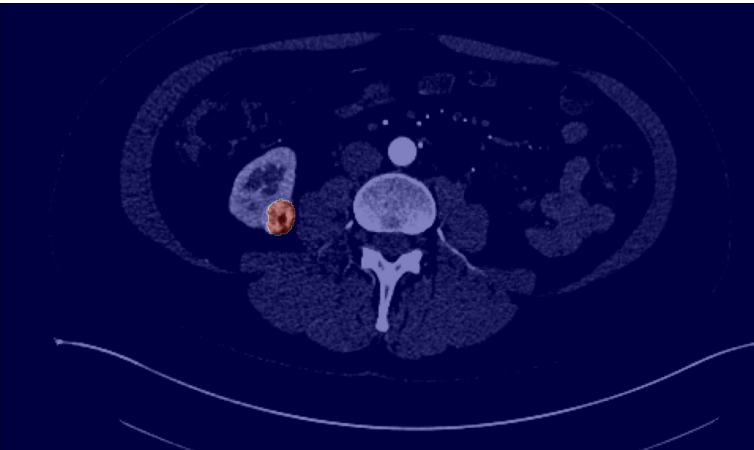
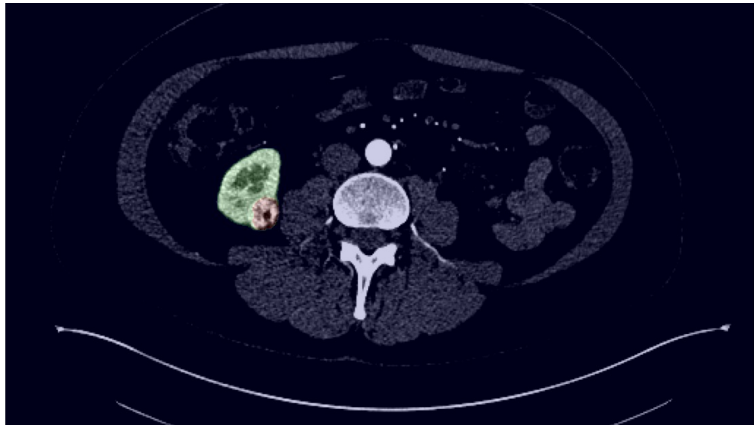
Input Image



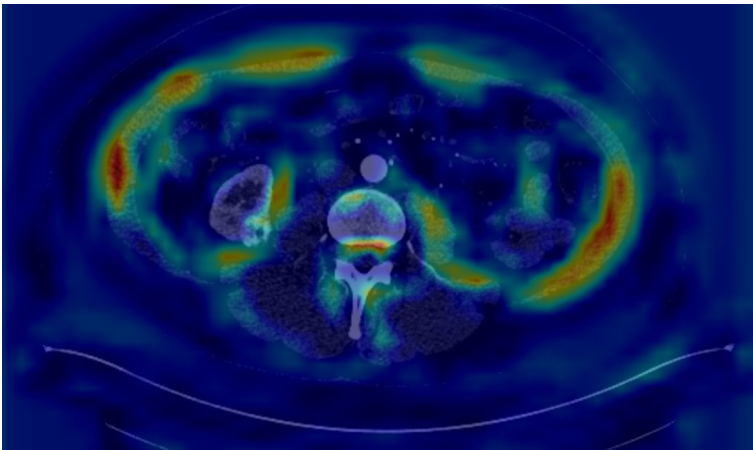
Ground Truth



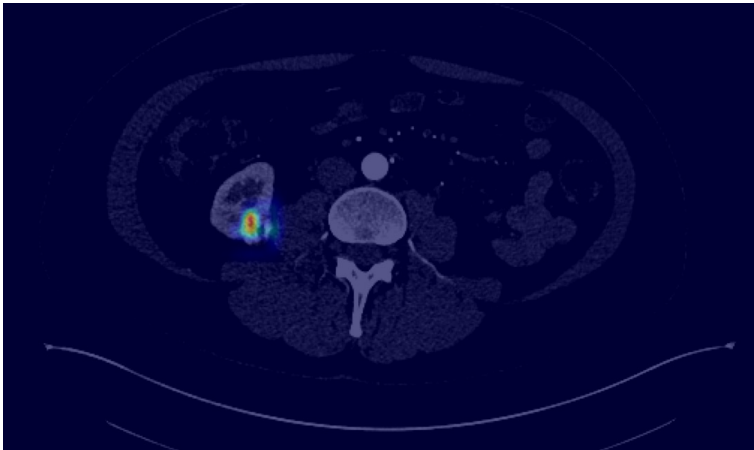
Prediction (U-Net [8])



Segmentation to be explained



Seg-Grad CAM [2]



Seg-HiRes-Grad CAM