Leveraging CAM Algorithms for Explaining Medical Semantic Segmentation

Tillmann Rheude[1 *], Andreas Wirtz[2], Stefan Wesarg[2], Arjan Kuijper[1,2]

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

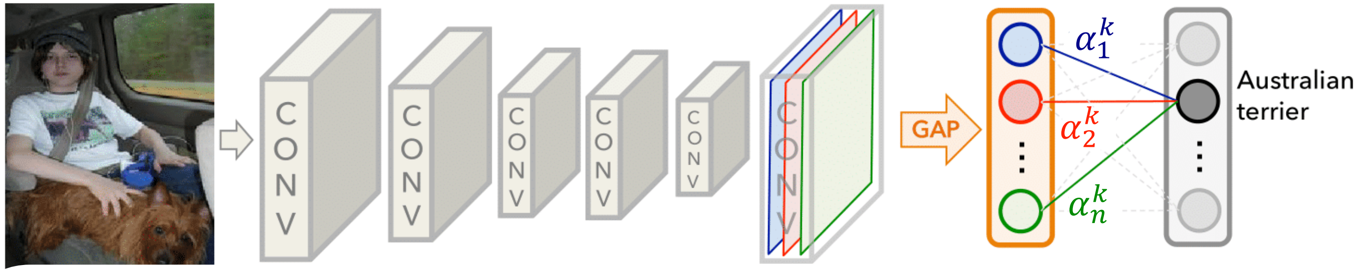
[*] meanwhile at Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

MICCAI 2023 Vancouver CANADA

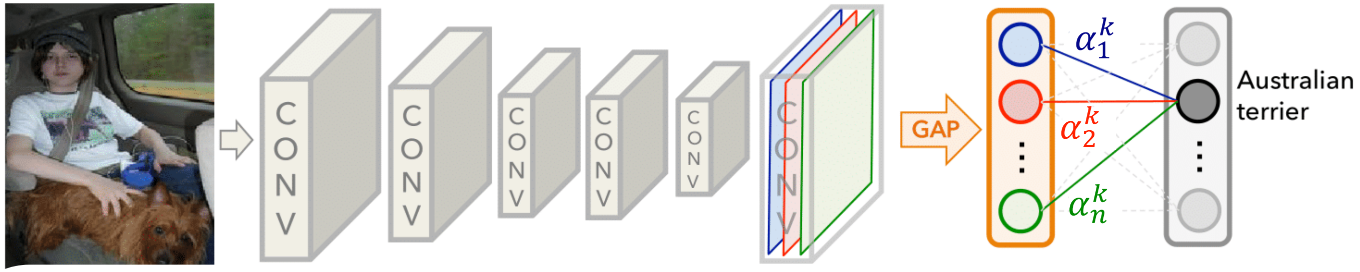
TECHNISCHE UNIVERSITÄT DARMSTADT

Fraunhofer IGD

CAM – Motivation with Classification Tasks

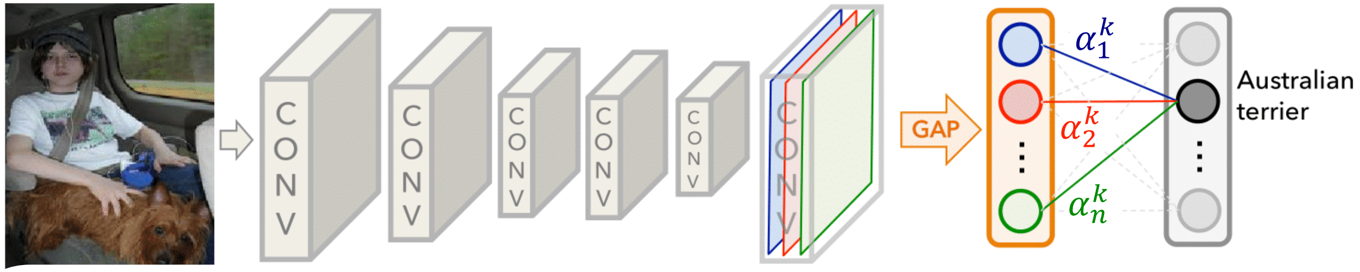


CAM – Motivation with Classification Tasks



Where did the CNN look?

CAM – Motivation with Classification Tasks

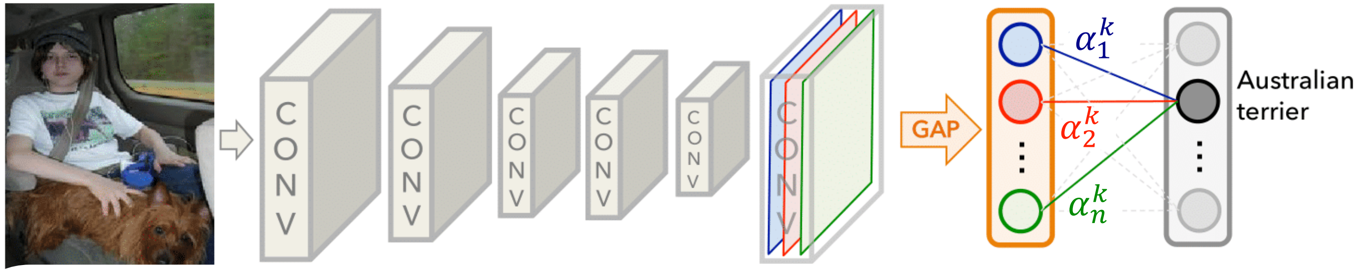


Where did the CNN look?

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

CAM [1]

CAM – Motivation with Classification Tasks



Where did the CNN look?

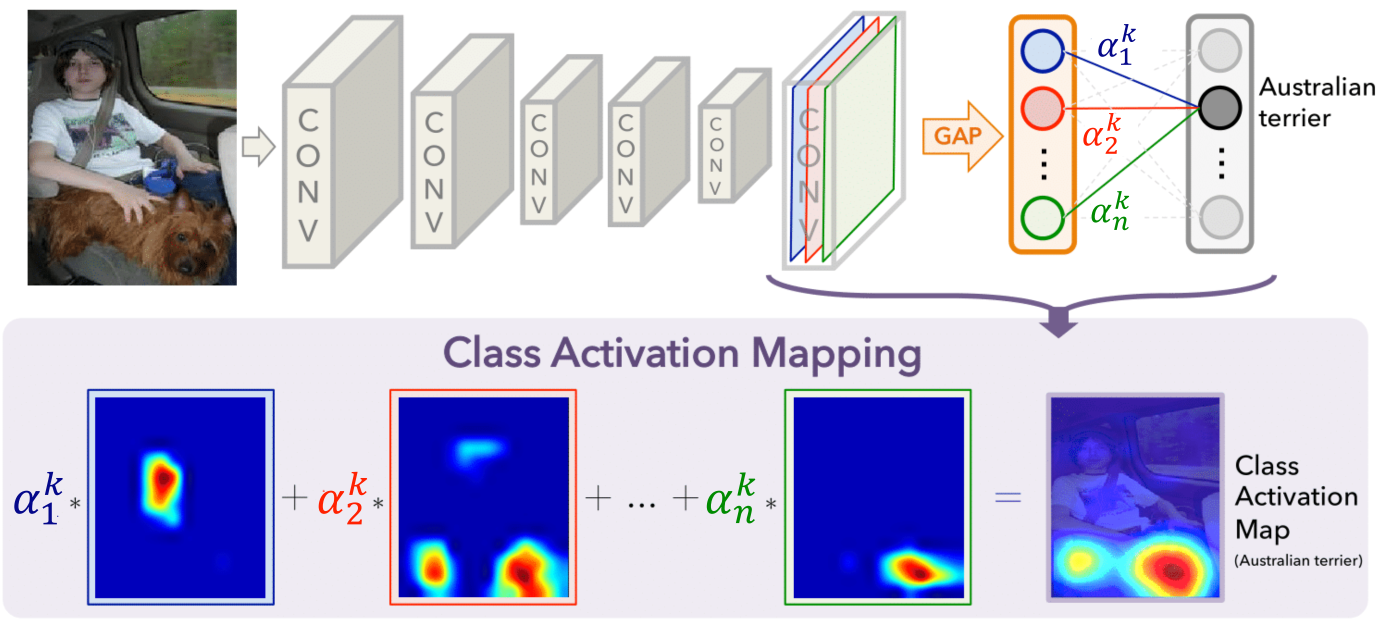
$$L^c = \underbrace{\sum_k}_{} \alpha_c^k A^k$$

Weights

Activation Maps

CAM [1]

CAM – Motivation with Classification Tasks



Where did the CNN look?

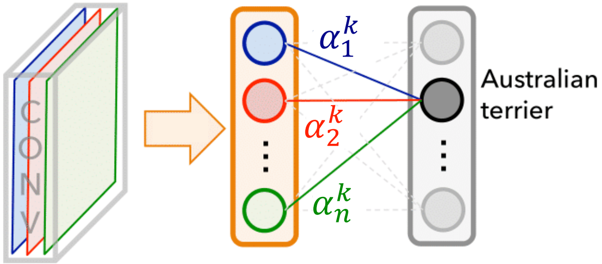
$$L^c = \sum_k \alpha_c^k A^k$$

Weights

Activation Maps

CAM [1]

HiRes / Grad CAM – Extensions to CAM

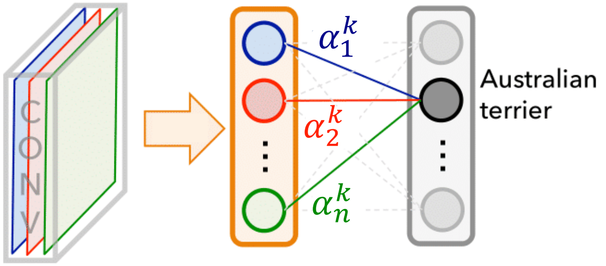


$$\alpha_c^k = weights$$

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

CAM [1]

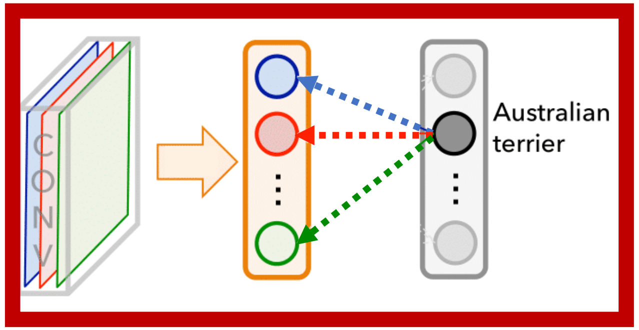
HiRes / Grad CAM – Extensions to CAM



$$\alpha_c^k = weights$$

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

CAM [1]

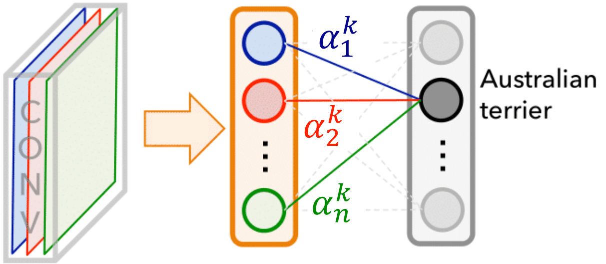


$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

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

Grad CAM [3]

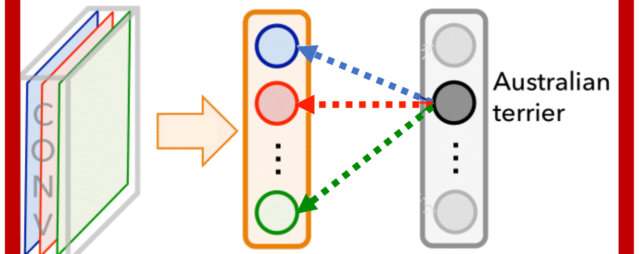
HiRes / Grad CAM – Extensions to CAM



$$\alpha_c^k = weights$$

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

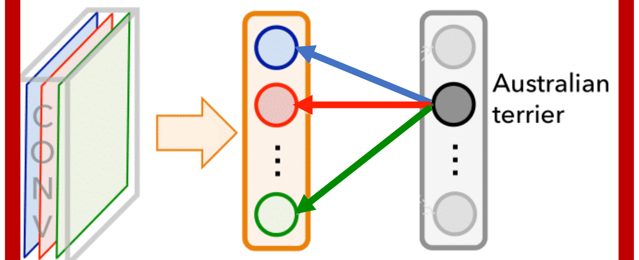
CAM [1]



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

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

Grad CAM [3]

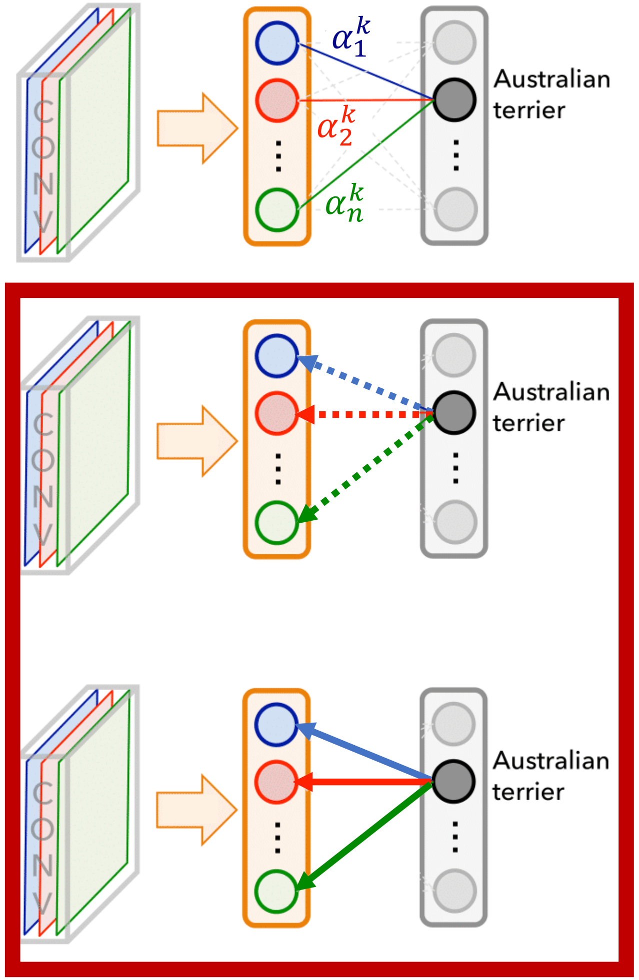


$$\alpha_c^k = \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

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

HiRes CAM [4]

HiRes / Grad CAM – Extensions to CAM



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

CAM [1]

$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k} \qquad L^c = ReLU\left(\sum_k \alpha_c^k A^k\right)$$

Grad CAM [3]

Fix inaccuracies

$$\alpha_c^k = \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k} \qquad L^c = ReLU\left(\sum_k \alpha_c^k A^k\right)$$

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]

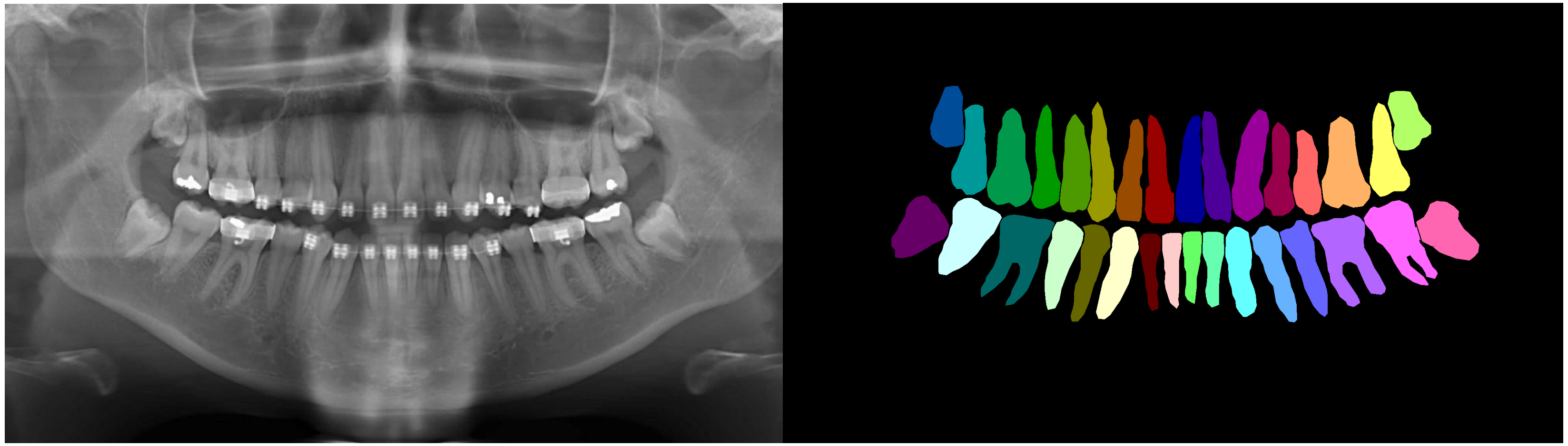
Overview – CAMs for Classification and Segmentation Tasks

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

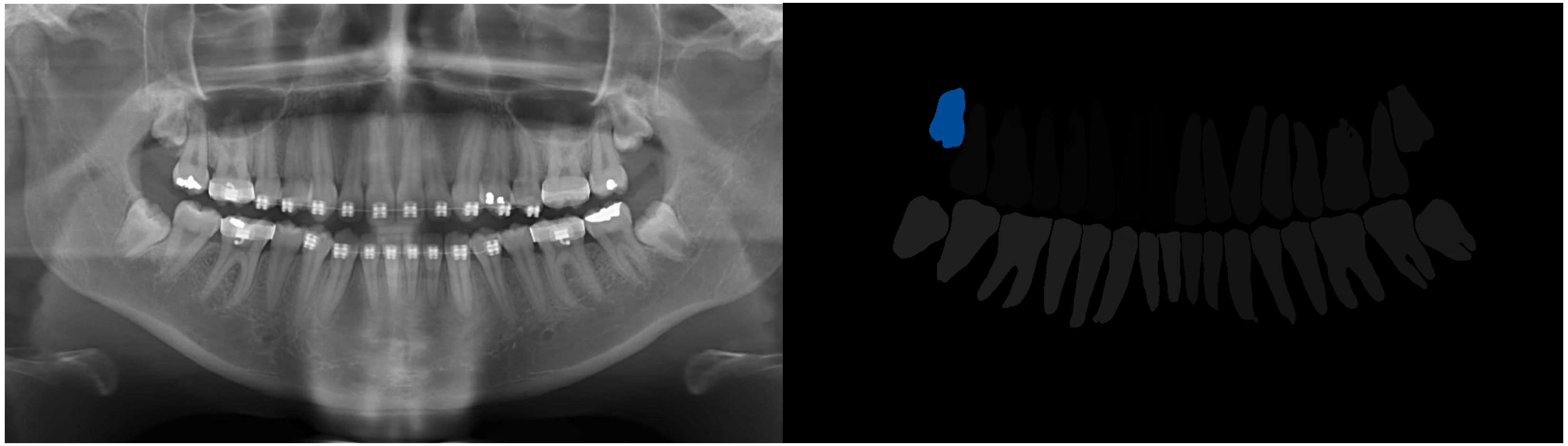
Overview – CAMs for Classification and Segmentation Tasks

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

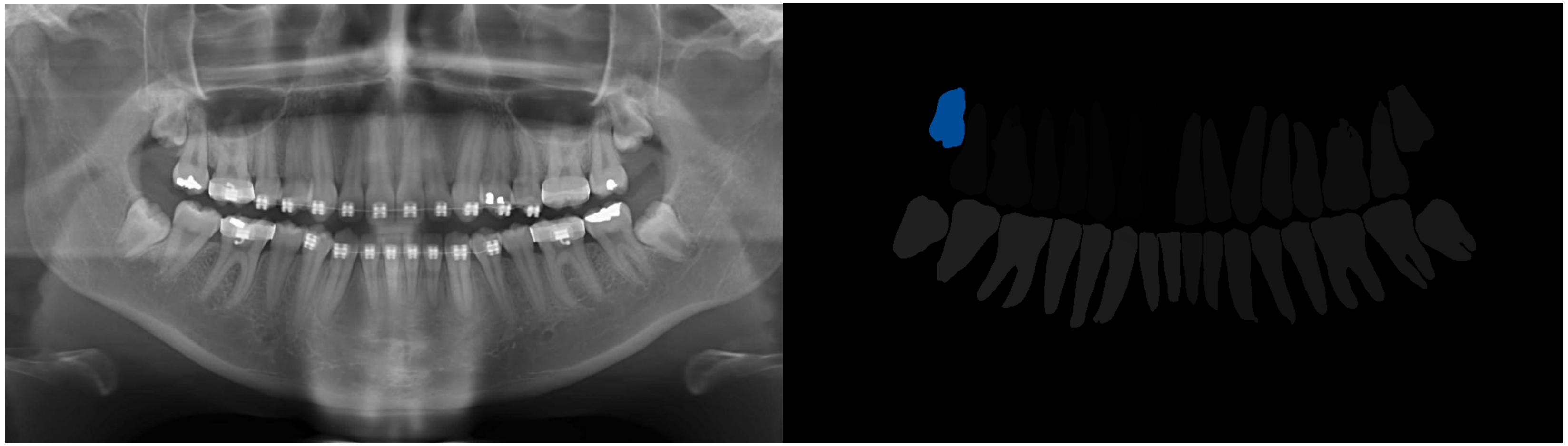
Seg-Grad CAM – Transfer to Segmentation Tasks



Seg-Grad CAM – Transfer to Segmentation Tasks

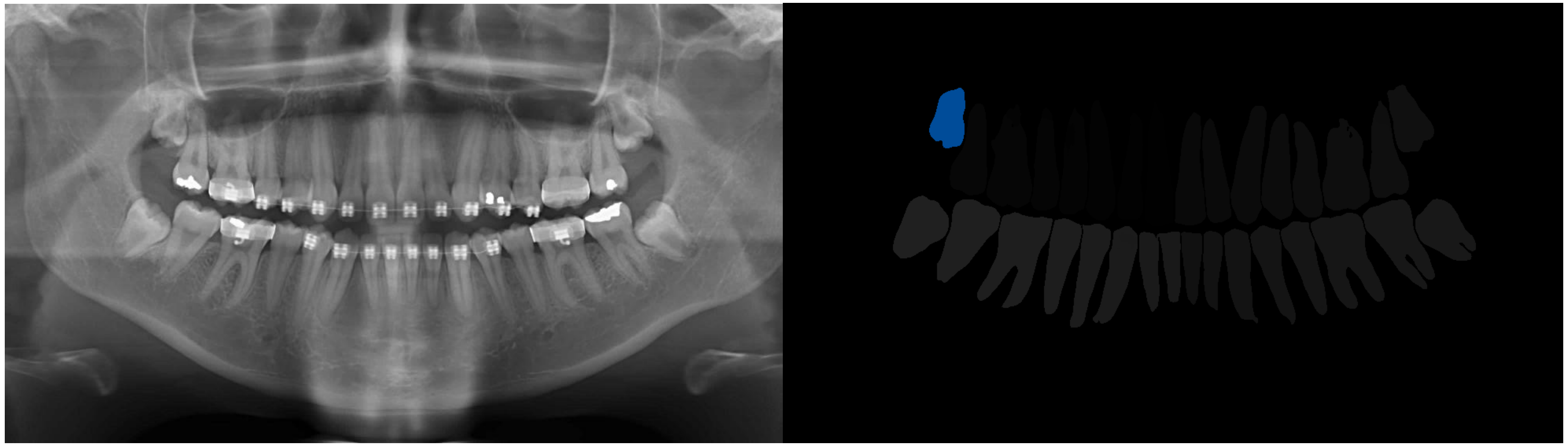


Seg-Grad CAM – Transfer to Segmentation Tasks



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

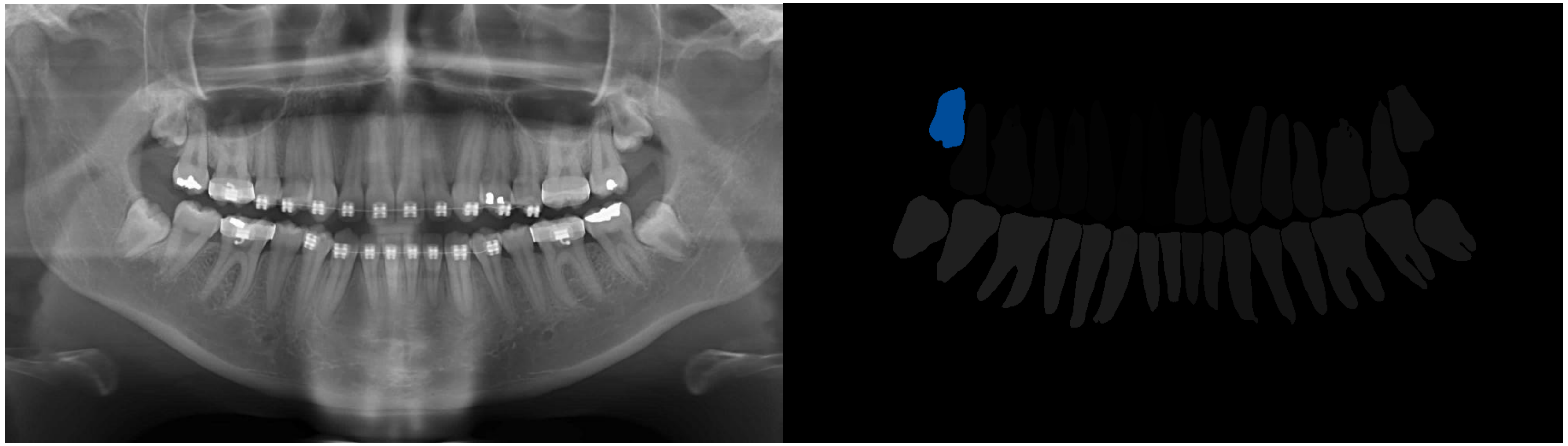
Seg-Grad CAM – Transfer to Segmentation Tasks



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

Output for a certain class

Seg-Grad CAM – Transfer to Segmentation Tasks

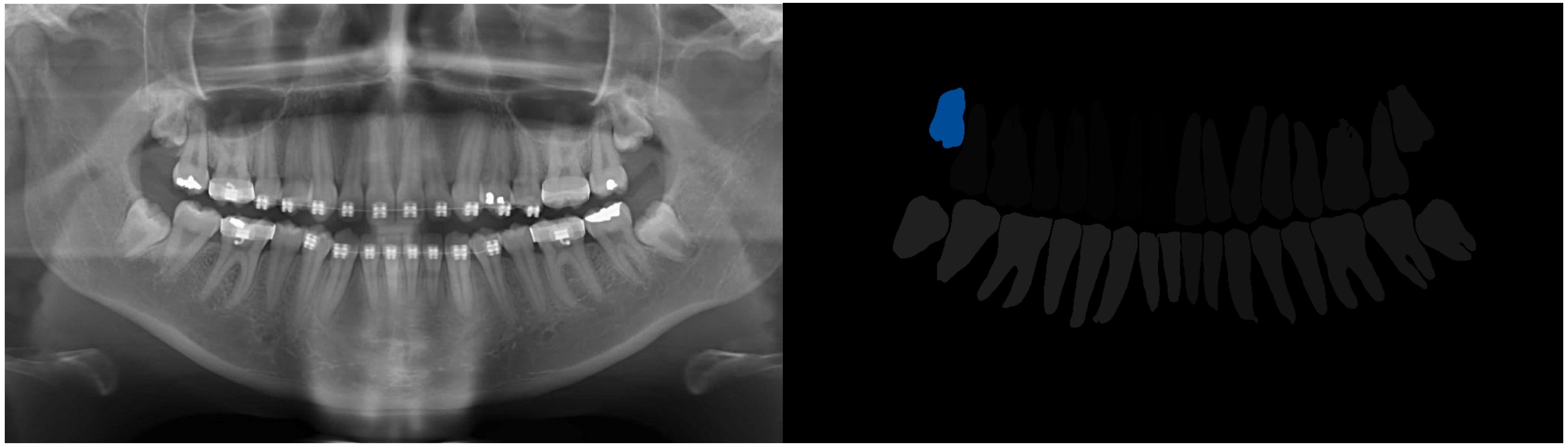


$$\alpha_c^k = \frac{1}{N}\sum_{u,v}\frac{\partial y^c}{\partial A_{uv}^k}$$

Output for a certain class

Feature Maps for each channel and each pixel

Seg-Grad CAM – Transfer to Segmentation Tasks



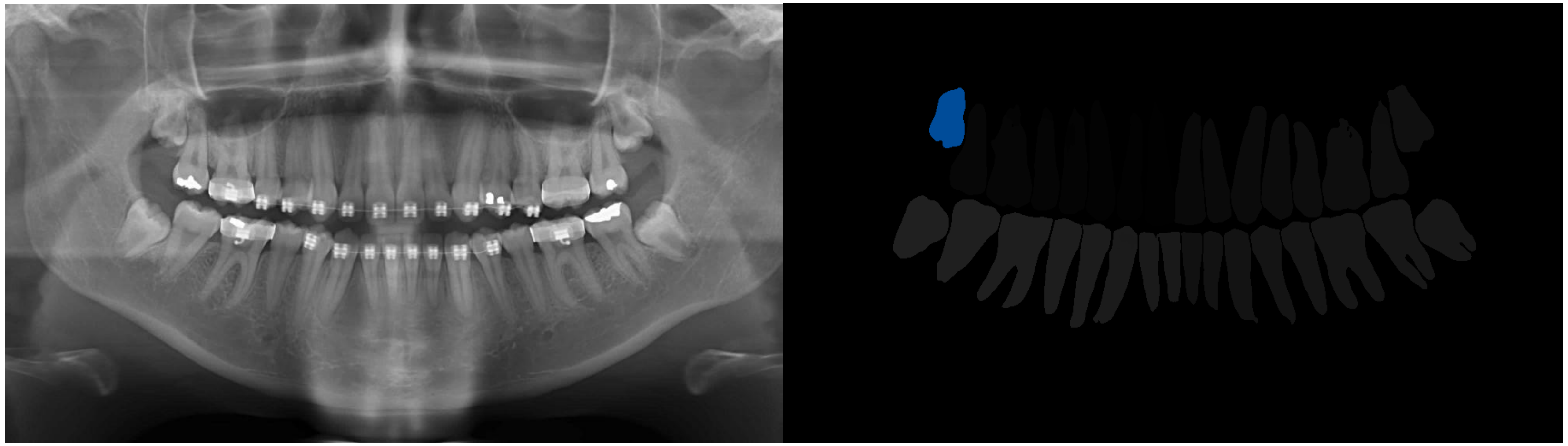
$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

Output for a certain class

Number of Pixels

Feature Maps for each channel and each pixel

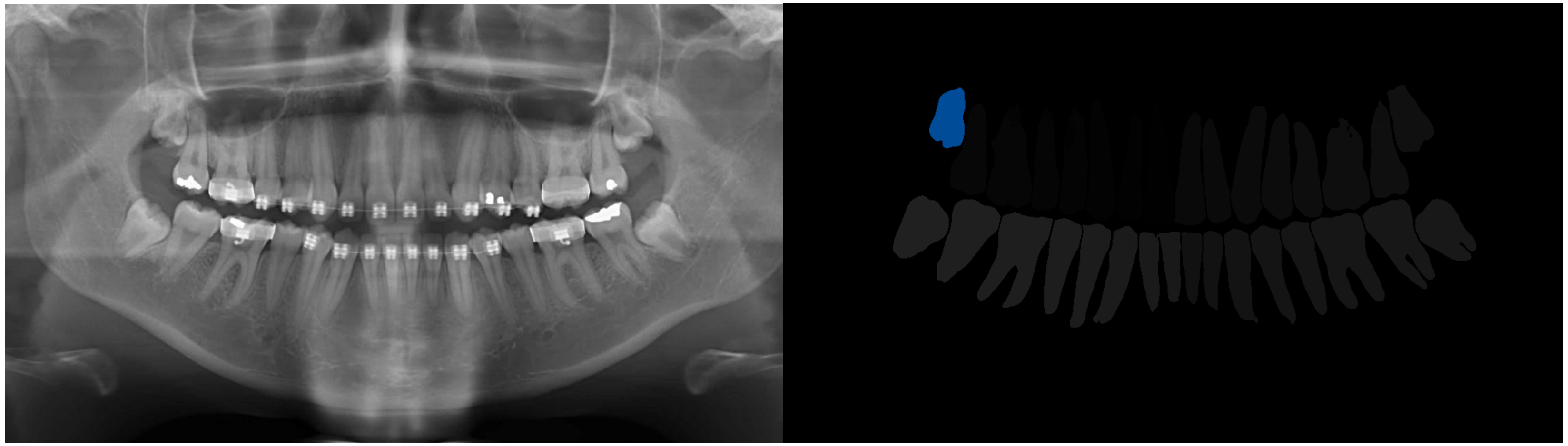
Seg-Grad CAM – Transfer to Segmentation Tasks



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

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

Seg-Grad CAM – Transfer to Segmentation Tasks

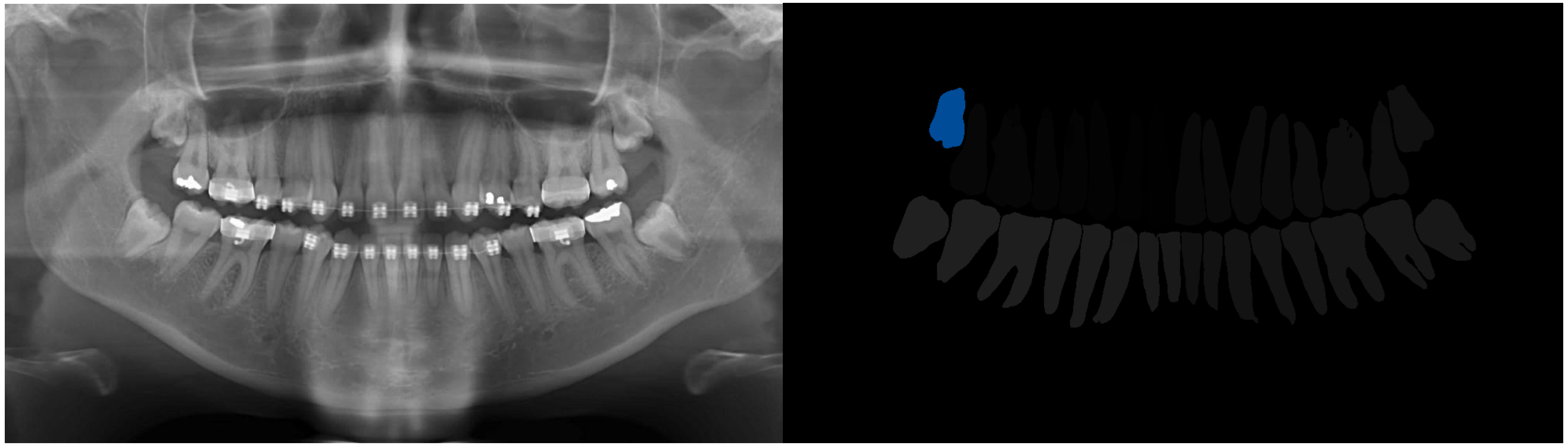


$$\alpha_c^k = \frac{1}{N} \sum_{u,v} \frac{\partial y^c}{\partial A_{uv}^k}$$

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

Set of Pixel Indices

Seg-Grad CAM – Transfer to Segmentation Tasks

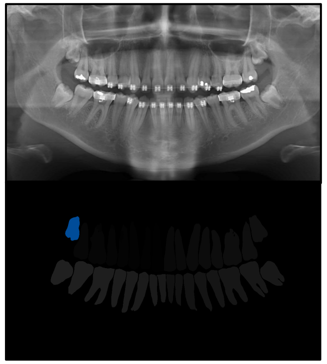


$$\alpha_c^k = \frac{1}{N}\sum_{u,v}\frac{\partial y^c}{\partial A_{uv}^k}$$

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

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

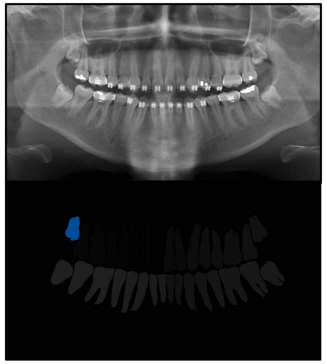
Seg-Grad CAM – Transfer to Segmentation Tasks



Input and Pixelset $\mathcal{M}$

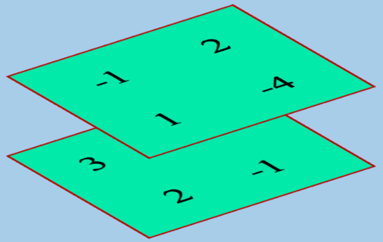
$$L^c = ReLU\left(\sum_k \alpha_c^k A^k\right)$$

Seg-Grad CAM – Transfer to Segmentation Tasks



Input and Pixelset $\mathcal{M}$

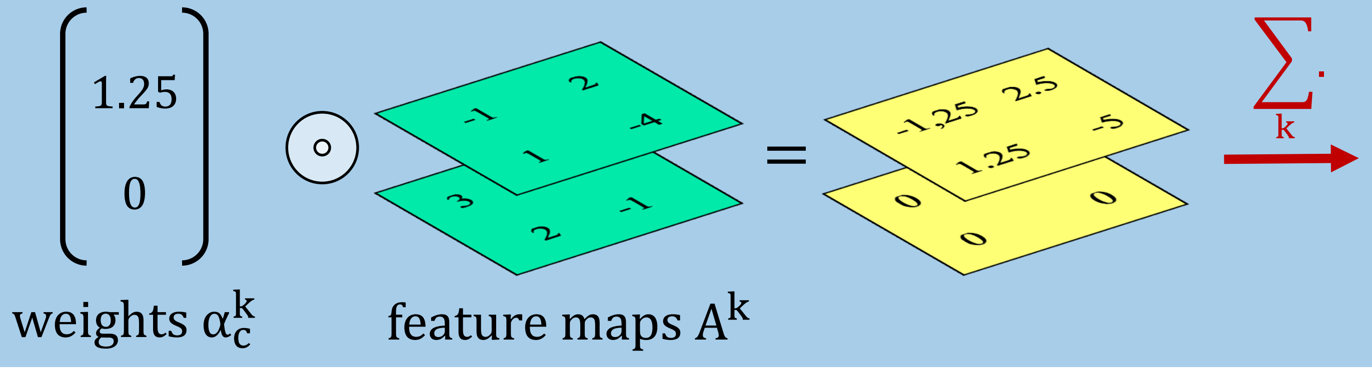
$$\begin{bmatrix} 1.25 \\ 0 \end{bmatrix} \odot$$

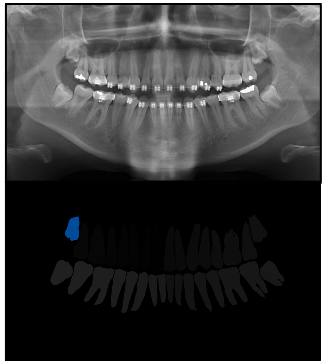


weights α_c^k feature maps A^k

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

Seg-Grad CAM – Transfer to Segmentation Tasks

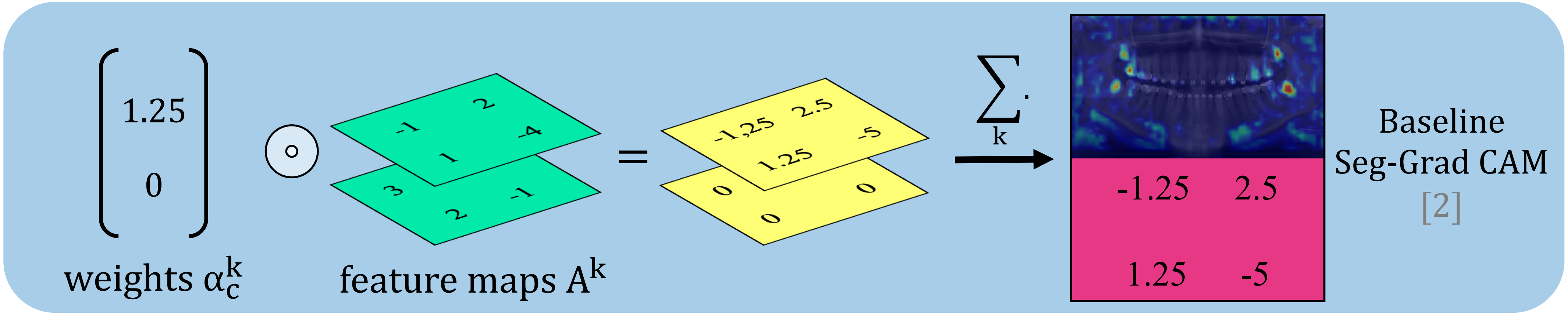


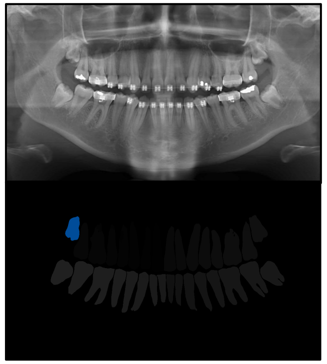


Input and Pixelset $\mathcal{M}$

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

Seg-Grad CAM – Transfer to Segmentation Tasks





Input and Pixelset $\mathcal{M}$

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

* *ReLU not visualized to ensure a better comparability later on*

Overview – CAMs for Classification and Segmentation Tasks

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

Overview – CAMs for Classification and Segmentation Tasks

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

Overview – CAMs for Classification and Segmentation Tasks

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

!

Fix inaccuracies

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

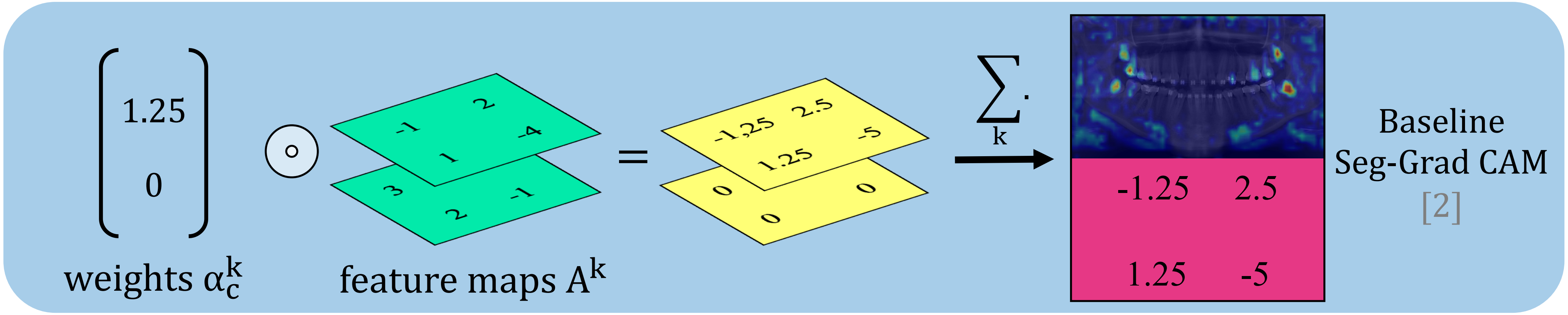
Classification	**Segmentation**
CAM [1]	*n/a*
Grad CAM [3]	Seg-Grad CAM [2]
Grad CAM++ [9]	*n/a*
XGrad CAM [10]	*n/a*
...	...
Eigen CAM [11]	*n/a*
Layer CAM [12]	*n/a*
HiRes CAM [4]	**Seg-HiRes-Grad CAM**

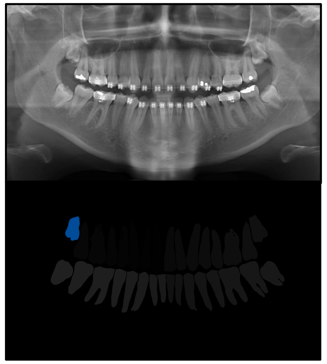
!

Fix inaccuracies

Fix inaccuracies

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





Input and
Pixelset $\mathcal{M}$

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

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

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

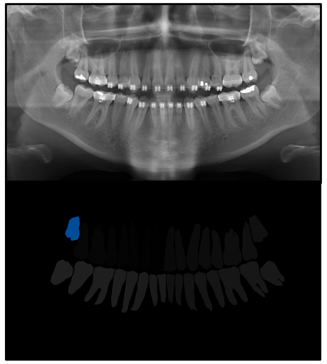
** ReLU not visualized to ensure a better comparability later on*

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

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

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

$$\alpha_c^k = \frac{\sum_{i,j\in\mathcal{M}} y_{i,j}^c}{\partial A^k}$$



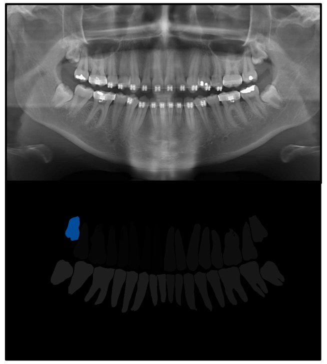
Input and Pixelset $\mathcal{M}$

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

$$\alpha_c^k = \cancel{\frac{1}{N}} \cancel{\sum_{u,v}} \frac{\partial y^c}{\partial A^k_{\cancel{u,v}}}$$

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

$$\alpha_c^k = \frac{\sum_{i,j \in \mathcal{M}} y_{i,j}^c}{\partial A^k}$$



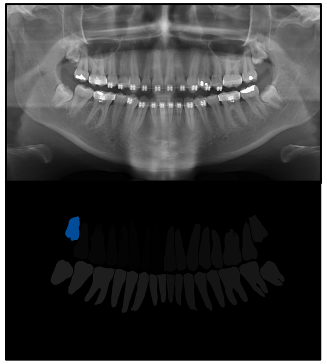
Input and Pixelset $\mathcal{M}$

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

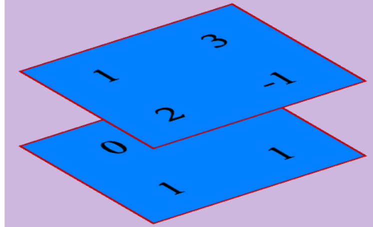
$$\alpha_c^k = \cancel{\frac{1}{N}}\cancel{\sum_{u,v}} \frac{\partial y^c}{\partial A^k_{\cancel{u,v}}}$$

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

$$\alpha_c^k = \frac{\sum_{i,j \in \mathcal{M}} y^c_{i,j}}{\partial A^k}$$



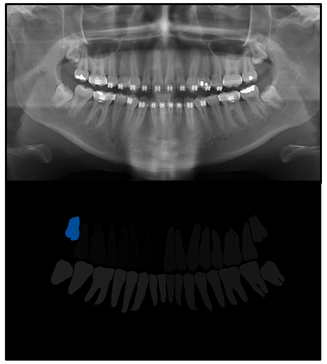
Input and Pixelset $\mathcal{M}$



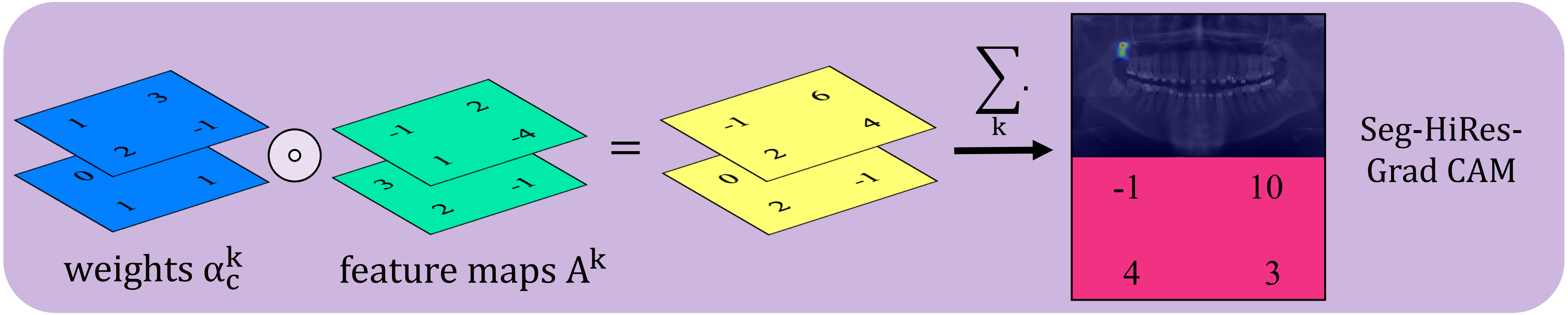
weights α_c^k

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

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



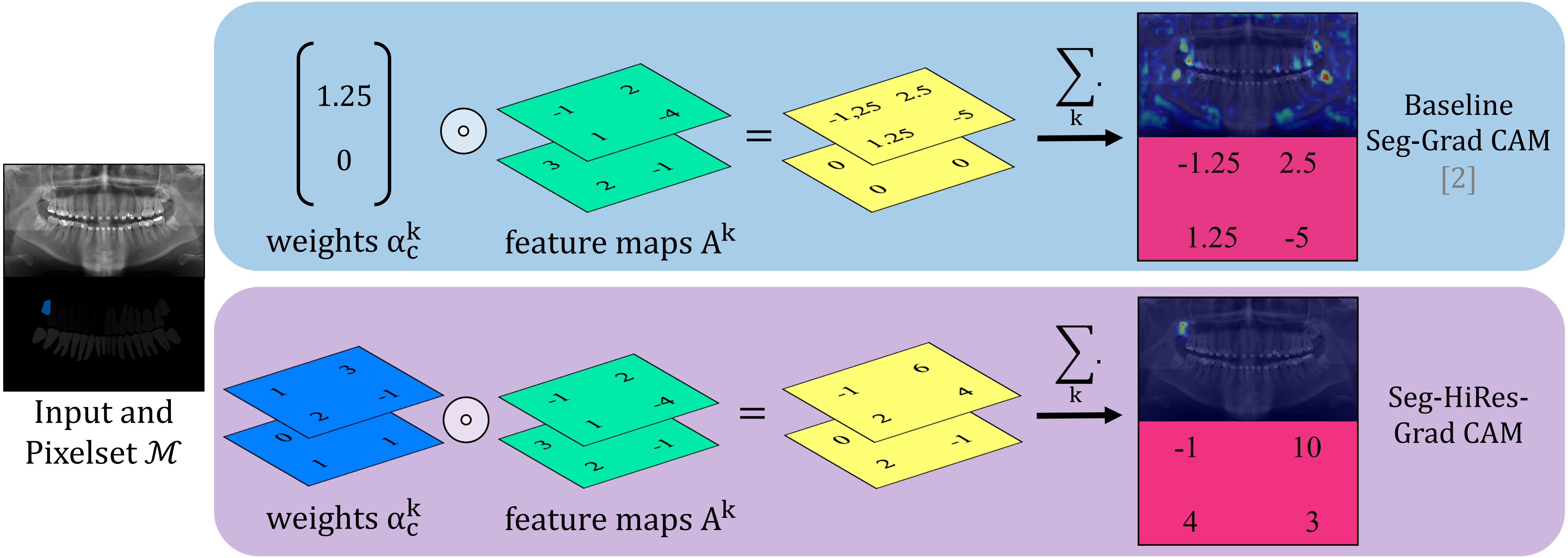
Input and Pixelset $\mathcal{M}$



Seg-HiRes-Grad CAM

** ReLU not visualized to ensure a better comparability later on*

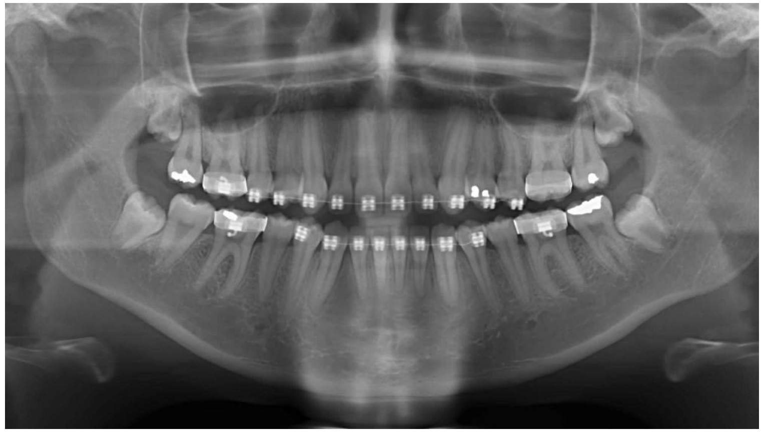
Seg-HiRes-Grad CAM – Extension to Seg-Grad CAM



* ReLU not visualized to ensure a better comparability

Examples with OPG [5]

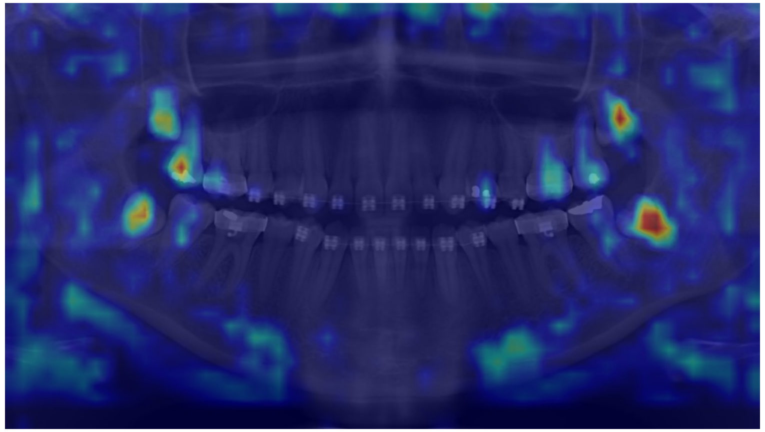
Input Image



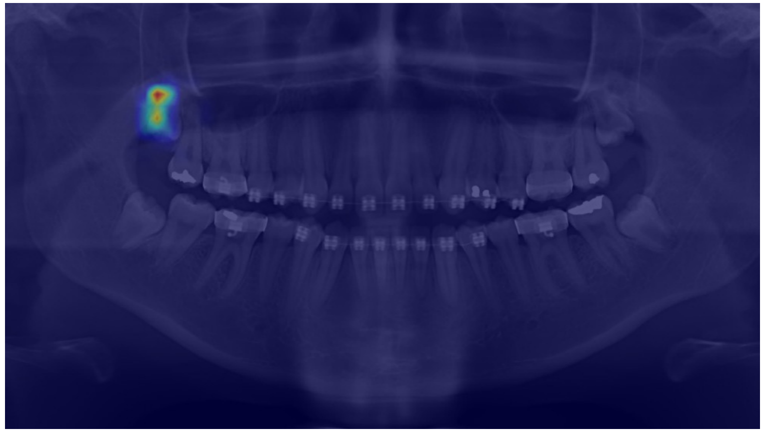
Ground Truth

Prediction with U-Net [8]

Segmentation to be explained



Seg-Grad CAM [2]

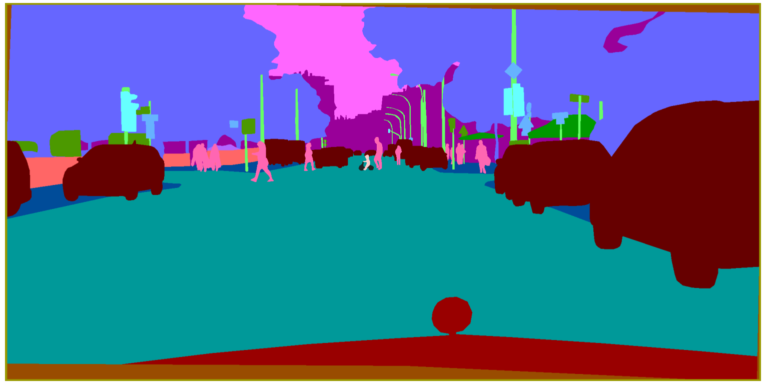


Seg-HiRes-Grad CAM

Examples with Cityscapes [6]

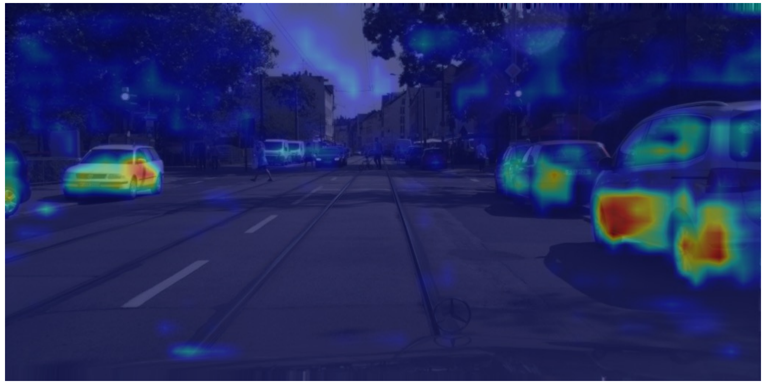
Input Image

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

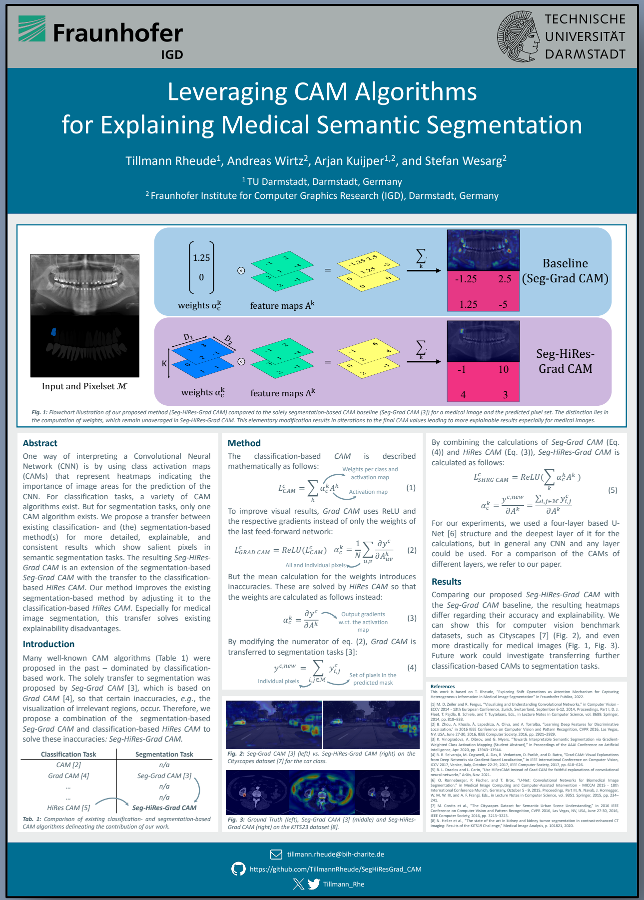
- Examples which are not explainable at all
- Dimensions of feature map must be large enough
- Future Work could include …
  - … different CNNs
  - … 3D segmentation tasks
  - … transfer to further classification-based CAM methods

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!



Code available on

Tillmann_Rhe

tillmann.rheude@bih-charite.de

MICCAI 2023 Vancouver CANADA

TECHNISCHE UNIVERSITÄT DARMSTADT

Fraunhofer IGD

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. Chattopadhay, 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.

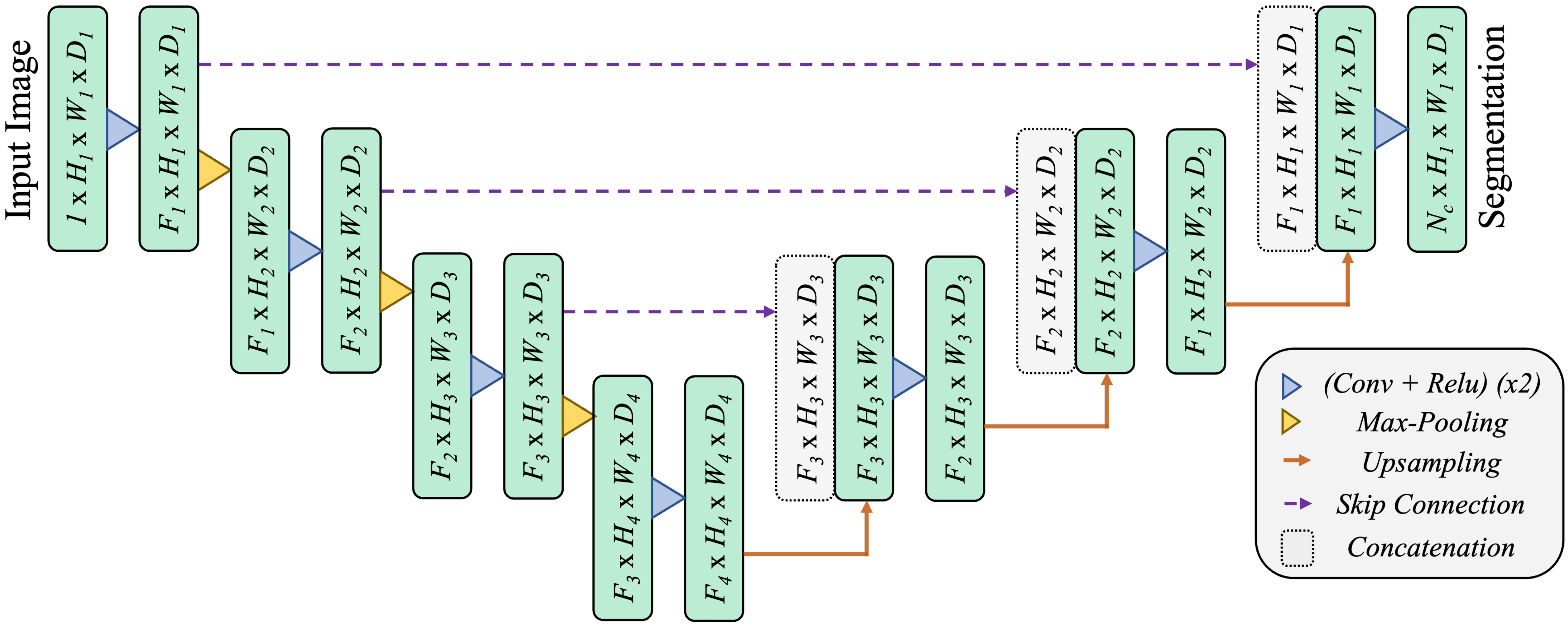
[11] M. B. Muhammad and M. Yeasin, “Eigen-CAM: Class Activation Map using Principal Components,” in 2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020, IEEE, 2020, pp. 1–7.

[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

TECHNISCHE
UNIVERSITÄT
DARMSTADT

U-Net



Input Image
$1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_4 \times W_4 \times D_4$
$F_4 \times H_4 \times W_4 \times D_4$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_2 \times H_3 \times W_3 \times D_3$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_1 \times H_2 \times W_2 \times D_2$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$N_c \times H_1 \times W_1 \times D_1$
Segmentation
(Conv + Relu) (x2)
Max-Pooling
Upsampling
Skip Connection
Concatenation

Examples with KITS23 [7]