IEEE - ISBI 2021
International Symposium
on Biomedical Imaging

SMOCAM:
SMOoth Conditional Attention Mask for 3D-Regression Models

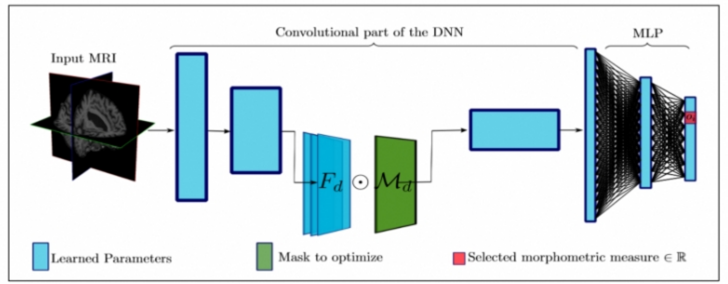
S. KONATE[‡,+], L. LEBRAT[*,+], R. SANTA CRUZ[*,+], P. BOURGEAT[*], V. DORE[*,⋆], J. FRIPP[*], A. BRADLEY[+], C. FOOKES[+], O. SALVADO[‡,+]

[*]CSIRO Health and Biosecurity, [‡]CSIRO Data61, [+]Image and Video Laboratory QUT, [⋆]Department of Nuclear Medicine and Centre for PET, Australia.

MOTIVATION AND OBJECTIVES

- We introduce **SMOCAM** (**SM**Ooth **C**onditional **A**ttention **M**ask), **an optimization method** that reveals the specific regions of the input image considered by the prediction of a trained neural network for **brain morphometric measurement**.
- SMOCAM performs **saliency analysis** for complex **regression** tasks for **3D medical image** with deep convolutional neural networks.
- SMOCAM optimizes an attention mask **at a given layer** of a convolutional neural network (CNN) in **40secondes**.
- This attention mask is point-wise multiplied for all the features at a certain depth of the CNN, we wish to find a mask that yields a minimal prediction error, and which is spatially smooth given a limited L^2 budget.
- SMOCAM can help to identify **neural network's limitations** when cases are underrepresented as cases with large volume asymmetry.

METHOD



- F_d a feature map at a depth d of dimension $L_d \times H_d \times W_d \times D_d$.
- **Element wise multiply** F_d with a mask M_d of dimension $1 \times H_d \times W_d \times D_d$.
- Pass the masked feature map $F_d \odot Md$ **through the remaining layers** of the network.
- The modified predictions $\boldsymbol{DNN_{M_d}}$ is therefore obtained for a **selected brain morphometric measurement** output o_i.
- **Upsample the mask** $\boldsymbol{M_d}$ to the size at the input image.
- Given a fixed L^2 budget $\|M_d\|_2 = 1$, find a mask M_d such that the prediction after masking the feature map DNN_{Md} is as close as possible to its original value DNN:

$$\boldsymbol{\arg\min_{M_d} d\left(DNN_{M_d}, DNN\right)}, \text{ with } \boldsymbol{d} \text{ the normalized distance}$$

- Problem solves using its **Tikhonov form** by adding a regularization term $\mathcal{R}_s$ that promotes spatial smoothness of the mask:

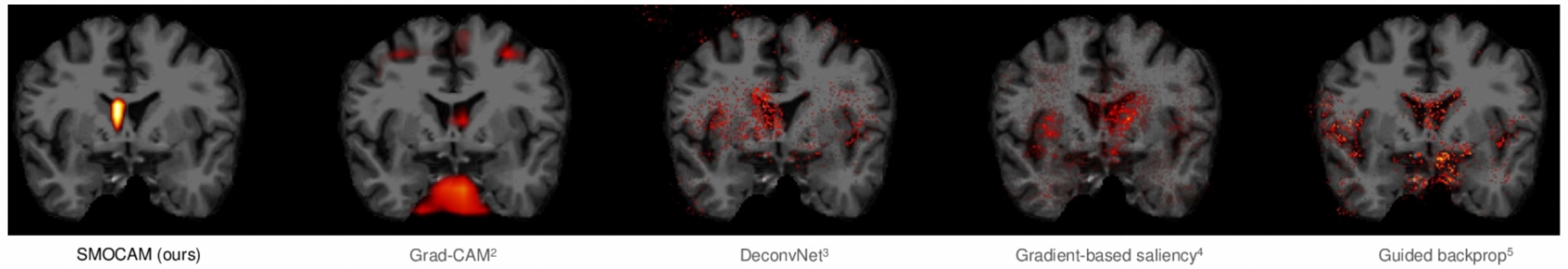
$$\arg\min_{M_d} \boldsymbol{d}\left(DNN_{M_d}, DNN\right) + \lambda \mathcal{R}_{\|\cdot\|_2}(Md) + \gamma \mathcal{R}_s(M_d), \text{ where}$$

$d(a+b) = \sigma^{-1}\|a-b\|_1$, $\mathcal{R}_{\|\cdot\|_2}(M_d) = |1 - \|M_d\|_2|$,
$\mathcal{R}_s(M_d) = \frac{1}{3}(\|\nabla_x M_d\|_1 + \|\nabla_y M_d\|_1 + \|\nabla_z M_d\|_1)$,
with σ the standard deviation of the output measurement. γ and λ two hyper-parameters,

RESULTS

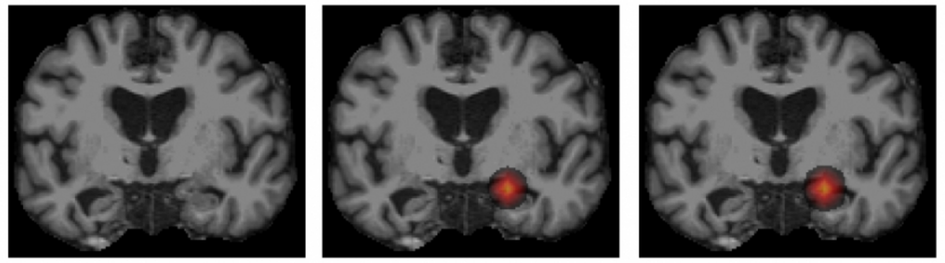
- Dataset:T1 weighted 3D-MRI scans
- Model: regression model based on the VGG framework train[1] with Mean square error and produce 165 morphometric measurements.

Different methods of saliency maps for the volume prediction of the right lateral ventricle



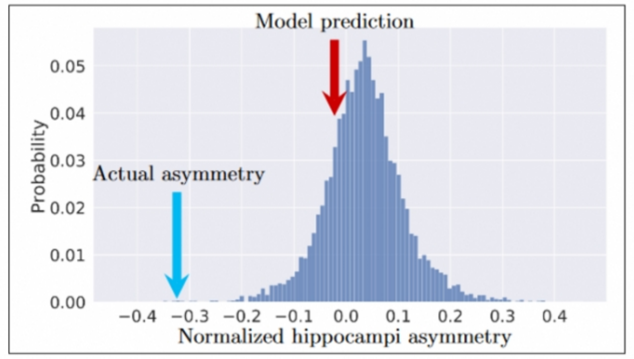
- The neural network estimated a hemispherical asymmetry of $0.10cm^3$ whereas the ground truth variation was 10 times bigger. We hypothesized that only one hippocampus was used for both left and right predictions because on average both sides are well correlated and that large asymmetries are underrepresented in our training dataset.

Attention map for the prediction of both hippocampi for a patient with large asymmetry



Patient with large asymmetry of the hippocampi $1.12cm^3$(left), attention mask for the right hippocampus volume (middle), attention mask for left hippocampus volume (right).

Distribution of the hippocampi asymmetries of the training dataset



Normalized hippocampi asymmetries (right hemisphere minus left hemisphere volume), position of the individual within the distribution of the training dataset (blue) and prediction made by the deep-learning model (red).

CONCLUSIONS

- SMOCAM performs a quantitative analysis of brain morphometric measurements by optimizing the mask for brain parcellation volumes and intersecting the generated attention map with all brain regions to obtain their respective attention.
- SMOCAM produces localised attention maps relatively fast for 3D regression models.
- For predicting a volume, SMOCAM can highlight the region of interest as well as its neighbour regions.
- Our approach can be used to detect some limitations of end-to-end deep-learning models like its poor performance on anatomical areas where the data is under-represented.

REFERENCES

[1] M. Rebsamen, Y. Suter, R. Wiest, M. Reyes, and C. Rummel, "Brain morphometry estimation: From hours to seconds using deep learning," Frontiers in neurology, vol. 11, p. 244, 2020.
[2] 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 Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
[3] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision. Springer, 2014, pp. 818–833.
[4] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps,"arXiv preprintarXiv:1312.6034, 2013.
[5] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences,"arXiv preprint arXiv:1704.02685, 2017.

ACKNOWLEDGEMENTS

Thanks to Maxwell plus http://maxwellplus.com/.

CONTACT INFORMATION

Salamata.konate@hdr.qut.edu.au