




Understanding Graph Convolutional Networks to detect Brain Lesions from Stroke

Ariel Iporre-Rivas¹ , Gerik Scheuermann¹  and Christina Gillmann¹ 

¹Leipzig University, Germany

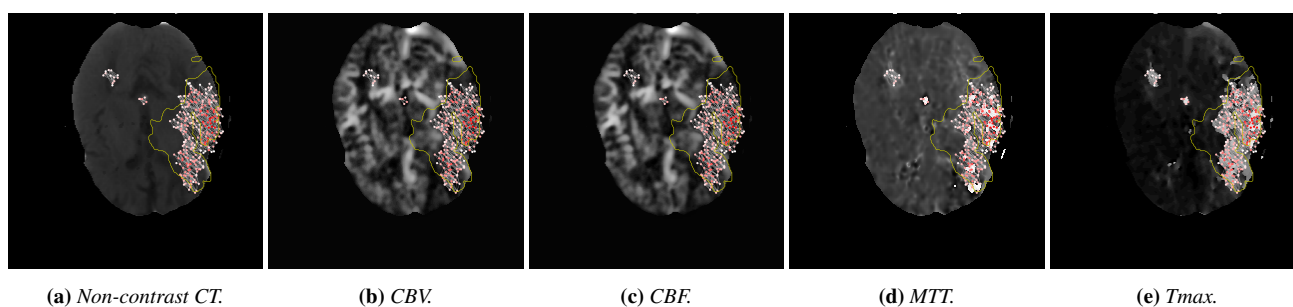


Figure 1: Results of the presented methods. The presented visualization was applied to a) a non-contrast CT and several parameter maps resulting from perfusion CT from b) to e). b) shows the applied visualization applied to cerebral blood volume, c) to cerebral blood flow, d) to mean transit time and e) to time to peak.

Abstract

Brain lesions derived from stroke episodes can result in disabilities for a patient. Therefore, the segmentation of brain lesions is an important task in neurology. Recently this task has been mainly tackled by machine learning approaches that demonstrated to be very successful. One of these approaches is Graph Convolutional Networks (GCN), where the input image is interpreted as a graph structure. As usual for neural networks, the interpretability is hard due to their black-box nature. We provide an interactive visualization of the activation inherent in the GCN, which is map from the original dataset. We visualize the activation values of the underlying graph network on top of the input image. We show the usability of our approach by applying it to a GCN that was trained on a real-world dataset.

1. Introduction

Stroke is the second leading cause of disability worldwide [SMS*20]. It cause by a reduction or blockage of blood supply to brain tissue, which leads to cell death and thus cognitive impairment. This scenario is referenced as an ischaemic stroke lesion. In order to provide suitable therapy for patients, the detection of the location of these lesions is crucial, which is usually accomplished by medical imaging.

Graph Convolutional Networks (GCN) are a special form of neural networks, that process graph inputs with specific convolutional operations. In image segmentation, GCNs are an interesting approach, and it has been shown that they can outperform other machine learning methods, such as convolutional neural networks [LCZC19, LCZ*21]. Brain imaging processing requires encoding the information of long range connections without losing the pixel

locations [DGP15], this motivates the idea of using graph convolution, as graphs allow features to be placed in between pixels. Graph convolution requires the input to be transformed into a graph.

Despite of the benefits of using GCNs for segmentation, the GCNs have the same problem of explainability as neural network approaches, and they behave as black boxes [GSG*21]. As a result, it is difficult to understand the reasoning behind the processing of the outputs, for example: what is the neural network learns? and what features are relevant in the feed-forward process [GSWS21]. Particularly in the segmentation of brain lesions, a deeper understanding of the internal processes in the neural networks could provide new insights about brain lesion itself. As shown in Section 2, there is no visualization approach in the literature that assists in the visual inspection of a GCN for brain lesions.

This work presents a visualization approach for understanding of

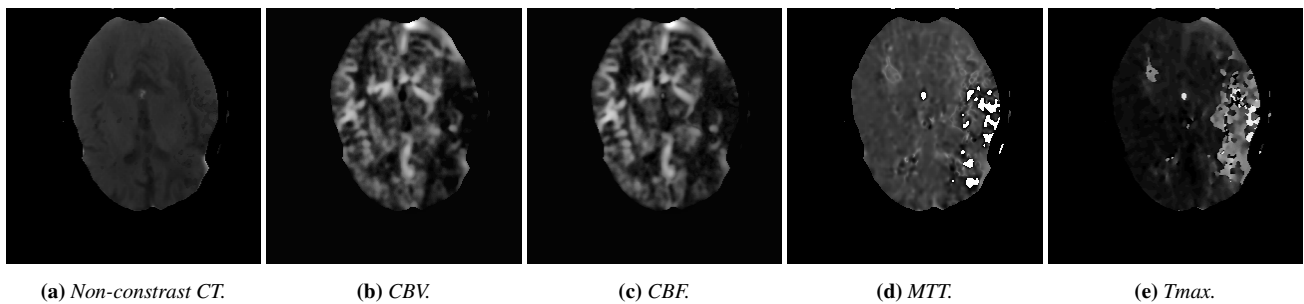


Figure 2: Input modalities in the presented case. a) original CT scan. The remaining modalities are computed based on the b) cerebral blood volume (CBV), c) cerebral blood flow (CBF), d) mean transition time (MTT) and e) time to peak (Tmax).

the relevance of features in deeper layers in a GCN architecture (see Section 3). We provide a visualization of the coarser graph topology processed by the GCN in the so-called latent space [VLBM08]. In the visualization, the graph is displayed overlaying the original image with a color-coding indicating average activation values in the latent space. These activation values show what parts of the input image are most important to the network. In addition, we provide an interactive filtering of interesting activation values.

The contributions of this work are:

- An interactive visualization of activation in GCN
- A case study on the use of our proposed visualization

The effectiveness of the presented approach is shown with an real-world example of brain lesion detection resulting from stroke (see Section 4). This work is discussed (see Section 5), concluded and future directions will be given in Section 6.

2. Related Work

The role of visual analytics when examining machine learning models was highlighted by Liu et al. [LWLZ17] as most machine learning approaches form a black-box. Here, the idea is to use visualization or visual analytics approaches to look inside the model and understand what features are relevant.

Explainable artificial intelligence approaches are available for standard convolutional neural network approaches that tackle Alzheimer’s [ER], brain tumors [STK*18] autism [ERS20] or general anatomical changes [BCT*19]. Although this shows the importance of a transparent machine learning approach, brain lesions from stroke have a different morphology and need therefore be treated differently. In this work, we aim for a specialized visualization that provides insight into a GCN detecting brain lesions from stroke.

Gillmann et al. [GPS*21] presented an activation and occlusion map-based visualization approach to understand what part of the brain has the most effect on the neural network. Although this is a suitable technique to understand what features will be learned in convolutional neural networks, this technique cannot be applied to GCNs.

Baldassarre and Azizpour [BA19] categorized potential approaches for explainable GCN in three categories: sensitivity analysis [GDL03], guided backpropagation [SDBR14], and layer-wise

relevance propagation [MLB*17]. Our visualization approach is closer to the Taylor expansion presented in [MLB*17], where each node is granted the same contribution weight in the brain lesion probability. This paper is also related to the [PKR*19] and serves as a starting point for the explainability of GCNs in the segmentation of brain lesions.

3. Methods

As shown in Section 2, a suitable visualization approach for activation maps in GCNs is missing in the literature, hence the motivation of this work is to propose a method of visualization for this problem. In the following sections, we provide a quick overview of the GCN model and how the visualization of the activation maps is generated. Furthermore, we provide detailed information on our visualization design.

3.1. Dataset and Model

The dataset used to train the network was the ISLES-2018 Challenge dataset [MMvdG*17, KBP*13], as shown in Figure 2. The dataset consists of a non-contrast CT and CT-perfusion (CTP) parameters maps, which we denoted as:

- Non-contrast Computed Tomography (CT)
- Cerebral Blood Flow (CBF)
- Cerebral Blood Volume (CBV)
- Mean Transit Time (MTT)
- Time to peak (Tmax)

The CTP parameters maps are obtained by modeling the flow of contrast agent in a series of CT scans denominated CTP. The maps yield different information about the flow of the contrast agent through the brain tissue. As the interpretation of the original CTP is a hard task, computing parameter maps is a state-of-the-art approach in clinics. The ground truth is obtained from a Diffusion Weighted Image (DWI) obtained a few hours after the CTP took place. The DWI is a gold standard method to detect the core lesion, i.e. the dead tissue, from where the lesion mask is delimited. The original dataset consists of 94 samples with mask information and 63 without mask information. We used only samples with mask so we can fully assess the results.

The model used is based on the architecture of a Fully-Convolutional-Network [LSD15]. We employ spline convolutions

Layer	Output Topology	N ^o features	N ^o nodes
conv 1 a	\mathcal{V}_0	32	65536
conv 1 b	\mathcal{V}_0	32	65536
pool 1	\mathcal{V}_1	32	34697
conv 2 a	\mathcal{V}_1	64	34697
conv 2 b	\mathcal{V}_1	64	34697
pool 2	\mathcal{V}_2	32	18254
conv 3 a	\mathcal{V}_2	128	18254
conv 3 b	\mathcal{V}_2	128	18254
pool 3	\mathcal{V}_3	32	9605
conv 4 a	\mathcal{V}_3	256	9605
conv 4 b	\mathcal{V}_3	256	9605
pool 4	\mathcal{V}_4	256	5048
score fr	\mathcal{V}_4	32	5048
unpool 4	\mathcal{V}_3	32	9605
score 3	\mathcal{V}_3	32	9605
Σ	\mathcal{V}_3	32	9605
unpool 3	\mathcal{V}_2	32	18254
score 2	\mathcal{V}_2	32	18254
Σ	\mathcal{V}_3	32	18254
unpool 2	\mathcal{V}_1	32	34697
unpool 1	\mathcal{V}_0	32	65536
spline conv out	\mathcal{V}_0	1	65536

Table 1: GCN architecture. Notations for the spline convolutional layers is Conv {block number} {index in block}; for pooling layers is pool {index}. Scores are spline convolutional layers with notation: score fr for the latent space; and score {index} for the corresponding graph topology $\mathcal{V}_{\text{index}}$. The column N^o features describes the dimensionality of the feature vector at each node. The column N^o nodes describes the number of nodes for the corresponding graph topology.

[FLWM18], graph pooling [DGK07, DBV16] and unpooling layers to modify the position and value of the node features. The input considered is transformed into a grid graph with nodes connected to eight neighbor nodes. Each node has five channel values corresponding to a pixel value from the CT, CBV, CBF, MTT, and TMax inputs. Therefore, the input is a 8-neighbor interconnected graph from a 256×256 image with five channels.

As shown in Table 1, the architecture of the GCN model is divided into two paths: the downsampling path, and the upsampling path. The downsampling path of the model is constructed with four coarsened graph topologies $\{\mathcal{V}_i\}_{i \in [1,4]}$, where nodes are relocated by the max pooling operations of clustered sub-graphs. Pooling layers are termed *pool i*, with $i \in [1, 4]$ for each new graph topology.

The upsampling path recovers the initial grid topology matching with the ground truth mask. The upsampling path consists of a combination of unpooling layers and convolutional blocks, denoted as *unpool i* and *score i* with $i \in [1, 4]$, respectively. Skip connections are accompanied by a convolution block denoted *score* that normalizes previous outputs to the same number of features to make the sum feature in each node consistent. The non-linear activation function used is an exponential linear unit [CUH15].

The model was trained over 200 epochs using 65 training sam-

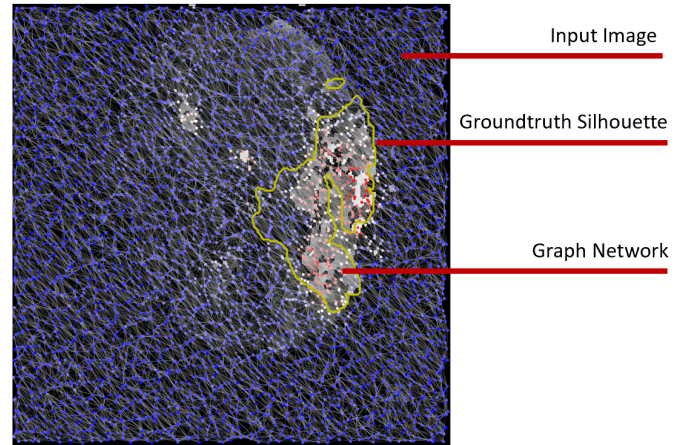


Figure 3: Visualization design of the presented approach. The underlying image modality is overlaid by the silhouette of the ground-truth and the graph structure of the neural network. As it can be seen, no thresholding is applied.

ples and 23 testing samples. We used 6 additional samples for validation and model selection. The evaluation uses a 4-fold cross-validation splits. We used an Adam optimizer [KB15] and learning rate of $1E - 5$. The training was done using a GEFORCE RTX 2080 TI of 11Gb of memory, an Intel(R) Xeon(R) CPU E5-2665 0 @ 2.40GHz, and 124Gb of RAM

3.2. Activation Maps for GCN

The activation maps are computed as the average of activation on each node. Each node $u \in \mathcal{V}$ has an feature vector $x(u) \in \mathbb{R}^d$ of dimension d , therefore the activation map is simply the average value of the vector components $x^i(u) \in \mathbb{R}$:

$$\bar{x}(u) = \frac{1}{d} \sum_{i=1}^d x^i(u) \quad (1)$$

The final calculation is normalized as min-max to have linear range of values and visualize the activation maps as heat-maps with a diverging blue to red color scale.

3.3. Visualization Approach

In order to understand what parts of the input image have higher activation values in the latent space of the GCN, we aimed for an interactive visualization approach. Our visualization consists of three parts, as shown in Figure 3: the input images, a silhouette of the lesion ground-truth, and the graph extracted from the GCN.

The input images are required to provide anatomical context. For the perfusion maps, the color scheme is also changed to greyscale for consistency, and avoid cluttering the color-scale. Here, lighter areas mean higher values in the parameter map and are usually subject to further investigation in clinics. Users are enabled to enable and disable certain image modalities in the visualization.

To highlight the area that should have been detected by the neural

network in a bright yellow. We solely show the silhouette to allow an unoccluded view of the original image.

On top of that, we render the graph network that is inherent in the GCN. Nodes and edges are both visualized. The nodes are color-coded according to the activation value that has been computed. The color-scheme ranges from blue (0 activation) over white (0.5 activation) to red (1.0 activation). As the graph network contains many nodes, we allow interactive thresholding of nodes and connections. The threshold can be set between 0 and 1 and removes all nodes that have an activation value below the threshold as well as all connecting edges to this node.

4. Case Study

In the following, we aim to use our proposed visualization in a real-world scenario to show its effectiveness. Figure 1 shows our visualization applied to one patient selected from the ISLES challenge. The figure includes all image modalities while using a threshold of 0.5 to remove nodes in the graph with less activation. Here, we are looking at the output of the last downsampling layer **pool4**, as highlighted in Table 1. This layer is of particular interest, as the detected features have removed most of the redundant information which results in the most compressed representation.

Figure 1a shows our visualization applied to the input CT scan. When reviewing the original CT scan, it is hard to recognize what parts of the brain are affected by the stroke. There exist minor clusters in the right lobe of the brain that indicate the stroke. When comparing this to the labeled ground-truth we can see, that these parts do not reflect the entire stroke lesion. When comparing these areas to activated areas in the neural network, we can identify a large overlap. This indicates, that the network was able to classify these areas as stroke lesions. Interestingly, larger parts of the resulting lesion do not create activation in the network.

Figure 1b and 1c show very similar results. This is due to the nature of these parameter maps. While CBV aims to capture the amount of blood that flows through the vascular system of the brain, CBF aims to capture the strength of this flow. Naturally, these activation maps are highly interconnected. When reviewing the original images, we identify a dark spot in the area of the stroke lesion. This means, that there is almost no or no blood flow detected. When reviewing the graph network, we can identify that the network gets activated by these areas, which shows that the network successfully identified these features.

Regarding the mean transit time (MTT), Figure 1d provides a very distinct activation of light areas. The mean transit time described the average time, in seconds, that red blood cells spend within a determinate volume of capillary circulation. High values mean that the blood is not transported properly, which is the direct result of the stroke. In contrast to other image modalities, the network makes a clear separation between these areas and surrounding pixels. This means, that the network successfully learns to use these features to detect a stroke.

Figure 1e shows the activation map regarding the Tmax parameter map. This image modality encodes the time to the maximal contrast concentration to each voxel. For interpretation, high values indicate a slow blood flow and therefore poor oxygenation. In

a similar way, the graph network apparently follows this pattern, as shown in the higher activation values.

Interestingly, the network also reacts to two small spots located outside the actual lesion. Here, our visualization indicated that these spots cause false attention in the neural network.

In summary, we can see that our visualization approach is able to reason the decision-making process of the GCN that we considered.

5. Discussion

In the discussion of the presented visualization approach, we consider the tasks in medical practice and research [EMML22], namely: *overview*, *contextualization*, *quantification*, *navigation* and *comparison*.

Our visualization approach provides *Overview* over the data and the activations in the GCN, as it combines the original image, the ground-truth, and the activation of the graph network. As shown in Section 4, we are able to *identify* the features that are detected by the neural network and also outliers. The latter are identified as activation regions in the neural network outside the targeted lesion. *Contextualization* is provided as we show the activation graphs on all image modalities. *Quantification* is provided by using and visualizing activation map values. The interactive thresholding of nodes values in the graph network allows users to *navigate* to interesting areas in the network. For *comparison*, the presented visualization is provided for each of the used modalities. As shown, the side-by-side views allow understanding of the differences in activation of modalities.

Our approach is limited to two-dimensional data. Further, each imaging modality is considered individually. The activation maps in this visualization approach can be used to debug the network and get an insight what features are relevant. For example, a successful segmentation should have high activation values within the ground truth, and this information can be used to refine the dataset or to identify labeling problems.

6. Conclusion

In this work, we provide an interactive visualization that provides an overview of activation in a GCN. Our visualization fuses the input image, the targeted ground-truth, and the graph network that is color-coded to indicate activation of each node with an intuitive color-map. In addition, interaction is provided to filter nodes with high activation values in the GCN. We show the effectiveness of the presented approach by successfully applying it to a real-world case study.

For future work, we aim to extend the presented approach into the three-dimensional space and provide an open-source tool, where users can upload their models and explore them.

References

- [BA19] BALDASSARRE F., AZIZPOUR H.: Explainability techniques for graph convolutional networks, 2019. URL: <https://arxiv.org/abs/1905.13686>, doi:10.48550/ARXIV.1905.13686. 2

- [BCT*19] BIFFI C., CERROLAZA J. J., TARRONI G., BAI W., DE MARVAO A., OKTAY O., LEDIG C., FOLGOC L. L., KAMNITSAS K., DOUMOU G., DUAN J., PRASAD S. K., COOK S. A., O'REGAN D. P., RUECKERT D.: Explainable anatomical shape analysis through deep hierarchical generative models, 2019. URL: <https://arxiv.org/abs/1907.00058>, doi:10.48550/ARXIV.1907.00058. 2
- [CUH15] CLEVERT D.-A., UNTERTHINER T., HOCHREITER S.: Fast and accurate deep network learning by exponential linear units (elus), 2015. URL: <https://arxiv.org/abs/1511.07289>, doi:10.48550/ARXIV.1511.07289. 3
- [DBV16] DEFFERRARD M., BRESSON X., VANDERGHEYNST P.: Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems* 29 (2016). 3
- [DGK07] DHILLON I. S., GUAN Y., KULIS B.: Weighted graph cuts without eigenvectors a multilevel approach. *IEEE transactions on pattern analysis and machine intelligence* 29, 11 (2007), 1944–1957. 3
- [DGP15] DESPOTOVIC I., GOOSSENS B., PHILIPS W.: MRI segmentation of the human brain: Challenges, methods, and applications. *Computational and Mathematical Methods in Medicine* 2015 (2015). 1
- [EMML22] EULZER P., MEUSCHKE M., MISTELBAUER G., LAWONN K.: Vessel maps: A survey of map-like visualizations of the cardiovascular system. *Computer Graphics Forum* 41 (06 2022). 4
- [ER] EITEL F., RITTER K.: URL: <https://arxiv.org/abs/1909.08856>, doi:10.48550/ARXIV.1909.08856. 2
- [ERS20] ESLAMI T., RAIKER J. S., SAEED F.: Explainable and scalable machine-learning algorithms for detection of autism spectrum disorder using fMRI data, 2020. URL: <https://arxiv.org/abs/2003.01541>, doi:10.48550/ARXIV.2003.01541. 2
- [FLWM18] FEY M., LENSSEN J. E., WEICHERT F., MÜLLER H.: Splinecn: Fast geometric deep learning with continuous b-spline kernels. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 869–877. 3
- [GDL03] GEVREY M., DIMOPOULOS I., LEK S.: Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 160, 3 (2003), 249–264. Modelling the structure of aquatic communities: concepts, methods and problems. URL: <https://www.sciencedirect.com/science/article/pii/S0304380002002570>, doi:https://doi.org/10.1016/S0304-3800(02)00257-0. 2
- [GPS*21] GILLMANN C., PETER L., SCHMIDT C., SAUR D., SCHEUERMANN G.: Visualizing multimodal deep learning for lesion prediction. *IEEE Computer Graphics and Applications* 41, 05 (2021), 90–98. 2
- [GSG*21] GILLMANN C., SMIT N. N., GRÖLLER E., PREIM B., VILANOVA A., WISCHGOLL T.: Ten open challenges in medical visualization. *IEEE Computer Graphics and Applications* 41, 5 (2021), 7–15. doi:10.1109/MCG.2021.3094858. 1
- [GSWS21] GILLMANN C., SAUR D., WISCHGOLL T., SCHEUERMANN G.: Uncertainty-aware visualization in medical imaging - a survey. *Computer Graphics Forum* 40, 3 (2021), 665–689. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14333>, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14333, doi:https://doi.org/10.1111/cgf.14333. 1
- [KB15] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. iclr. 2015. *arXiv preprint arXiv:1412.6980* 9 (2015). 3
- [KBP*13] KISTLER M., BONARETTI S., PFAHRER M., NIKLAUS R., BÜCHLER P., ET AL.: The virtual skeleton database: an open access repository for biomedical research and collaboration. *Journal of medical Internet research* 15, 11 (2013), e2930. 2
- [LCZ*21] LU Y., CHEN Y., ZHAO D., LIU B., LAI Z., CHEN J.: Cnn-g: Convolutional neural network combined with graph for image segmentation with theoretical analysis. *IEEE Transactions on Cognitive and Developmental Systems* 13, 3 (2021), 631–644. doi:10.1109/tcds.2020.2998497. 1
- [LCZC19] LU Y., CHEN Y., ZHAO D., CHEN J.: Graph-fcn for image semantic segmentation. *Advances in Neural Networks – ISNN 2019* (2019), 97–105. doi:10.1007/978-3-030-22796-8_11. 1
- [LSD15] LONG J., SELHAMER E., DARRELL T.: Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440. 2
- [LWLZ17] LIU S., WANG X., LIU M., ZHU J.: Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics I*, 1 (2017), 48–56. URL: <https://www.sciencedirect.com/science/article/pii/S2468502X17300086>, doi:https://doi.org/10.1016/j.visinf.2017.01.006. 2
- [MLB*17] MONTAVON G., LAPUSCHKIN S., BINDER A., SAMEK W., MÜLLER K.-R.: Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition* 65 (2017), 211–222. URL: <https://www.sciencedirect.com/science/article/pii/S0031320316303582>, doi:https://doi.org/10.1016/j.patcog.2016.11.008. 2
- [MMvdG*17] MAIER O., MENZE B. H., VON DER GABLENTZ J., HÄNI L., HEINRICH M. P., LIEBRAND M., WINZECK S., BASIT A., BENTLEY P., CHEN L., ET AL.: Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical image analysis* 35 (2017), 250–269. 2
- [PKR*19] POPE P. E., KOLOURI S., ROSTAMI M., MARTIN C. E., HOFFMANN H.: Explainability methods for graph convolutional neural networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 10764–10773. doi:10.1109/CVPR.2019.01103. 2
- [SDBR14] SPRINGENBERG J. T., DOSOVITSKIY A., BROX T., RIEDMILLER M.: Striving for simplicity: The all convolutional net, 2014. URL: <https://arxiv.org/abs/1412.6806>, doi:10.48550/ARXIV.1412.6806. 2
- [SMS*20] SCHARDT K., MAACK R. G. C., SAUER D., HAGEN H., SCHEUERMANN G., GILLMANN C.: Multi-modal visualization of stroke lesion ct-imaging, Nov 2020. URL: osf.io/qk39a, doi:10.31219/osf.io/qk39a. 1
- [STK*18] STOYANOV D., TAYLOR Z., KIA S. M., OGUZ I., REYES M., MARTEL A., MAIER-HEIN L., MARQUAND A. F., DUCHESNAY E., LÖFSTEDT T., LANDMAN B., CARDOSO M. J., SILVA C. A., PEREIRA S., MEIER R. (Eds.): *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer International Publishing, 2018. URL: <https://doi.org/10.1007/978-3-030-02628-8>, doi:10.1007/978-3-030-02628-8. 2
- [VLBM08] VINCENT P., LAROCHELLE H., BENGIO Y., MANZAGOL P.-A.: Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning - ICML '08* (2008). doi:10.1145/1390156.1390294. 2