Machine Learning for Health
Workshop, NIPS 2016

# Understanding Anatomy Classification using Visualization

Devinder Kumar[1], Vlado Menkovski[2]

[1]Philips Research, [2]Eindhoven University of Technology

## Introduction

- Understanding the decision process of a deep neural network model for classification can be challenging due to the very large number of parameters and model's tendency to represent the information internally in a distributed manner.
- Distributed representations have significant advantages for the capability of the model to generalize well [3], however the trade-off is the difficulty in communicating the model's reasoning In other words,
- One direction towards understanding how CNNs processes the information internally is through visualization. The work of Zeiler et. al [2] & Mahendran et. al [1], have shown that the inner working of the CNN can be projected back to the image space in a comprehensible way to a human expert.
- We build on the work of [2] and present an approach to understand the decision making process of these networks through visualizing the information used as a part of this process. We apply this approach to the anatomy classification problem using X-ray images.

## Method

- We construct an informed approach to designing models by building three different deep CNN models with different architectures and hyper-parameters: shallow CNN, deeper CNN without data augmentation and deeper CNN with data augmentation inspired by the work of Razavian et al. [3].
- It is done by visualizing the top-n most *activated* activation filters of the last conv layer in the above described models.

**Tabel 1 Accuracy in percent for three different networks trained on the imageClef 2009 annotation task**

| Shallow Network | Deeper Network | Deeper Net + data aug |
|---|---|---|
| 71.1 | 90.36 | **95.62** |

- Top *n* filters are used to visualize the parts of the input image that the network considers *important*.
- The back projection to input space is achieved by using the fractionally strided convolutional technique [2] through another parallel network with transposed convolution filters and switches for un-pooling.
- Next, we then examine the correlation of those regions obtained through visualization with identified regions and shapes of image landmarks that are mentioned in the medical radiology literature.

## Results

We train the three different networks for all the 24 anatomy classes simultaneously from ImageClef dataset. The results obtained by training the three models are shown on the other side.

## Conclusion

- We propose an approach that allows for evaluating the decision making process of CNNs.
- We show that the design of the model architectures for deep CNN and the training procedure does not necessarily need to be a trial-and-error process, solely focused on optimizing the test set accuracy. Through visualization we managed to incorporate domain knowledge.
- Furthermore, visually understanding the information involved in the model decision allows for more confidence in its performance on unseen data.

**Figure 1: Focus area of the top 5 last conv layer of the shallow network**

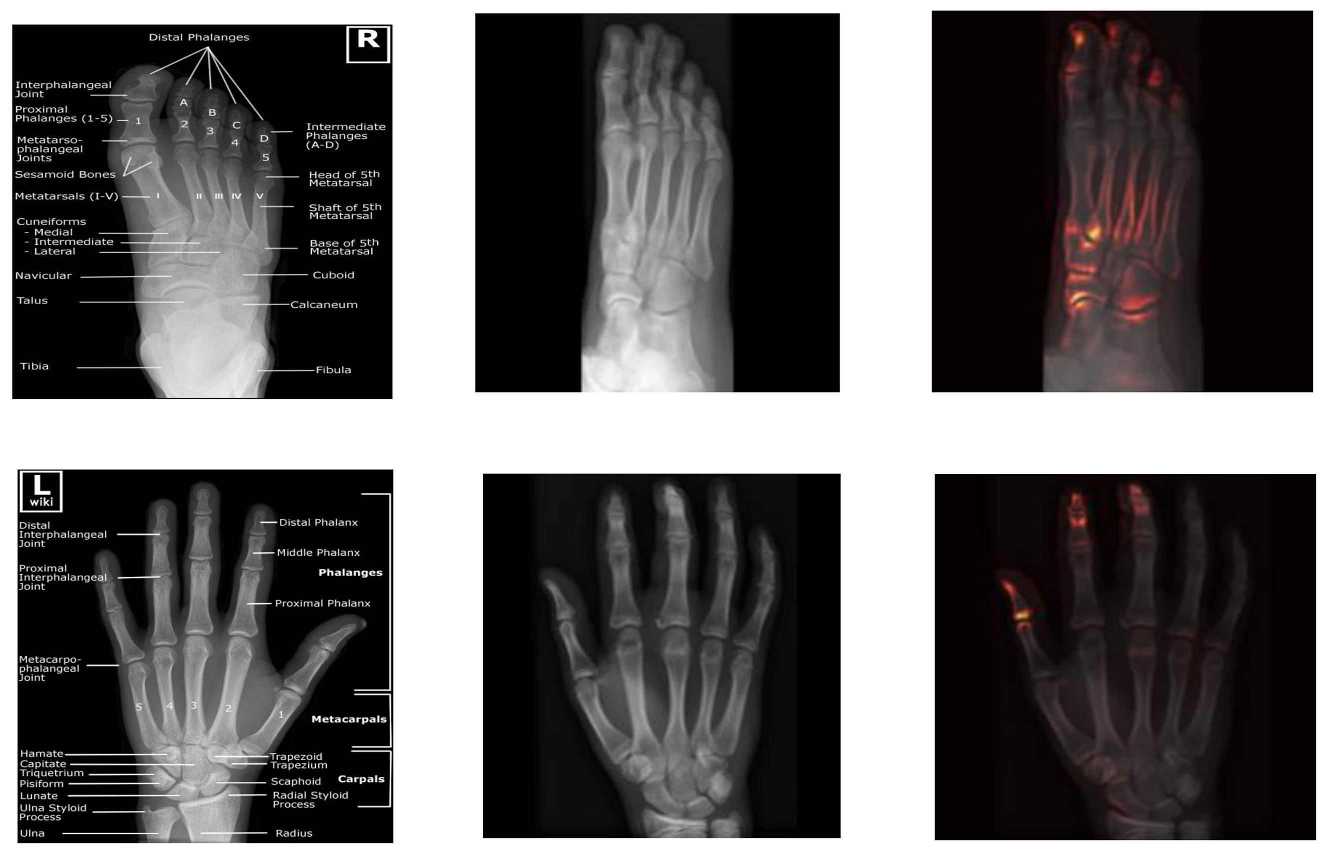

**Figure 2: Correspondence between anatomical description of found in literature that are used by human experts original images & the heat maps overlaid on the original images**

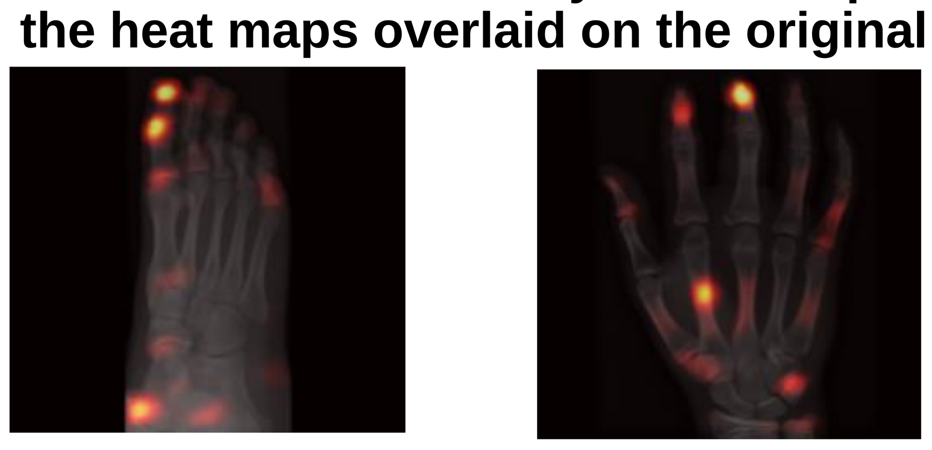

**Figure 3: Heat maps overlaid on the original images from the last conv layer of the deeper network with no data aug.**

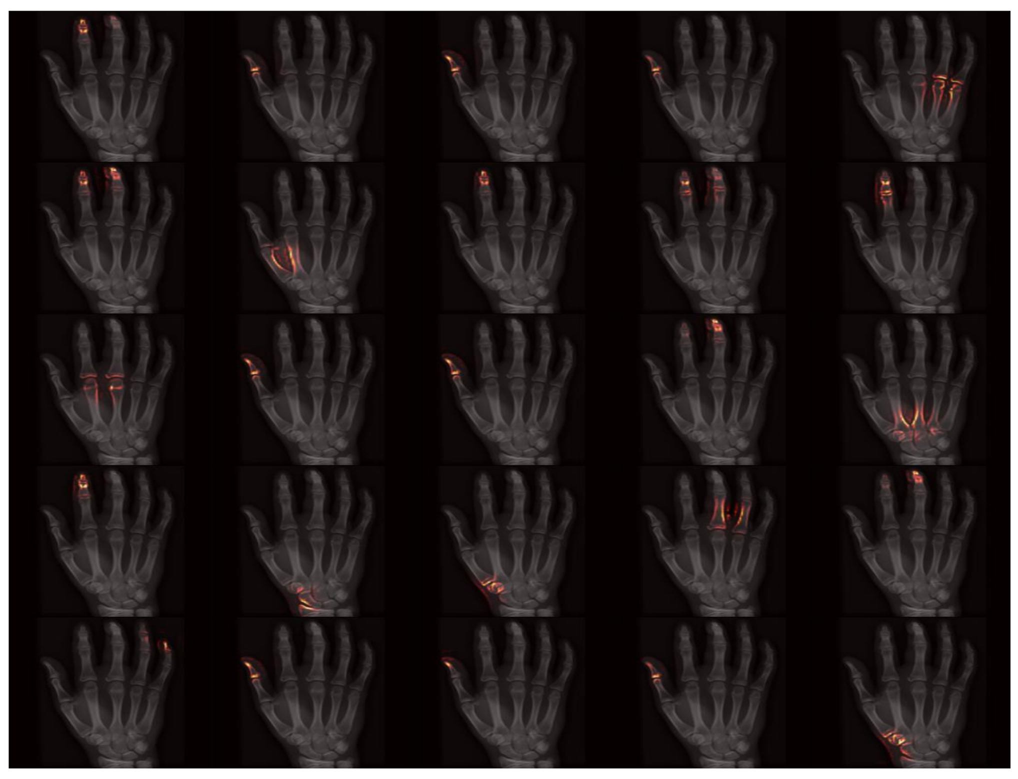

**Figure 4: Focus areas for top 25 filter in the last conv layer of trained Net.**

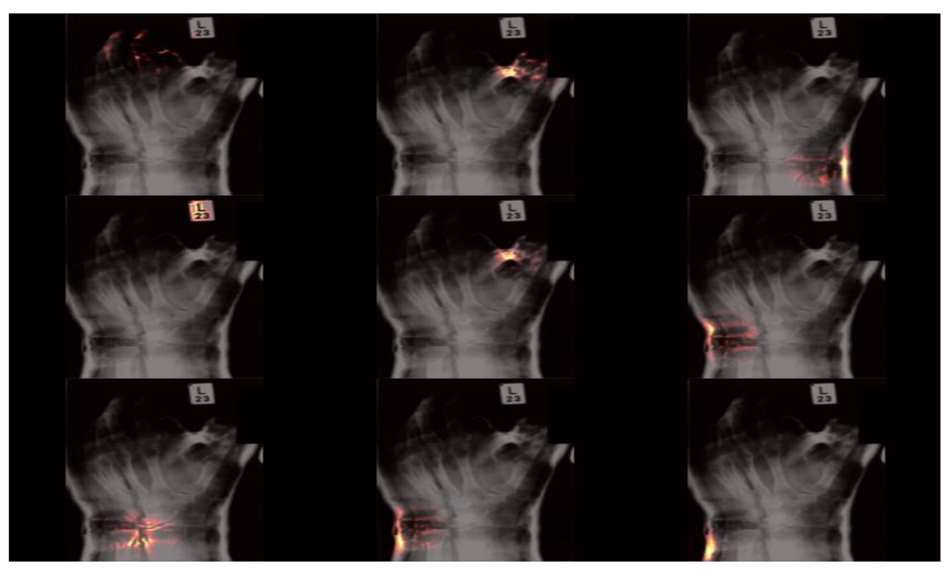

**Figure 5: Results from the last conv layer of the deeper network with aug for hand class mis-classified as cranium.**

## References

[1] A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In IEEE conf. on computer vision and pattern recognition (CVPR), pages 110 5188–5196, 2015.
[2] M. D Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In European Conf. on Computer Vision, Springer, 2014.
[3] A. Razavian, J. Sullivan, A. Maki, and S. Carlsson. A baseline for visual instance retrieval with deep convolutional networks. arXiv:1412.6574, 2014