Technical Report:
Combining knowledge from Transfer Learning during training and Wide Resnets

Wolfgang Fuhl
Mühringen, 72160
wolfgang-fuhl@gmx.de

June 20, 2022

Abstract

In this report, we combine the idea of Wide ResNets and transfer learning to optimize the architecture of deep neural networks. The first improvement of the architecture is the use of all layers as information source for the last layer. This idea comes from transfer learning, which uses networks pre-trained on other data and extracts different levels of the network as input for the new task. The second improvement is the use of deeper layers instead of deeper sequences of blocks. This idea comes from Wide ResNets. Using both optimizations, both high data augmentation and standard data augmentation can produce better results for different models. Link: https://github.com/wolfgangfuhl/PublicationStuff/tree/master/TechnicalReport1/Supp

1 Introduction

DNNs [30] have found their way into a variety of fields [1, 27]. In eye tracking [31, 49], they are already used for scanpath analysis [32, 33, 4], as well as other approaches based on machine learning [24, 2, 3, 12], feature extraction [49] such as pupil [28, 18, 25, 48, 47, 46, 11, 26, 16, 10, 8, 45, 6], iris [9, 19, 5] and eyelids [22, 21, 23], eyeball estimation [7], and also for eye movement classification [41, 42, 14]. In recent years, there have been a number of new large data sets [38] that have also been annotated using modern machine learning approaches. Another important aspect is the anonymisation of the data [34, 20], in which more and more research is done. Other areas include human computer interaction, medicine, robotics, industrial sensor evaluation, and many more.

A lot of research is also being done in the field of visualization [15, 17, 40]. Here one tries to illustrate the data with different statistics and distributions. The visualizations also provide a way to better understand deep neural networks and other machine learning methods. In the field of data analysis, machine learning is often used to visualize values as functions and to reduce high-dimensional feature spaces. Based on the undestanding, it is also important to validate models [13, 43, 44].

With all the applications and improvements achieved by machine learning and in particular by deep neural networks, this research field has established itself as state of the art. Research in the field of machine learning is also continuously delivering new improvements such as tree convolution [39], neural pooling [35], rotating convolution operators [36] and many more [37]. In this work, we describe the combination of two existing ideas that allow to obtain better generalized networks and to reduce the loss of information due to spatial reduction. Together with the idea of wide ResNets, parameters and operations in networks can be saved.

2 Method

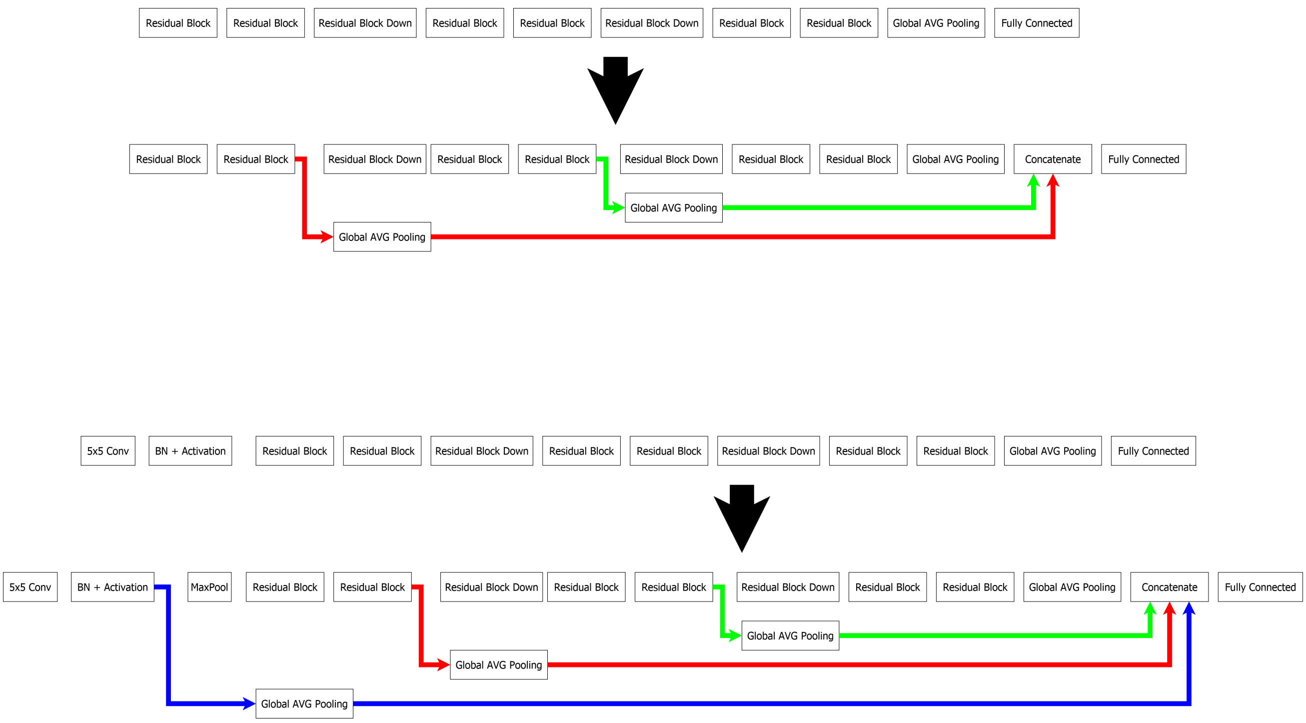

Figure 1: Shows the basic procedure for optimizing an architecture with the presented method. Before each residual block with a spatial reduction, an additional connection is added. This connection performs an average pooling along the depth of the layer and provides the result to the last layer as an additional source of information. The residual nets are shown here only partially since to reduce the size of the image.

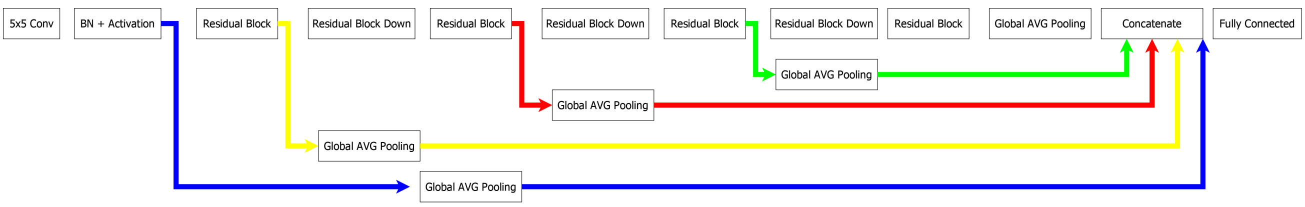

Figure 2: The proposed architecture. Each residual stage has double the depth of the standard ResNet-32 and the propagation to the last fully connected layer in combination with the depthwise pooling. Additionally, we have a convolution + BN + activation at the beginning, as it is usually done for large image inputs like image net.

In Figure 1 it is shown how to transform nets with residual [50] or maximum propagation [29] connections to the proposed architecture. This can be applied for all networks and steams from the idea of using multiple layers of a feature extraction network.

In Figure 2 a new architecure is shown. It has one additional convolution in the beginning (Similar to resnets for larger inputs like image net). Compared to ResNets all levels are decreased to only contain two residual blocks. Additionally, each level has a depthwise pooling and forward propagation integrated.

3 Evaluation

Table 1: The results of the proposed approch to standard models on Cifar 100 [51]. We used the standard train and validation split as well as we used the same resolution of 32 by 32. We also used only one GPU (RTX 3050 ti 8GB RAM).
Shift + Flip: Cropping a image region (32 by 32) out of the original image with a possible shift of 4 in each direction. Pixels outside the original image are filled with zeros. Flipping was done left right only. We used the standard scheduler reduction of 0.1 after each 100 epochs, an initial learning rate of 0.01 and a batch size of 10 with the SGD and Momentum optimizer. Each net was trained for 400 epochs.
For WideResNet [53] and FastAutoAugment [52] we used the implementation provided here https://github.com/kakaobrain/fast-autoaugment. They use the sinus scheduler and many additional data augmentations. We had to reduce the batch size to half due to the available RAM on the GPU.

Model	Augmentation	Cifar 100 Accuracy
ResNet32	Shift + Flip	76.7
ResNet32 + Proposed	Shift + Flip	77.6
ResNet32 (Double depth)	Shift + Flip	76.9
ResNet32 (Double depth) + Proposed	Shift + Flip	78.11
NewNet (Figure 2)	Shift + Flip	78.94
NewNet (Figure 2, double depth)	Shift + Flip	79.67
WideResNet 28x10	FastAutoAugment	82.01
WideResNet 28x10 + Proposed	FastAutoAugment	83.52

4 Acknowledgements

I know that evaluations on multiple data sets as well as more models are neccessary to show the improvement. But since I have to pay for the energy and I am not employed at a research department, I leave this to others if they are interested.

References

[1] W. Fuhl. *Image-based extraction of eye features for robust eye tracking*. PhD thesis, University of Tübingen, 04 2019.

[2] W. Fuhl, N. Castner, and E. Kasneci. Histogram of oriented velocities for eye movement detection. In *International Conference on Multimodal Interaction Workshops, ICMIW*, 2018.

[3] W. Fuhl, N. Castner, and E. Kasneci. Rule-based learning for eye movement type detection. In *International Conference on Multimodal Interaction Workshops, ICMIW*, 2018.

[4] W. Fuhl, N. Castner, T. C. Kübler, A. Lotz, W. Rosenstiel, and E. Kasneci. Ferns for area of interest free scanpath classification. In *Proceedings of the 2019 ACM Symposium on Eye Tracking Research & Applications (ETRA)*, 06 2019.

[5] W. Fuhl, N. Castner, L. Zhuang, M. Holzer, W. Rosenstiel, and E. Kasneci. Mam: Transfer learning for fully automatic video annotation and specialized detector creation. In *International Conference on Computer Vision Workshops, ICCVW*, 2018.

[6] W. Fuhl, S. Eivazi, B. Hosp, A. Eivazi, W. Rosenstiel, and E. Kasneci. Bore: Boosted-oriented edge optimization for robust, real time remote pupil center detection. In *Eye Tracking Research and Applications, ETRA*, 2018.

[7] W. Fuhl, H. Gao, and E. Kasneci. Neural networks for optical vector and eye ball parameter estimation. In *ACM Symposium on Eye Tracking Research & Applications, ETRA 2020*. ACM, 01 2020.

[8] W. Fuhl, H. Gao, and E. Kasneci. Tiny convolution, decision tree, and binary neuronal networks for robust and real time pupil outline estimation. In *ACM Symposium on Eye Tracking Research & Applications, ETRA 2020*. ACM, 01 2020.

[9] W. Fuhl, D. Geisler, W. Rosenstiel, and E. Kasneci. The applicability of cycle gans for pupil and eyelid segmentation, data generation and image refinement. In *International Conference on Computer Vision Workshops, ICCVW*, 11 2019.

[10] W. Fuhl, D. Geisler, T. Santini, T. Appel, W. Rosenstiel, and E. Kasneci. Cbf: Circular binary features for robust and real-time pupil center detection. In *ACM Symposium on Eye Tracking Research & Applications*, 06 2018.

[11] W. Fuhl, D. Geisler, T. Santini, and E. Kasneci. Evaluation of state-of-the-art pupil detection algorithms on remote eye images. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct publication – PETMEI 2016*, 09 2016.

[12] W. Fuhl and E. Kasneci. Eye movement velocity and gaze data generator for evaluation, robustness testing and assess of eye tracking software and visualization tools. In *Poster at Egocentric Perception, Interaction and Computing, EPIC*, 2018.

[13] W. Fuhl and E. Kasneci. Learning to validate the quality of detected landmarks. In *International Conference on Machine Vision, ICMV*, 11 2019.

[14] W Fuhl and E Kasneci. A multimodal eye movement dataset and a multimodal eye movement segmentation analysis. *arXiv preprint arXiv:2101.04318*, 01 2021.

[15] W. Fuhl, T. C. Kübler, H. Brinkmann, R. Rosenberg, W. Rosenstiel, and E. Kasneci. Region of interest generation algorithms for eye tracking data. In *Third Workshop on Eye Tracking and Visualization (ETVIS), in conjunction with ACM ETRA*, 06 2018.

[16] W. Fuhl, T. C. Kübler, D. Hospach, O. Bringmann, W. Rosenstiel, and E. Kasneci. Ways of improving the precision of eye tracking data: Controlling the influence of dirt and dust on pupil detection. *Journal of Eye Movement Research*, 10(3), 05 2017.

[17] W. Fuhl, T. C. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci. Arbitrarily shaped areas of interest based on gaze density gradient. In *European Conference on Eye Movements, ECEM 2015*, 08 2015.

[18] W. Fuhl, T. C. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci. Excuse: Robust pupil detection in real-world scenarios. In *16th International Conference on Computer Analysis of Images and Patterns (CAIP 2015)*, 09 2015.

[19] W. Fuhl, W. Rosenstiel, and E. Kasneci. 500,000 images closer to eyelid and pupil segmentation. In *Computer Analysis of Images and Patterns, CAIP*, 11 2019.

[20] W Fuhl, N Sanamrad, and E Kasneci. The gaze and mouse signal as additional source for user fingerprints in browser applications. *arXiv preprint arXiv:2101.03793*, 01 2021.

[21] W. Fuhl, T. Santini, D. Geisler, T. C. Kübler, and E. Kasneci. Eyelad: Remote eye tracking image labeling tool. In *12th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, 02 2017.

[22] W. Fuhl, T. Santini, D. Geisler, T. C. Kübler, W. Rosenstiel, and E. Kasneci. Eyes wide open? eyelid location and eye aperture estimation for pervasive eye tracking in real-world scenarios. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct publication – PETMEI 2016*, 09 2016.

[23] W. Fuhl, T. Santini, and E. Kasneci. Fast and robust eyelid outline and aperture detection in real-world scenarios. In *IEEE Winter Conference on Applications of Computer Vision (WACV 2017)*, 03 2017.

[24] W. Fuhl, T. Santini, T. Kuebler, N. Castner, W. Rosenstiel, and E. Kasneci. Eye movement simulation and detector creation to reduce laborious parameter adjustments. *arXiv preprint arXiv:1804.00970*, 2018.

[25] W. Fuhl, T. Santini, T. C. Kübler, and E. Kasneci. Else: Ellipse selection for robust pupil detection in real-world environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 123–130, 03 2016.

[26] W. Fuhl, T. Santini, C. Reichert, D. Claus, A. Herkommer, H. Bahmani, K. Rifai, S. Wahl, and E. Kasneci. Non-intrusive practitioner pupil detection for unmodified microscope oculars. *Elsevier Computers in Biology and Medicine*, 79:36–44, 12 2016.

[27] Wolfgang Fuhl. From perception to action using observed actions to learn gestures. *User Modeling and User-Adapted Interaction*, pages 1–18, 08 2020.

[28] Wolfgang Fuhl. 1000 pupil segmentations in a second using haar like features and statistical learning. *arXiv preprint arXiv:2102.01921*, 2021.

[29] Wolfgang Fuhl. Maximum and leaky maximum propagation. *arXiv preprint arXiv:2105.10277*, 2021.

[30] Wolfgang Fuhl. Tensor normalization and full distribution training. *arXiv preprint arXiv:2109.02345*, 2021.

[31] Wolfgang Fuhl. Groupgazer: A tool to compute the gaze per participant in groups with integrated calibration to map the gaze online to a screen or beamer projection. *arXiv preprint arXiv:2201.07692*, 2022.

[32] Wolfgang Fuhl. Hpcgen: Hierarchical k-means clustering and level based principal components for scan path genaration. *arXiv preprint arXiv:2201.08354*, 2022.

[33] Wolfgang Fuhl, Efe Bozkir, Benedikt Hosp, Nora Castner, David Geisler, Thiago C Santini, and Enkelejda Kasneci. Encodji: encoding gaze data into emoji space for an amusing scanpath classification approach. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–4, 2019.

[34] Wolfgang Fuhl, Efe Bozkir, and Enkelejda Kasneci. Reinforcement learning for the privacy preservation and manipulation of eye tracking data. In *International Conference on Artificial Neural Networks*, pages 595–607. Springer, 2021.

[35] Wolfgang Fuhl and Enkelejda Kasneci. Multi layer neural networks as replacement for pooling operations. *arXiv preprint arXiv:2006.06969*, 08 2020.

[36] Wolfgang Fuhl and Enkelejda Kasneci. Rotated ring, radial and depth wise separable radial convolutions. *arXiv preprint arXiv:2010.00873*, 08 2020.

[37] Wolfgang Fuhl and Enkelejda Kasneci. Weight and gradient centralization in deep neural networks. *arXiv preprint arXiv:2010.00866*, 08 2020.

[38] Wolfgang Fuhl, Gjergji Kasneci, and Enkelejda Kasneci. Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 367–375. IEEE, 2021.

[39] Wolfgang Fuhl, Gjergji Kasneci, Wolfgang Rosenstiel, and Enkeljda Kasneci. Training decision trees as replacement for convolution layers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3882–3889, 2020.

[40] Wolfgang Fuhl, Thomas C Kübler, Thiago Santini, and Enkelejda Kasneci. Automatic generation of saliency-based areas of interest for the visualization and analysis of eye-tracking data. In *VMV*, pages 47–54, 2018.

[41] Wolfgang Fuhl, Yao Rong, and Kasneci Enkelejda. Fully convolutional neural networks for raw eye tracking data segmentation, generation, and reconstruction. *arXiv preprint arXiv:2010.00821*, 08 2020.

[42] Wolfgang Fuhl, Yao Rong, and Kasneci Enkelejda. Fully convolutional neural networks for raw eye tracking data segmentation, generation, and reconstruction. In *Proceedings of the International Conference on Pattern Recognition*, pages 0–0, 2020.

[43] Wolfgang Fuhl, Yao Rong, Thomas Motz, Michael Scheidt, Andreas Hartel, Andreas Koch, and Enkelejda Kasneci. Explainable online validation of machine learning models for practical applications. In *Proceedings of the International Conference on Pattern Recognition*, pages 0–0, 2020.

[44] Wolfgang Fuhl, Yao Rong, Thomas Motz, Michael Scheidt, Andreas Hartel, Andreas Koch, and Enkelejda Kasneci. Explainable online validation of machine learning models for practical applications. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3304–3311. IEEE, 2021.

[45] Wolfgang Fuhl, Thiago Santini, and Enkelejda Kasneci. Fast camera focus estimation for gaze-based focus control. *arXiv preprint arXiv:1711.03306*, 2017.

[46] Wolfgang Fuhl, Thiago Santini, Gjergji Kasneci, and Enkelejda Kasneci. Pupilnet: Convolutional neural networks for robust pupil detection. *arXiv preprint arXiv:1601.04902*, 2016.

[47] Wolfgang Fuhl, Thiago Santini, Gjergji Kasneci, Wolfgang Rosenstiel, and Enkelejda Kasneci. Pupilnet v2. 0: Convolutional neural networks for cpu based real time robust pupil detection. *arXiv preprint arXiv:1711.00112*, 2017.

[48] Wolfgang Fuhl, Marc Tonsen, Andreas Bulling, and Enkelejda Kasneci. Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art. In *Machine Vision and Applications*, pages 1–14, 06 2016.

[49] Wolfgang Fuhl, Daniel Weber, and Enkelejda Kasneci. Pistol: Pupil invisible supportive tool to extract pupil, iris, eye opening, eye movements, pupil and iris gaze vector, and 2d as well as 3d gaze. *arXiv preprint arXiv:2201.06799*, 2022.

[50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[51] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[52] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. *Advances in Neural Information Processing Systems*, 32, 2019.

[53] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.