# Data augmentation for deep learning-based gaze estimation.

Author: Jorn Dijk, jwdijk@student.tudelft.nl
Supervisor: Lingyu Du, lingyu.du@tudelft.nl - Responsible professor: Guohao Lan g.lan@tudelft.nl
EEMCS, Delft University of Technology, The Netherlands

## Background

Gaze can be used in many applications:

- Human-Robot interaction [1]
- Predicting actions from humans [2]
- Processing social signals [3]

Data augmentation is often used to improve the accuracy convolutional neural networks, but is not yet proven to work on the regression problem of gaze estimation.

**"What effect do different data augmentations on images have on the mean angular error of gaze estimation using convolutional neural networks?"**

## Methodology

### Baseline model (B)

Results of applying data augmentations are compared to baseline models, which are altered versions of AlexNet [4] and ResNet18 [5] shown in figure 1. The convolutional layers of both models stay the same.

### Data augmentations

Training the baseline model on images altered with data augmentations to see it's effects.

### Geometric transformations:

- Random rotation (R) with (+L) and without (-L) rotating the labels
- Random flipping (F) with (+L) and without (-L) flipping the labels
- Circular shifting (S)
- Random cropping (C)
- Random translation (T)

### Appearance tranformations:

- Gaussian noise injection (N)
- Colour jitter (J)
- Gaussian blur (B)
- Erasing (E)

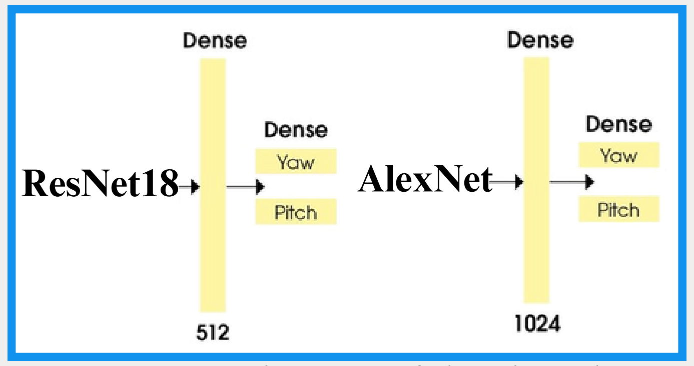


*Figure 1: Visualization of the altered part of the two used models*

## Visualisation of data augmentations

*Figure 2: Random rotation*

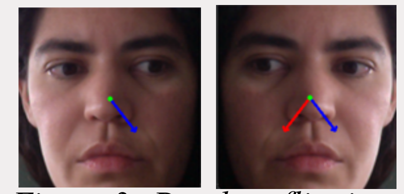

*Figure 3: Random flipping*

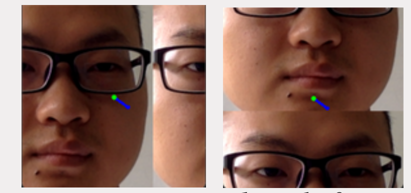

*Figure 4: Circular shifting*

*Figure 5: Random cropping*

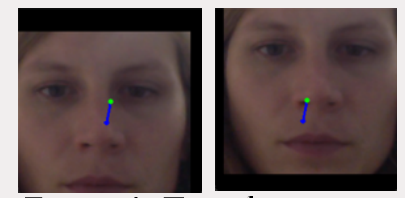

*Figure 6: Translation*

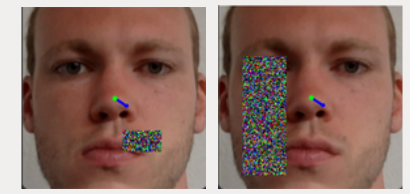

*Figure 7: Random erasure*

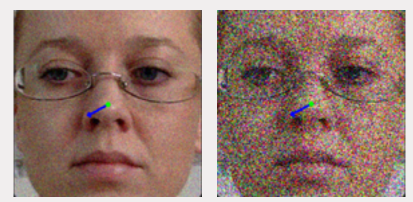

*Figure 8: Noise injection*

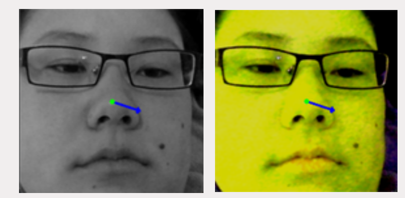

*Figure 9: Color jitter*

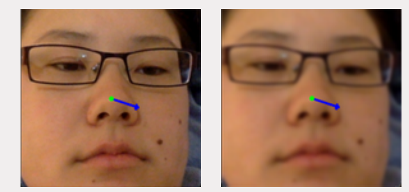

*Figure 10: Gaussian blurring*

## Results

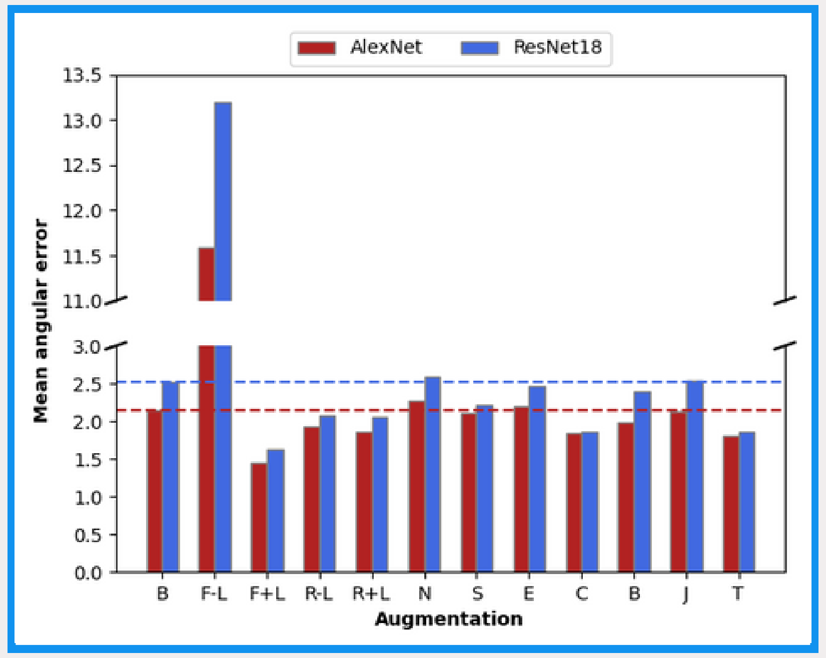


*Figure 11: Results of all data augmentation methods*

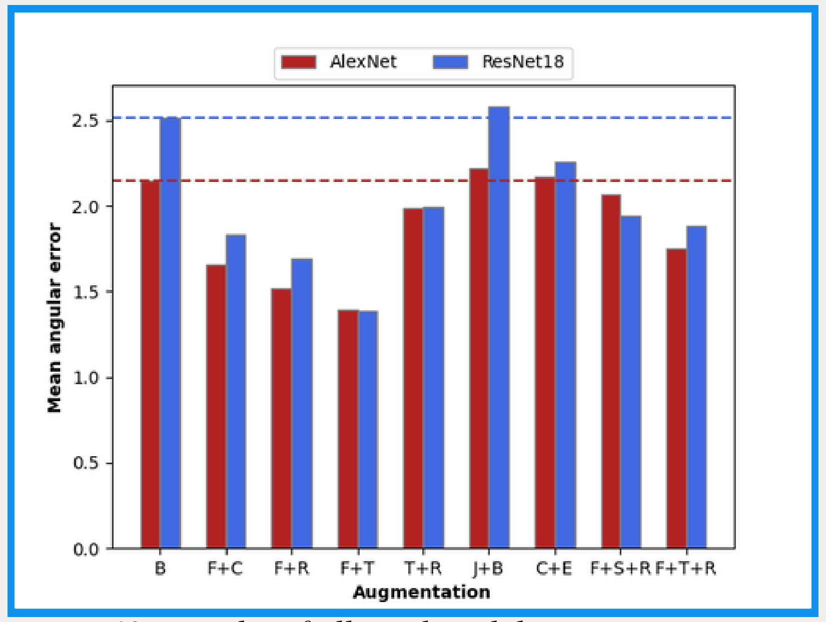


*Figure 12: Results of all combined data augmentation methods*

## Conclusion

- **Flipping when flipping the labels accordingly** gives the best results for individually applied data augmentation methods with improvements of 33% for AlexNet and 34% for ResNet18.
- Changes caused by **small geometric transformations** give the best improvements in comparison to the baseline models.
- The results of the applied combinations in figure 12 show that flipping together with translation is the only combination that gives better results than its augmentations applied individually.
- **Flipping with translation** gives the best result with accuracies of 1.396 for AlexNet and 1.389 for ResNet18.
- Applying data augmentations can have a positive impact on gaze estimation when applied correct.

## Future work

- Their exists many more data augmentation techniques and combinations that could be studied.
- Delve deeper in why a certain augmentation gives results and look at the influence of the parameters of a augmentation, such as the rotation degrees or scale of cropping.
- Study why certain convolutional neural network give more improvements than others. What underlying features have the most influence on the reaction on data augmentations?

## References

[1] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. pages 61–68, 03 2009

[2] Vesna Novak and Robert Riener. Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection. IEEE ... International Conference on Rehabilitation Robotics : [proceedings], 2013:1–6, 06 2013.

[3] Alessandro Vinciarelli, Maja Pantic, Herve Bourlard, and Alex Pentland. Social signal processing: State-of-the-art and future perspectives of an emerging domain. MM'08 - Proceedings of the 2008 ACM International Conference on Multimedia, with co-located Symposium and Workshops, 10 2008.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems, volume 25. Curran Associates, Inc., 2012

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CoRR, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385