Where and Why Infants Look:
A recurrent neural network for the development of visual attention

Yukie Nagai (NICT, Bielefeld University) and Niyati Rawal (Osaka University)

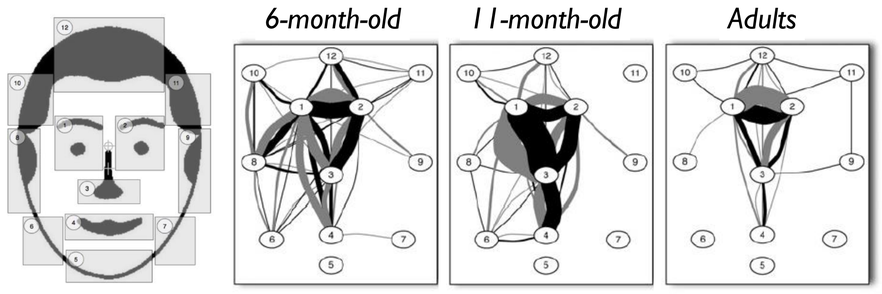
Development of Visual Attention

- Random scan paths in young infants during the observation of a person's face become simpler and more similar with age [Kato & Konishi, 2013].

Our Hypotheses about Underlying Mechanism

(a) The visual information obtained from the previous time steps needs to be integrated in order to predict the subsequent attention location.

(b) Attention shift among the eyes and mouth is driven by both featural processing and configural processing of input images.

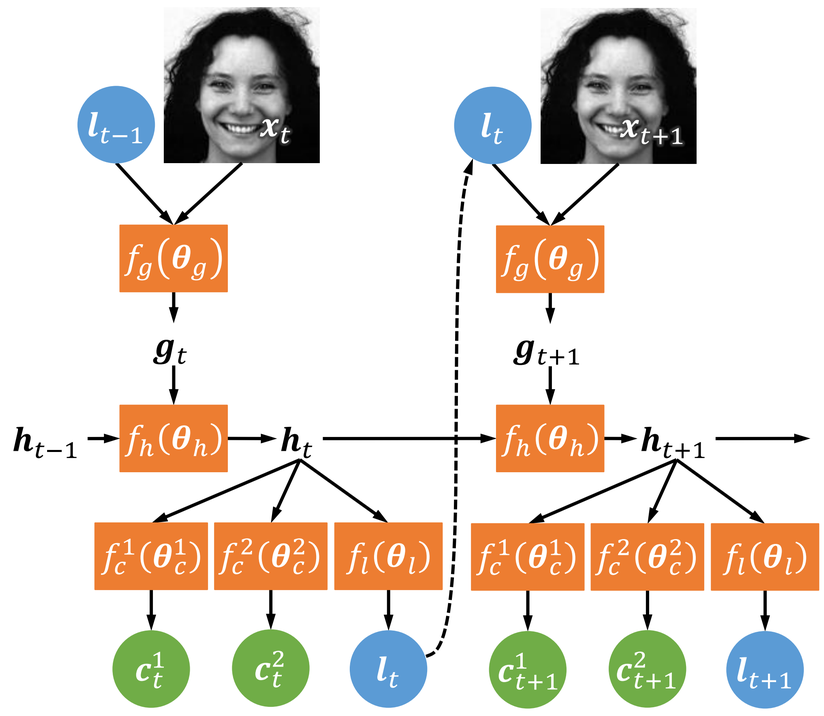


[Kato & Konishi, 2013]

A Computational Model for Visual Attention

Key idea: A recurrent neural network (hypo (a)) learns to optimize attention locations through learning of featural and configural classification tasks (hypo (b)) of input images.

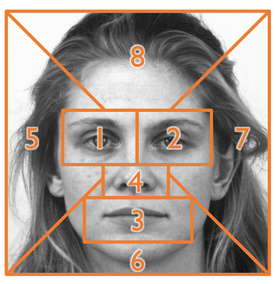
- **Input:** a face image $\boldsymbol{x}_t$, the attention location at previous time step $\boldsymbol{l}_{t-1}$
- **Output:** the next attention location $\boldsymbol{l}_t$, classifications $\boldsymbol{c}_t^1$ for a *featural* task and $\boldsymbol{c}_t^2$ for a *configural* task
- **Four network modules:**
  - Glimpse network $f_g(\boldsymbol{\theta}_g)$: To extract a glimpse representation $\boldsymbol{g}_t$ corresponding to the fovea
  - Internal network $f_h(\boldsymbol{\theta}_h)$: To integrate $\boldsymbol{g}_t$ with the internal representation $\boldsymbol{h}_{t-1}$ with a recurrent connection
  - Classification networks $f_c^1(\boldsymbol{\theta}_c^1)$ and $f_c^2(\boldsymbol{\theta}_c^2)$: To estimate the categories c_t^1 and c_t^2 of $\boldsymbol{x}_t$
  - Location network $f_l(\boldsymbol{\theta}_l)$: To determine the next attention location $\boldsymbol{l}_t$
- **Training through reinforcement learning:**
  - Reward: 1 for a correct classification, 0 otherwise



(Modified from [Mnih et al., 2014])

Experimental Setting

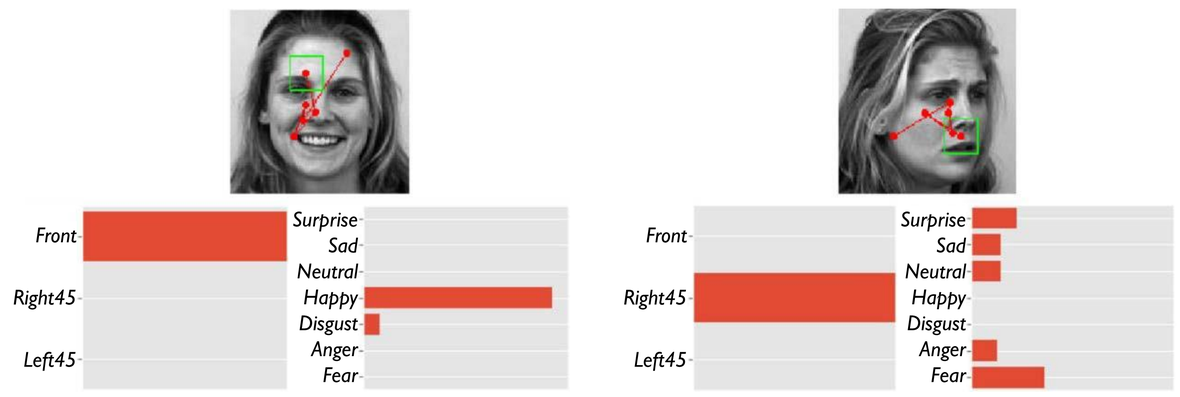
- **Two classification tasks:**
  - Featural task: *Emotion* estimation
  - Configural task: Estimation of *head orientation*
- **Training conditions:**
  - KDEF dataset (600 pictures with seven types of emotion and three head orientations) [Lundqvist et al., 1998]
  - Image size: 128 x 128 [pixels]
  - Glimpse size: 26 x 26 [pixels]
  - Max. number of attention shift: 6
  - Learning period: 1,000 epochs



Eight regions used for analysis

Exp. 1: Accuracy of Featural and Configural Tasks

- Emotion estimation (i.e., featural task): 39%
- Estimation of head orientation (i.e., configural task): 97%

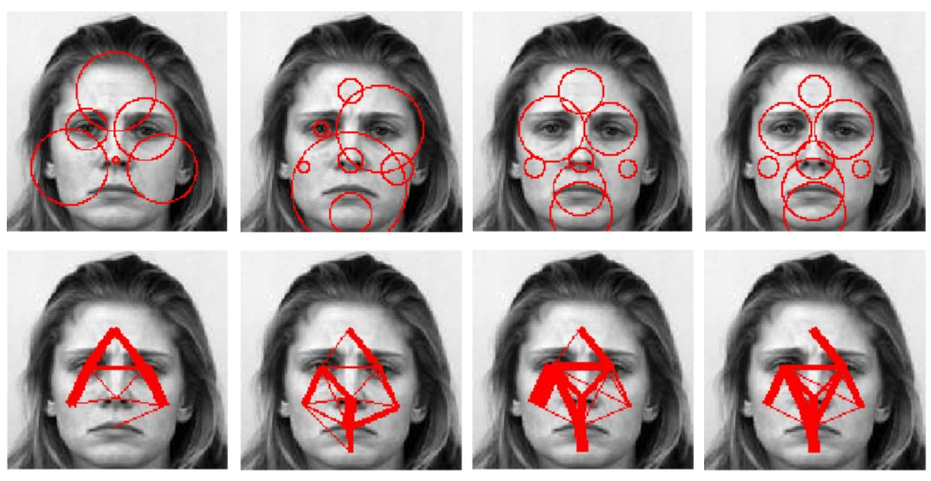


Exp. 2: Development of Visual Attention

- The neural network reproduced the developmental change similar to that observed in infants.

Glimpse locations
(diameter of circles = frequency of attention)

Scan paths
(thickness of lines = frequency of attention shift)



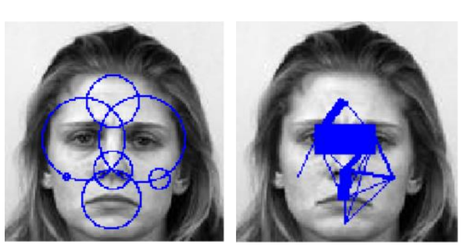
0 epoch *200 epoch* *500 epochs* *1000 epochs*

Exp. 3: Complementary Roles of Featural and Configural Tasks

- The emotion estimation induced closer attention to salient facial features (i.e., *featural* processing) .

Learned only for emotion

- The estimation of head orientation required attention shift among facial regions (i.e., *configural* processing).



Learned only for head orientation

Conclusion

- The recurrent neural network reproduced infant-like development of visual attention.
- Featural and configural processing played complementary roles in the development of visual attention.

Future Issue

- Integration of bottom-up saliency-based attention

References

- M. Kato and Y. Konishi, Infant Behavior and Development, 36(1): 32–41, 2013.
- S. Carey and R. Diamond, Science, 195(4275): 312–314, 1977.
- J. Sergent, British Journal of Psychology, 75(2): 221–242, 1984.
- V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, in Advances in Neural Information Processing Systems 27: 2204-2212, 2014,
- D. Lundqvist, A. Flykt, and A. Ohman, http://kdef.se/home/aboutKDEF.html, 1998.

This work was supported by JST CREST "Cognitive Mirroring" (Grant Number: JPMJCR16E2), Japan.