社団法人 人工知能学会
Japanese Society for
Artificial Intelligence

人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B101-8 (5/4)

Analogies between the autistic brain and the state of the art artificial autonomous agent

Norbert Michael Mayer[1], Horng Wu[1], Ian Fasel[2], and Minoru Asada[3]

[1]National Chung Cheng University, Min-Hsiung, Chia-Yi, Taiwan
[2]University of Arizona, Tucson, USA,[3]Osaka University and JST ERATO Asada Project, Osaka, Japan

1 abstract

The basic idea of this work is to find a analogies between clinical phenomena as Asperger Syndrome and Autism Spectrum Disorders (ASD) and problems in state of the art autonomous agents in dynamic environments. Autistic individuals show a variety of curious behavioural and perceptual patterns which generally seem to lack a common underlying cause. At another place [1] we propose a theory of learning and perception which suggests that the general impairment of the autistic brain may be a restriction on the class of connectivity patterns (or "features") that can be utilised for learning perceptual and cognitive tasks. In particular, we argue that the autistic brain does not make proper use of features which pool information over larger areas of the input space, which would allow them to make use of symmetries and develop invariants to permutations. In this work we briefly introduce the basic concept of our Autism theory and then go to cognitive problems in autonomous agents; we test the hypothesis that these in both cases similar causes result in certain impairments both seen in artificial autonomous agents and autistic individuals.

2 Introduction

The CDC list of diagnostic criteria [2] lists a total of 8 possible symptoms of ASD with regard to social and communication impairments, as lack of speech, 'a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people', and others.

However, psychological signs and symptoms exceed problems of social interaction and communication. A wide variety of symptoms can be seen in children with ASD that are not directly related to those, but express some very particular relation to objects and order in the real world.

Here the CDC list more possible symptoms that indicate ASD: *Restricted repetitive and stereotyped patterns of behaviour, interests, and activities, as manifested by at least one of the following:*

- *encompassing preoccupation with one or more stereotyped and restricted patterns of interest that*

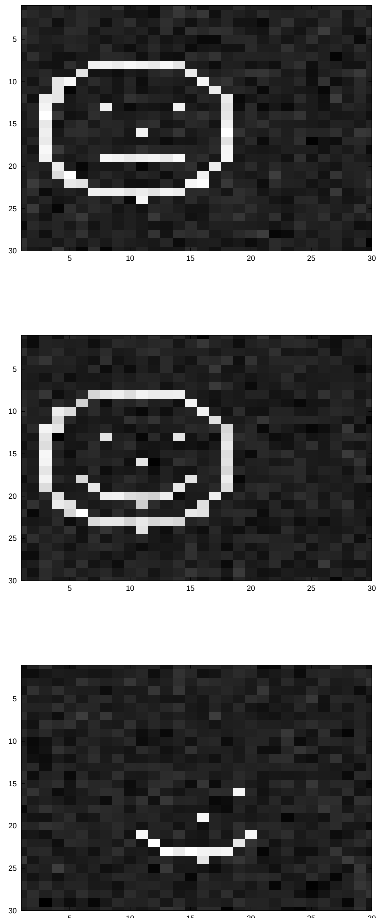

Figure 1: Different stimulus types in a toy world. The agent should be trained to distinguish the smile from the non-smile face. The faces can appear at any position in the input space. Left and middle: Stimuli used for training, the non-smiling and smiling simley have to be discriminated by the network. Right side: Example for incomplete simleys. In one additional task, the trained network performances are tested by these incomplete smileys.

is abnormal either in intensity or focus

- *apparently inflexible adherence to specific, non-functional routines or rituals*
- *stereotyped and repetitive motor manners (e.g., hand or finger flapping or twisting, or complex whole-body movements)*
- *persistent preoccupation with parts of objects*

Causes of autism are not well understood although for many years theories exist. One standard approach sees autism as a combination of both behavioural and perceptive problems caused by an overdose of testosterone during some critical developmental period [3]. In that sense autism is seen as an extremely male brain, which means pathological low levels of empathy. For the treatment the main issue is that autistic individuals lack to some degree a theory of mind, and their social behaviour is impaired.

Ramachandran [4] popularised the fact that one very essential defect lays in the mirror neuron system. This pins down the psychological problems to differences between non-autistic and autistic individuals on a physiological level. Problems in the mirror neuron system have been shown by EEG [5] experiments.

Since one can not record the auditory and visual sensations inside a human, it is hard to distinguish perceptive impairments that result in unusual behaviours, and unusual behaviours while perception works impaired. Already in the 1980s it had been suggested that the behavioural problems are by a "weak central coherence" (WCC) in perception (for a review confer [6]).

Recently autism has become considered an interesting subject for research in artificial intelligence and developmental robotics. Asada et al. [7] introduce a new integrated framework to understand the sequential development of sensory, sensory-motor, motor skills and –finally– social behaviour. The hot spot of development is also located anatomically as a process that starts in the occipital lope and then moves frontal via the parietal cortex, finally reaching the frontal lope. In this work, among other things, Asada et al. propose that diseases as ASD on one hand and Williams syndrome (extreme socially and empathic behaviour, [8]) on the other hand can be modelled as failures of the last stage of this type of sequential developmental model.

We follow the WCC idea of as general theory of perception that comprises but is not restricted to social interaction or imitation and the mirror neuron system. Thus, we focus on potential sensory deficits rather than an initial behavioural impairment and we see the pathological behaviour of ASD patients as a secondary effect of the underlying sensory problems. Different from the initial WCC theory we focus here particular on impairments with regard to detect invariants.

In the context of this workshop we would like to discuss similar problems that appear in intelligent autonomous agents. It is almost trivial common knowledge that tasks that appear simple to human intelligence, as common object recognition are difficult to achieve in automated recognition systems.

Autonomous agents tent to have similar problems as autistic individuals in real world environments. Weaknesses but also strengths of autistic individuals seem to be some extend surprisingly congruent to artificial intelligent machines. While it is no problem to reproduce photo-realistic impressions it is much harder to detect objects on a table, to grab things, to imitate. We want to emphasise the role of the detection of invariants. Basically, the phenomenon is well known to the community, in some sense this also has let to the foundation of the RoboCup [9], which suggests soccer as a more realistic benchmark, because demands in soccer a similar to real world environments. The purpose of this work is to find analogies between our autonomous agents and the autism and so find a new perspective to a set of well known problems in our community.

As mentioned above we focus here on face recognition. Many computer vision algorithms make use of carefully crafted translational invariant features, or instead use weight-sharing convolutional or bag of words models in order to enforce strong translation invariance constraints. David Lowe [10], and more recently Serre, Wolf, and Poggio [11], have drawn on neuroscience evidence to back up the claims from their models that various kinds of local pooling which result in translational, scale, and rotational invariants are a common feature throughout primate visual cortex.

We speculate that, if these translational invariance mechanisms are taken away, then an autonomous system will start to make mistakes similar to autistic mistakes. On the other hand we see potentially great advantages if the detection of invariants can be further improved, in fact this could potentially close the gap between artificial and natural intelligent agents.

In detail, real world invariants are a hard to detect, because they are covered behind of several stages of processing. We focus here on the relative simple lateral invariants. These are probably the simplest examples. The problem is essence is identical to dimensionality reduction in machine learning: The knowledge of invariants is the tool to reduce the search space for imagination and planning tasks, it is such a very essential problem in real world intelligent agents, which has been addressed at many occasions [12].

In some sense mirror neurons and imitation also fit into the context of invariants. The knowledge of which actions of others are equivalent to one's own actions (that is an invariance) is necessary to do so and thus deal with a much more complicated invariants of perception. In consequence, body parts of the other persons have to identified with own body parts. Nonetheless, the detection of these invariants seems to work to some extend at the time of birth [13].

As an example that we relate to our toy model are experiments [14], that show that in spite of having a lower over all recognition rate for faces autistic individuals tent to have relative high recognition rates from parts of faces in comparison to non-autistic persons.

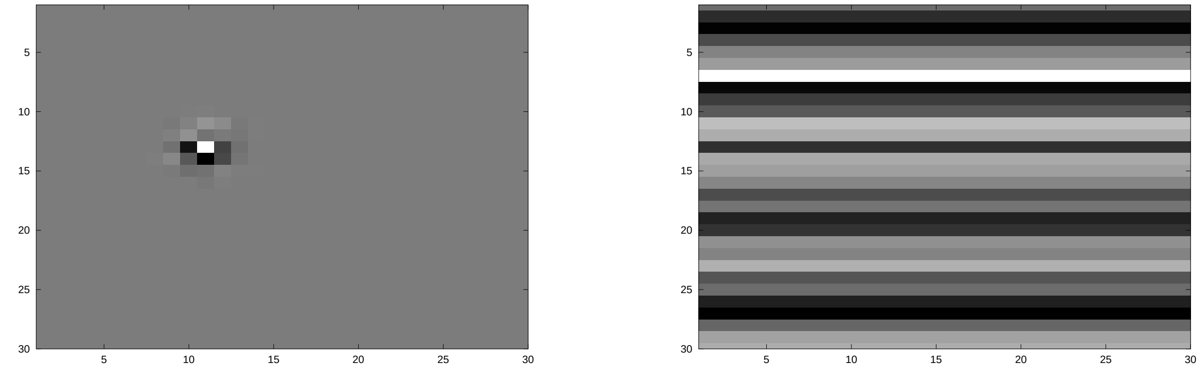

Figure 2: Receptive fields for the model for the autistic (left) and non-autistic individual (right).

3 Feature Invariants

An important part of the "art" of pattern recognition research is the construction of features that are invariant to the aspects of the world that are unimportant for solving the task at hand. An early example of this approach in computer vision was the idea of first extracting edges, and then discarding the rest of the information on the assumption that smooth gradients, color, or the overall intensity of light on either side of an edge, are generally irrelevant for the task of recognising shape. Research involving edge or local gradient type features has since since evolved to the use of collections of edges at multiple scales, for instance Gabor jets [15], or the SIFT features of Lowe [10] which consist of histograms of gradient orientations in a grid of cells in image patches. Provided powerful enough features, problems of interest then become quite trivial in the transformed feature space, leading one leading computer vision researcher to argue that the single most important thing for future computer vision research will be the development of better features [16]. Also in the field of robotics invariants have often been applied (see for example [17]).

4 Toy model

The toy model illustrates advantages of a build in lateral symmetry constraint. We want to compare in this model that computational costs in the case when the symmetry constraint is implemented (non-autistic) and when the symmetry constraint is implemented (autistic). As a measurement for the computational cost serves the number of neurons that are necessary to fulfil a discrimination task.

The basic model consists of a two layer neuronal model. The interpretation of the 'neurons' would not be real biological neurons but rather functional entities that serve our task. In the first layer we have set of model neurons with fixed random receptive fields. Each neuron i has the activation function:

$$A_{i,t} = tanh(I_t \cdot R_i), \quad (1)$$

where I_t represents the stimulus, t is the number of the trial, that is either a smiling or non-smiling smiley (see Fig. 1) at a random position in the visual field. In our model the receptive field has the size 30×30 pixel. The receptive field of each neuron R_i is individual for each neuron and chosen random with the following constraints:

- Unspecific receptive field type: Each pixel has a different random value equally distributed in the range -0.5 to 0.5. In the following we call this model autistic. Each pixel is multiplied with a 2 dimensional Gaussian distribution with one pixel standard deviation and a random position in the visual field.
- Lateral specific receptive fields: All pixel within one line have the same value in the range -0.5 to 0.5 (see Fig. 2). In the following non-autistic receptive fields.

The outputs of neurons in the first layer form a subspace of Hilbert space on I. The task is now to conduct a supervised discrimination task between smiling and non-smiling smileys independent from their position in visual field. The teaching signal s_t is -1 for a smiling smiley and 1 for a non-smiling smiley. The task is now to learn this identification task. The easiest to achieve this is to do a linear regression of the output signals of the neurons in order to separate the 2 outputs. The method is in this way similar to the principle in kernel methods and support vector machines. Technically the problem is solved by testing all possible positions of the smiley within the visual field. We record for each trial the vector $A_{i,t}$ and the training signal s_t. By calculating the pseudo-inverse of A we can get the connectivity with mean square error for solving task with

$$W = s \times pinv(A). \quad (2)$$

And the quality of the processing is tested by measuring the

$$corr =< s_{test} sgn(W \cdot A_{test}) >, \quad (3)$$

where s_{test} is the type of the stimulus that has been presented during the test phase, whereas A_{test} is the response of the neurons to the particular stimulus. In the test set the stimulus is chosen from a random position in both the x and y direction. Obviously, increasing the numbers of neurons in the first layer increases the

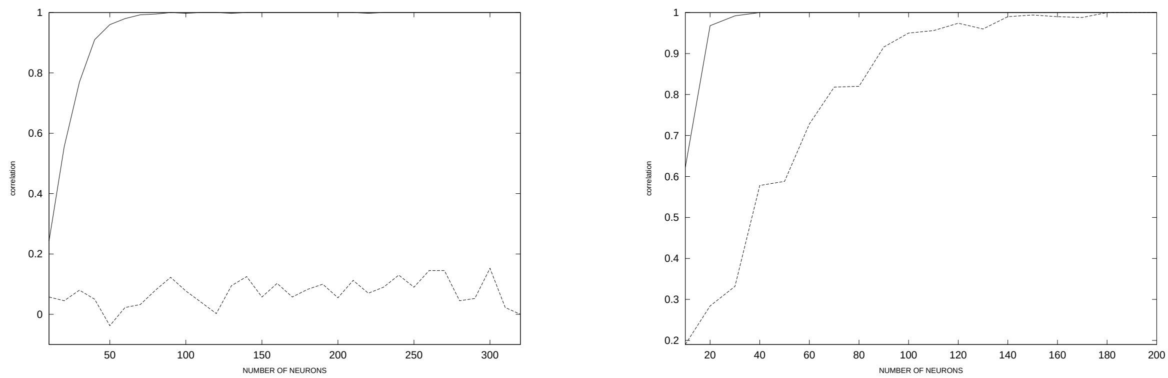

Figure 3: Results of simulation 1 and 2: Depicted are results of simulations with different numbers of neurons: The y-axes of the graphs depict the square error, x-axes the number of neurons in the first layer of network. Left side: One can see simulation 1, the network was trained and tested for the whole visual field. Right side: Simulation 2, the network was trained only for stimuli where the position of the smiley was varied only by 5 pixels in the horizontal and vertical direction the test. In both simulations the full line depicts the non-autistic version of the model, the dashed version shows the model with the autistic receptive fields.

number of dimensions in the Hilbert space and thus, increases in that way the accuracy of the response, i.e. the probability that the neural networks gives the correct response.

The only difference between the non-autistic approach and the autistic approach is the shape of receptive fields R_i. The non-autistic model is not able to detect lateral shifts of the stimulus in the visual field. This could be interpreted as a handicap of the non-autistic individual, since she/he is not able to perceive certain aspects of the stimulus. However, this build in handicap helps as we see later in the result part, to reduce the computational cost (i.e. the necessary number of neurons) to process the information in a proper way. Below one can see that the impairment pays off due to the lateral vertical invariance of the receptive fields. In order to get a better statistics, the simulations are repeated several times, the average is depicted in the figures.

This concept of "weight sharing" has often been used, for example in Yann LeCun's convolutional neural networks [18]. Technically even more useful are the connectionist style neural-network based face detector, from Rowley, Beluja, and Kanade [19], found it was beneficial to have multiple receptive fields that pooled information from different regions, such as patches of different sizes and shapes which also include horizontal bar receptive fields. Another recent approach by Serre, Wolf and Poggio's [11], and in the recent NIPS and ICML conferences Lee et al (with Andrew Ng) have introduced convolutional deep belief networks [20, 21]. Finally, many researchers in vision use *bag of words* models in which various features in an image are extracted and then pooled together into a single summary statistic, such as a histogram. This results in considerable translation and some scale invariance. Butko and Movellan (2009 CVPR) even built a convolutional POMDP controller for modeling robot eye movements, as did Fasel, Ruvolo, Wu, and Movellan [22].

5 Results of the toy model

In the following three numerical experiments have been conducted:

Simulations 1 The value of E depends on the different shape of receptive fields and the number of neurons in the first layer that has been used Fig. 3) depicts results from the following simulation: The training was conducted by teaching both the laughing and non-laughing test smiley at all possible positions in the visual field (14×14 different positions). The resulting network was tested with the 100 test samples that use either the laughing and non-laughing smiley with equal probability. The response error E was sampled. In order to smooth out the statistics of the result the numerical experiment was repeated 100 times with different, randomly chosen autistic and non-autistic receptive fields.

We see for all tested number of neurons a better performance of the non-autistic model than of the autistic model (see Fig. 3 left). . In both cases the performance of the network improves with the number of neurons. Already 50 neurons are sufficient to almost perfectly fulfil the task in the case of the non-autistic, horizontally homogeneous receptive fields.

In the autistic model more than 200 neurons are not sufficient to achieve a similar performance as the non-autistic model in the case of only 50 neurons. Thus, the blindness of the non-autistic case towards horizontal differences reduces significantly the computational costs and makes the network invariant against positional shifts in the horizontal direction.

Simulation 2: The next numerical experiment was that only a small part of the visual field (only 5×5 different positions instead of 14×14). This means the hor-

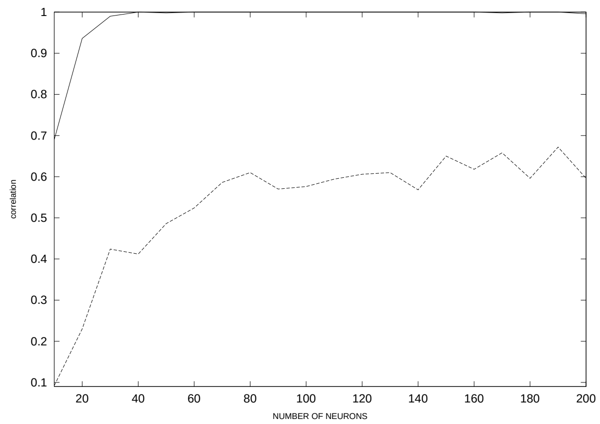

Figure 4: The figure depicts a simulation in which during the training the stimulus is only presented in a smaller range in the horizontal direction. In the test phase the network has to discriminate laughing and non-laughing stimuli in the whole area. As expected different from the autistic version of the model; due to the receptive field shape the non-autistic version can discriminate stimuli even if they are presented in the untrained range (Axes labels, line definitions same as in Fig. 3).

izontal positions of the smiley are only varied 5 pixels during learning. This has been done to test if the autistic version of the model is able to learn the discrimination at all. The results show that the although the autistic version of the model is able to learn the discrimination task, a much higher number of neurons is required to fulfil the task.

Simulation 3: In simulation 3 both networks are trained only using smileys that vary in the area 3×5; the test is then conducted in the area 5×5. Due to the symmetry hints the model related to the non-autistic individual can recognise stimuli even if they appear at positions which have never been seen before. This is reflected at the results (cf. 4 right). The non-autistic model achieves a performance of about the same level as the if whole area would be trained. The reason is that the non-autistic model is not able to distinguish the stimulus with respect to its horizontal position. Thus, a two different stimuli with the same vertical position and a different horizontal position result in the same response in the first layer. Thus, the response of the second layer is the same too.

Simulation 4: Finally results from simulation 1 were tested with incomplete stimuli (Fig. 1 right). The results show that in this case the autistic model shows a better performance than the non-autistic model. Although both stimuli and information processing in the model in much simpler than the according processes in the brain, we see here parallels to findings that autistic children show a better performance in recognising faces from parts of faces than non-autistic children [14].

6 Discussion

In conclusion, we look at ASD as the result of a broken version of a highly advanced and complicated cognition machine, that makes use of invariants in our real world environment. One important point is to look at possible genetic causes of autism and Asperger Syndrome. It is known that several genetic mutations can cause autism [23] in comparison to for example Williams Syndrome, where the syndrome is linked to a unique mutation on a single chromosome [8]. This could be a hint that autism is a phenomenon that is caused by the defect of highly advanced and complicated cognition machine, rather than just a wrong tuned parameter in a behavioural system.

Instead of wondering about the deficits in autistic individuals, designers of autonomous intelligent agents probably wonder more, why for non-autistic individuals it seems so easy to detect objects, recognise faces, find invariants in temporal sequences of task executions, because the very same tasks are relatively difficult to implement into robots.

- **Ability draw sketches:** As outlined before some autistic children show the ability to draw near photo realistic, at least naturalist stile, sketches of animals [24]. The explanation of here would be that in the autistic brain the representation of parts of a scene are not filtered away. During the drawing process it is necessary to reproduce details. In order to reproduce these details it is necessary to have an individual representation of each of these details. In the non-autistic version of the model we can see that these representations are reduced in favour of a invariance. We see here a potential analogy to the situation in the real non-autistic and autistic brain.

- **Perfect reproduction of the sounds:** In addition, there are reports that autistic children voice authentic reproductions of stopping trains, and others. We see here an analogy to the above mentioned ability drawing of sketches; the same arguments apply.

- **Interest in series:** Very well known is the interest of autistic children in temporal and spatial series of events and objects. Thus, children make perfect lines of objects. For example they are interested in every wheel of a toy locomotive. This can be also interpreted as a result from non-functioning filters that let pop up spots of interest that would be suppressed otherwise.

- **Increased ability to recognise the face from parts of the face:** We related the abilities [14] to a broken filter of translational invariance in our model. But: In our model we show only one way in which this ability can relate to in some way damaged filters.

- **Sticking to the temporal series in proceedures:** Autistic children tend to stick to temporal series of proceedures, as for example dressing. Thus, some insist to dress on the socks always before the trousers, etc.. In the context of the model is could be interpreted in that way, that the invariant of this temporal sequence cannot be detected, or for some reason it is not understood as important.

- **Impairments in imitation:** To imitate an action it is required to understand in what way the action has to be done invariant from the person, who is undertaking the action.

The experiment 3 shows also that the autistic model is not able extrapolate along the direction of the invariance. One example where this is important could be situations where certain experiences are linked to touching, and from there a relation to a person can be established. Since the person is experienced differently from far, the autistic individual cannot learn the emotional relation independent from the sight. This could affect for example bimodal receptive field (as presented in[25]).

Finally, we see similar issues in our intelligent autonomous agents. Although, in recent years very good software on face, object recognition, speech recognition have been developed we see several aspects that bring together the autonomous agent issues and models related to ASD. First of all see a real autonomous agent as a perfect model to verify WCC theory, and the concept of impairments of the recognition of invariants. In addition, we think these kind of models can improve our understanding of the recognition of invariants and to develop a general theory there.

Acknowledgements

We thank M. Ogino for helpful hints and discussions. The work has been supported by National Science Council grant number 98-2218-E-194-003 -MY2.

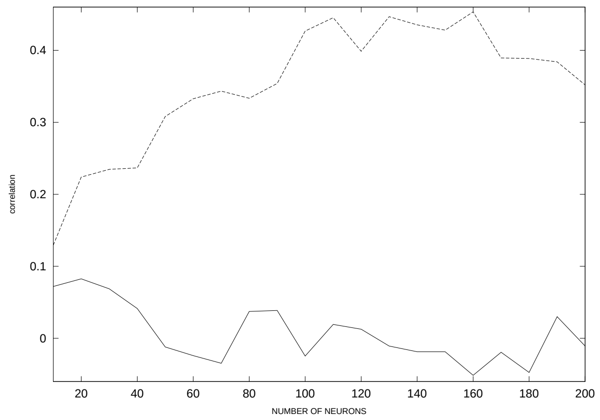

Figure 5: Results of the simulation 4: Here the trained network of simulation 2 is tested with incomplete stimuli. In this task we see better results in the autistic version than in the non-autistic model (Axes labels, line definitions same as in Fig. 3).

References

[1] NM Mayer and I Fasel. Title TBA. In *International Conference on Develo1978pment and Learning (ICDL)*, 2010. in preparation.

[2] CDC. Diagnostic criteria autism spectrum diorders. available at http://cdc.gov/ncbddd/autism/hcp-dsm.html.

[3] Simon Baron-Cohen. The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6:248–254, 2002.

[4] S. Vilayanur Ramachandran and Lindsay M. Oberman. Broken mirrors: A theory of autism. *Scientific American Magazine*, November 2006.

[5] LM Oberman, EM Hubbard, JP McCleery, EL Altschuler, VS Ramachandran, and JA Pineda. EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res*, 24(2):190–198, 2005.

[6] Natasja van Lang. *Autism spectrum disorders: a study of symptom domains and weak central coherence.* Groningen University, 2003. Dissertation, available online at http://irs.ub.rug.nl/ppn/25607075X.

[7] Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: A survey. *IEEE Trans. on Autonomous Mental Development*, 1(1):12–34, May 2009.

[8] MA Martens, SJ Wilson, and DC Reutens. Research review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychologcal Psychiatry*, 49(6):576–608, 2008.

[9] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. RoboCup: A Challenge AI Problem. *AI Magazine*, 18(1):73–85, 1997.

[10] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] T Serre, L Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*. IEEE Computer Society Press, San Diego, 2005.

[12] The curse of dimensionality problem: How does the brain solve it? In *Workshop of NIPS Conference (Whistler)*, 2009. Simon Haykin, Terry Sejnowski, Salk and Steven Zucker (Organizers).

[13] PF Ferrari, E Visalberghi, A Paukner, L Fogassi, and A Ruggiero. Neonatal imitation in rhesus macaques. *PLoS Biology*, 4(9):e302, 2006.

[14] T. Langdell. Recognition of faces: An approach to the study of autism. *Journal of Child Psychology and Psychiatry*, 19:255–268, 1978.

[15] Martin Lades, Jan C. Vorbrueggen, Joachim Buhmann, Joerg Lange, Christoph V. D. Malsburg, Rolf P. Wuertz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42:300–311, 1993.

[16] Ian Fasel. Automatic discovery of object categories. In *NIPS Workshop*, 2005.

[17] Yuichiro Yoshikawa. *Subjective Robot Imitation by Finding Invariance.* 2005. PhD-thesis, Osaka University.

[18] Y LeCun. Generalization and network design strategies. In R Pfeifer, Z. Schreter, F Fogelman, and L Steels, editors, *Connectionism in Perspective*. Elsevier, 1989.

[19] HA Rowley, S Baluja, and T Kanade. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 20:23–38.

[20] Ian J. Goodfellow, Quoc V. Le, Andrew M. Saxe, Honglak Lee, and Andrew Y. Ng. Measuring invariances in deep networks. In *Advances in Neural Information Processing Systems (NIPS) 22*, 2009.

[21] H Lee, R Grosse, R Ranganath, and A Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *In Proceedings of the 26th Annual international Conference on Machine Learning (Montreal, Quebec, Canada, June 14 - 18, 2009)*, 2009.

[22] Paul Ruvolo, Ian Fasel, and Javier R. Movellan. A learning approach to hierarchical feature selection and aggregation for audio classification. *Pattern Recognition Letters*, 2010. (in press).

[23] BS Abrahams and DH Geschwind. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet*, 9(5):341–55, 2008.

[24] David Pariser. Nadia's drawings: Theorizing about an autistic child's phenomenal ability. *Studies in Art Education*, 22(2):20–31, 1981.

[25] *Constructing bimodal receptive field based on visual attention by saliency*, 2008.