# Rotation-Invariant Neoperceptron

Beat Fasel
BIWI, ETH Zurich
Zurich, Switzerland
bfasel@vision.ee.ethz.ch

Daniel Gatica-Perez
IDIAP Research Institute
Martigny, Switzerland
gatica@idiap.ch

## Abstract

*Approaches based on local features and descriptors are increasingly used for the task of object recognition due to their robustness with regard to occlusions and geometrical deformations of objects. In this paper we present a local feature based, rotation-invariant Neoperceptron. By extending the weight-sharing properties of convolutional neural networks to orientations, we obtain a neural network that is inherently robust to object rotations, while still being capable to learn optimally discriminant features from training data. The performance of the network is evaluated on a facial expression database and compared to a standard Neoperceptron as well as to the Scale Invariant Feature Transform (SIFT), a-state-of-the-art local descriptor. The results confirm the validity of our approach.*

## 1. Introduction

For object recognition, many different approaches that allow for translation, scale and rotation invariant recognition have been investigated in the literature [6, 11, 4, 2, 8, 7]. An early method was the Fourier transform that allows for translation invariance. Combined with the logarithmic transform, also rotation and scale invariance can be achieved [6]. Rotation invariant object recognition can also be done with steerable filters [9]. These filters constitute a multiscale oriented image transform and have been deployed for various analysis and synthesis applications. Rotation invariance in steerable filters can be obtained by computing a main orientation of local regions and by steer the neighboring filter responses. In [11], three different methods have been investigated to estimate orientations: (a) gradient orientation at a center pixel location, (b) peaks in the orientation histogram of the local region, and (c) orientation of the eigenvector of the second moment matrix of the local region. The latter two methods are used as part of the Scale Invariant Feature Transform (SIFT) [4], and its variants, e.g. [2]. These descriptors have been successfully used for object recognition, and are increasingly applied to other computer vision problems [1, 10]. In this paper, we present a neural-network based approach that exploits, in a different way, the concepts of local features and invariance. Our approach is an extended Neoperceptron, that belongs to the family of convolutional neural networks [3] - which provide an efficient method to constrain the complexity of feedforward neural networks by means of weight-sharing. Our network has the advantage of being able to be inherently invariant to object orientations. In the context of neural networks, the rotation-invariant Neocognitron [8] and the rotation invariant neural face detector presented in [7] are the closest related works to ours. The rotation-invariant Neocognitron deploys cell-plane stacks and blurring layers to gain rotation invariance. In contrast, the rotation invariant face detector uses two distinct neural networks, one that estimates the orientation of a face candidate in an input image so that it can be rotated into up-right position. A second network then judges whether it is a face or not. Our approach is different in that only one network is used and we do not learn orientations from training data. In comparison to the Neocognitron, our Neoperceptron is trained in a supervised way and we do not rotate learned weight kernels with a rotation matrix but instead rotate input images and share weight kernels over orientations. In the remainder of the paper, we introduce our approach in details (Section 2), present and discuss results on a multi-class task (Section 3), and conclude with an overall discussion and some final remarks (Section 4).

## 2. Rotation-Invariant Neoperceptron

The proposed rotation-invariant Neoperceptron is a convolutional neural network that allows for invariant object recognition from one training example. Transformation invariance is not learned from training images, but the network architecture is inherently robust to input variations. The network architecture of the proposed rotation-invariant Neoperceptron is shown in Figure 1. It consists of two stages. The first one transforms an input image into vir-

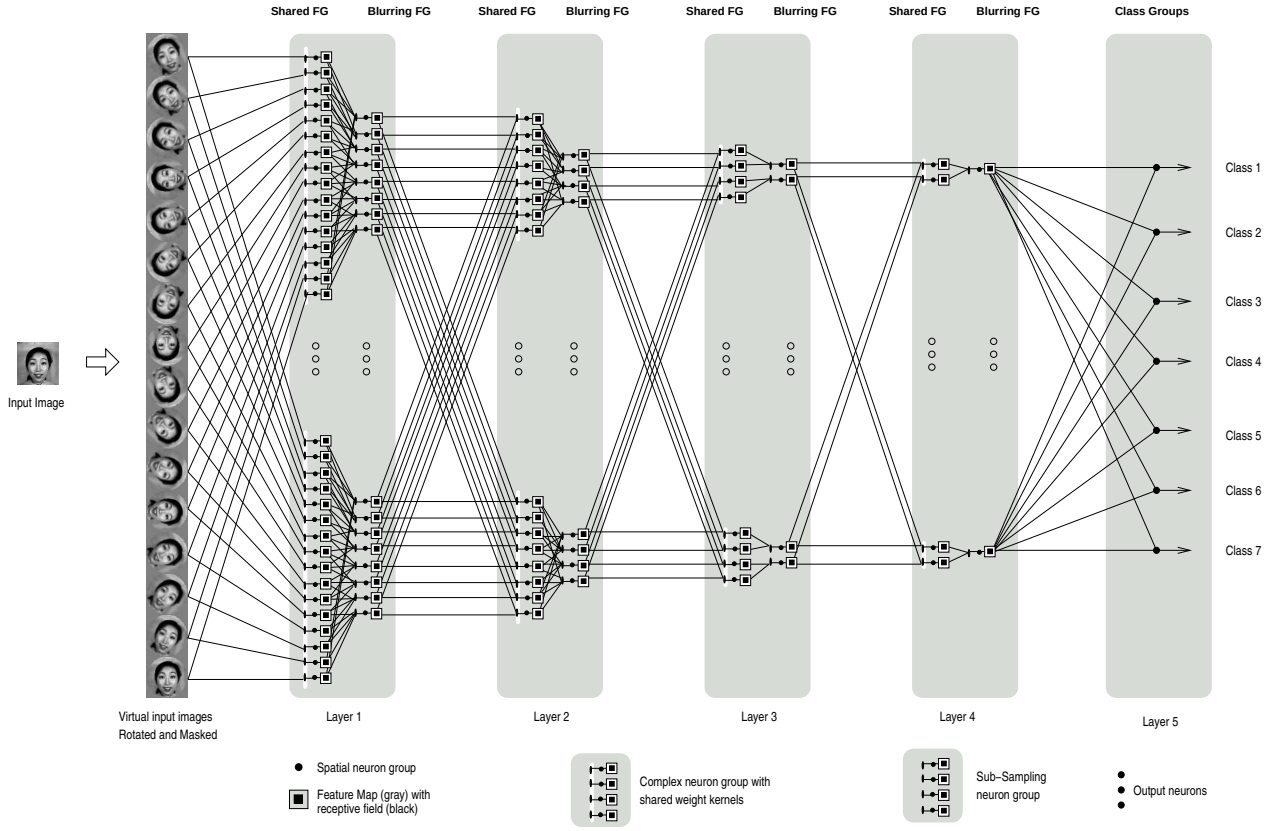


**Figure 1. Extended Neoperceptron Architecture: The network consists of two stages: virtual input image generator followed by a convolutional neural network with shared feature groups (Shared FG) and blurring feature group (Blurring FG) layers.**

tual images by generating rotated images. In the second stage, a convolutional neural network, consisting of weight-sharing feature groups organized together with blurring layers, selects features that are optimally discriminant for a given classification task with the training examples that are at disposition. How the network is able to gain rotation-invariance is shown in more detail in Figure 2. Three different feature groups are depicted. The simple weight-sharing feature group has only one weight kernel (w1) at disposition, and shares it over all rotations of the virtual input images. The blurring group reduces the number of virtual input image orientations that are seen by the network. This group uses a fixed Gaussian distribution of weights in order to blur and sub-sample the input (in the example, reducing the number of orientations from 8 to 4). Finally, a complex weight-sharing feature group has several weight kernels (w2 and w3) that are shared over 4 orientations. This feature group combines several simple features into a complex feature. Note that in Figure 2 we have left out the bias input into the neurons in order to simplify the illustration. The feature and blurring groups result in a massive weight-sharing. This has several advantages: (1) it reduces the capacity of the network and thus the network is less prone to overtraining, which in turn reduces its generalization performance. (2) Learned features are local and therefore less affected by occlusion. (3) Repeated local features extraction followed by blurring allows for object deformations produced by small affine transformations and viewpoint variations. The described Neoperceptron is not only robust to orientation changes, but it can also be made translation and scale-invariant. This can be achieved by creating a multiscale pyramid of the input image that is subsequently parsed by the network. Hereby, the Neoperceptron is replicated over space and scale. Similar to Multilayer Perceptrons (MLP), convolutional neural networks can be trained via the backpropagation algorithm. The risk of overtraining can be minimized by using validation sets and retaining the network weights at the training epoch with the smallest validation error. For the experiments in this paper we chose a rotation-invariant Neoperceptron with a network topology consisting of 6 simple weight-sharing feature groups in layer 1, 6 blurring neuron groups in layer 2, 12 complex

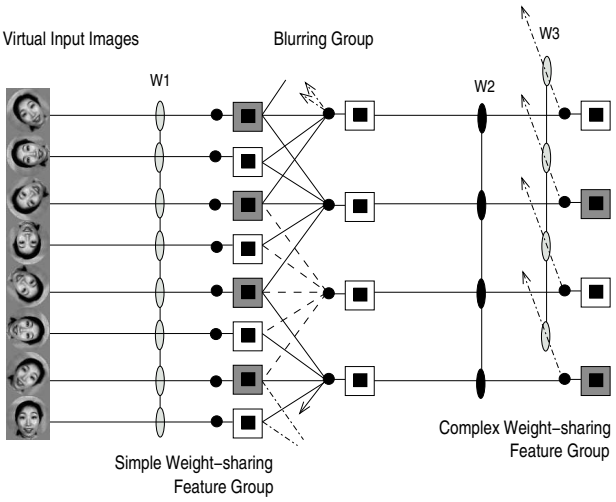


**Figure 2. Inherent Rotation invariance is gained by deploying weight-kernel sharing feature groups (both simple and complex) as well as orientation blurring layers.**

feature groups in layer 3 and 12 blurring neuron groups in layer 4.

## 3. Experiments and Results

We first describe the experimental setup before we demonstrate the rotation scale and translation invariance of the proposed neural network architecture. Furthermore, we compare the subject generalization performance of the proposed Extended Neoperceptron with the one of SIFT.

### 3.1. Experimental Setup

We demonstrate the performance of the proposed Extended Neoperceptron with facial expression images. For testing purposes, we employed the publicly available JAFFE database [5]. It consists of frontal face images of 10 Japanese female students that perform 6 posed emotional facial expressions (happiness, sadness, fear, anger, disgust, and surprise) as well as neutral displays. A total of 3x7=21 images per person were used for training, validation and testing, respectively. The employed grayscale images were originally of size $256 \times 256$ but reduced in scale to $64 \times 64$. Some sample images are shown in Figure 3.1. Note that the facial expressions were labeled categorically into 7 distinct classes. Before an image is processed by the described neural network, it is rotated (16 angles) by deploying bilinear interpolation. Hence, we have a feature detector at an angle of every 22.5 degrees. In order to be able to handle faces that appear at an angle in-between (e.g. 11.25 degrees)

**Figure 3. Sample Images of the JAFFE Facial Expression database.**

the learned features have to be invariant to small rotations - an assumption which we found to found to be reasonable. The deployed orientation blurring layers then interpolate the output of the detectors that are situated at fixed angles.

### 3.2. Rotation Invariance

The proposed extended Neoperceptron is to a high degree invariant to rotations. This is demonstrated in Figure 4. The figure shows the recognition accuracy for previously seen faces performing the six basic facial expression and neutral face displays. Depicted are the results for a standard Neoperceptron (Standard NP) [3], the rotation-invariant Neoperceptron (Rotinvar NP) as well as for SIFT. SIFT was deployed by matching all extracted keypoints of the training images against the ones of the test images (matches are identified by finding the two nearest neighbors of each keypoint). Clearly, both our approach as well as SIFT outperform the standard Neoperceptron when faces are rotated. Note that SIFT is more stable over angles and there is a peak for the rotation-invariant Neoperceptron at 45 degrees. We attribute both phenomena to the different virtual input image blurring occurring with the rotation of the input image and which depends on the rotation angle.

### 3.3. Subject Generalization Performance

An important issue is object categorization that allows for the generalization to objects that have not previously been seen by the system. We now compare the generalization performance obtained for the rotation-invariant Neoperceptron with the one obtained by SIFT. We use all images of one subject as test images (3x7=21) and the images of three other subjects as validation set (3x3x7=63). The reminding six subjects are used as a training set (6x3x7=126). The generalization performance for the 10 subjects is shown in Figure 5 for our approach and for SIFT. As can be seen, the recognition results vary considerably between subjects. This is due to the different quality of the posed expressions in the JAFFE database as well as the individual dif-

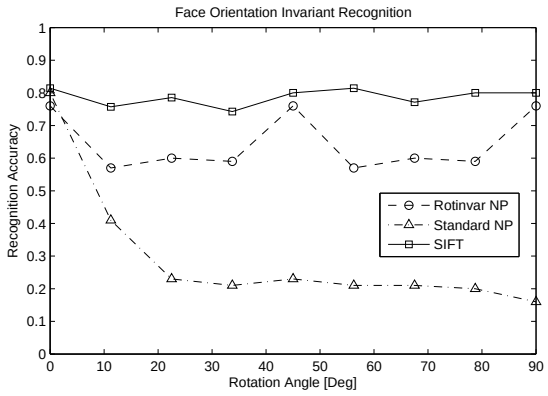


**Figure 4. Rotation Invariance. Shown is the recognition accuracy as a function of the rotation angle in the range of 0-90 degrees.**

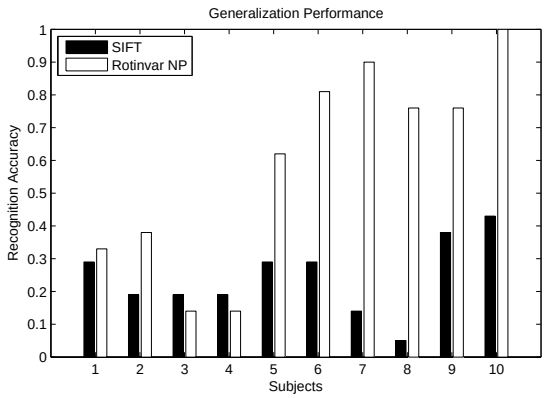


**Figure 5. Generalization Performance for Different Subjects. Shown are the results for the Extended Neoperceptron as well as raw SIFT.**

ferences in displaying facial expressions. The rotation-invariant Neoperceptron outperforms SIFT for most subjects (8 out of 10). Overall, the subject-based generalization performance for our method achieves 59%, while SIFT only achieves 24%. The poor performance of SIFT can be explained by the fact that, although the background of the images is relatively clean, some of the detected keypoints in the SIFT approach correspond to the background or to face physiognomy but in neither case contribute to the distinction of facial expressions.

## 4. Discussion and Concluding Remarks

In this paper we have proposed a novel convolutional neural network architecture that allows for invariance to object orientations without learning this property from training data. The rotation-invariant Neoperceptron has some properties in common with the steerable pyramid, especially the multiscale structure. In contrast to the latter, however, weight kernels are learned from data. Thus, the rotation-invariant Neoperceptron has the advantage to learn task-relevant discriminative filters that allow for a better generalization performance. Future work will include the comparison of the rotation-invariant Neoperceptron to a setup that clusters the variable number of SIFT descriptors into visual words that can be used to build object category models. It will be of interest to see whether the learned features of the Neoperceptron are more discriminant than the clustered SIFT features.

## References


[1] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Int. IEEE Conf. on Computer Vision And Pattern Recognition*, San Diego, June 2005.

[2] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. *Computer Vision and Pattern Recognition*, 2004.

[3] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995.

[4] D. G. Lowe. Distinctive image features from scale-invariant keypoints, cascade filtering approach. *Int. Journal of Computer Vision*, 60(2):91–110, January 2004.

[5] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 200–205, April 1998.

[6] E. Persoon and K. S. Fu. Shape discrimination using fourier descriptors. *IEEE Trans. Systems Man Cybernet*, 7:170–179, 1977.

[7] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[8] S. Satoh, J. Kuroiwa, H. Aso, and S. Miyake. Recognition of rotated patterns using neocognitron. In *ICONIP 1997*, volume Vol. 1, 1997.

[9] E. P. Simoncelli and H. Farid. Steerable wedge filters. In *Int. Conf. on Computer Vision (ICCV)*, Boston, MA, 1995.

[10] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering object categories in image collections. In *Int. IEEE Conf. on Computer Vision*, Beijing, China, October 2005.

[11] J. J. Yokono and T. Poggio. Rotation invariant object recognition from one training example. Technical Report AI Memo 2004-010, CBCL Memo 238, MIT Computer Science and Artificial Intelligence Laboratory, 2004.