Article:
Zhu, Xun, Watson, David M, Rogers, Daniel et al. (1 more author) (2025) View-symmetric representations of faces in human and artificial neural networks. Neuropsychologia. 109061. ISSN 0028-3932

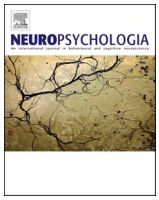
https://doi.org/10.1016/j.neuropsychologia.2024.109061

Neuropsychologia 207 (2025) 109061

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



View-symmetric representations of faces in human and artificial neural networks

Xun Zhu [1], David M. Watson [1], Daniel Rogers, Timothy J. Andrews [*]

Department of Psychology, University of York, YO10 4PF, UK

ARTICLE INFO

Keywords:
Face
Symmetry
Viewpoint
DCNN
fMRI

ABSTRACT

View symmetry has been suggested to be an important intermediate representation between view-specific and view-invariant representations of faces in the human brain. Here, we compared view-symmetry in humans and a deep convolutional neural network (DCNN) trained to recognise faces. First, we compared the output of the DCNN to head rotations in yaw (left-right), pitch (up-down) and roll (in-plane rotation). For yaw, an initial view-specific representation was evident in the convolutional layers, but a view-symmetric representation emerged in the fully-connected layers. Consistent with a role in the recognition of faces, we found that view-symmetric responses to yaw were greater for same identity compared to different identity faces. In contrast, we did not find a similar transition from view-specific to view-symmetric representations in the DCNN for either pitch or roll. These findings suggest that view-symmetry emerges when opposite rotations of the head lead to mirror images. Next, we compared the view-symmetric patterns of response to yaw in the DCNN with corresponding behavioural and neural responses in humans. We found that responses in the fully-connected layers of the DCNN correlated with judgements of perceptual similarity and with the responses of higher visual regions. These findings suggest that view-symmetric representations may be computationally efficient way to represent faces in humans and artificial neural networks for the recognition of identity.

1. Introduction

Recognising the identity of a familiar face is a simple and relatively effortless process for most human observers. However, the appearance of a face can change dramatically as a person moves their head. The visual system must take into account changes in the image that result from changes in viewpoint in order to recognise identity (Bruce and Young, 1986). Cognitive models of face perception suggest that this occurs through view-invariant representations (Bruce and Young, 1986; Burton et al., 1999; Young and Burton, 2017). A simple model for the emergence of view-invariant representations involves the convergence of multiple view-dependent representations (Serre et al., 2007; Rolls, 2012). However, the existence of view-symmetric representations for faces in behavioural and neural responses has led to the suggestion that this may be an important intermediate processing stage between view-specific and view-invariant representations (Freiwald and Tsao, 2010).

Behavioural support for the existence of view-symmetric representations in face processing comes from studies showing faces with symmetrical viewpoints (e.g. two profiles) are perceived to be more similar than non-symmetrical viewpoints (e.g. left profile and left ¾ view) and by face learning studies, which show a benefit in recognition when the test viewpoint is symmetrical to the learnt viewpoint (Troje and Bülthoff, 1998; Busey and Zaki, 2004; Favelle and Palmisano, 2018; Flack et al., 2019; Rogers and Andrews, 2022). Neurophysiological studies provide support for a functional hierarchy of facial viewpoint, with early stages of processing embodying view-specific representations, intermediate face regions showing more view-symmetric responses, and later face regions showing more view-invariance (Perrett et al., 1991; Freiwald and Tsao, 2010). Neuroimaging studies show a similar representational hierarchy with a transition from view-specific to view-symmetric representations (Axelrod and Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019; Rogers and Andrews, 2022).

The ability of artificial neural networks to recognise faces has improved significantly in recent years with the development of deep convolutional neural networks (DCNNs; Phillips et al., 2018; O'Toole et al., 2018; O'Toole and Castillo, 2021). However, the extent to which

* Corresponding author.
E-mail address: timothy.andrews@york.ac.uk (T.J. Andrews).
[1] Joint first authors.

https://doi.org/10.1016/j.neuropsychologia.2024.109061
Received 4 September 2024; Accepted 5 December 2024
Available online 5 December 2024

DCNNs provide good models of human face recognition remains a topic of debate (Cichy and Kaiser, 2019; O'Toole and Castillo, 2021). Although the fully-connected layers of the DCNN are able to decode face identity across image variation, they also contain information about the image, such as viewpoint (Parde et al., 2017; Hill et al., 2019; Abudarham et al., 2021). Recent studies using face recognition algorithms have found that view-symmetry is an emergent property for rotations along the yaw axis (Leibo et al., 2017; Yildirim et al., 2020; Farzmahdi et al., 2023). Because these algorithms have different architectures, this suggests that this reflects a general processing stage in artificial neural networks. Farzmahdi and colleagues (2024) found that view-symmetric responses to yaw emerged in the fully-connected layers of the DCNN and suggest that this reflects the pooling of mirrored representations. Although these studies provide some important insights into the emergence of mirror symmetry in artificial neural networks, there are a number of unanswered questions. For example, is view-symmetry evident for other rotations of the head (e.g. pitch and roll)? Does view-symmetry confer an advantage for the recognition of identity?

The aim of this study was to compare how viewpoint is represented in human and artificial neural networks. First, we measured the similarity between pairs of face images varying along 3 different axes of head rotation (yaw, pitch, and roll) using a DCNN trained to discriminate face identity (Parkhi et al., 2015). We asked if view-symmetrical patterns (i.e., greater similarity for symmetric compared to non-symmetric pairs) was evident in the output of the DCNN and whether this emerged in early (convolutional) or later (fully-connected) layers. Because of the symmetry of the face along the vertical axis, opposite head rotations in yaw (left-right) lead to mirror images (see Farzmahdi et al., 2023). However, because the face is not symmetrical along the horizontal axis, opposite head rotations in pitch (up-down) do not lead to mirror images. By comparing across different head rotations, we ask whether view-symmetry is only evident for rotations that lead to mirror symmetric representations. Second, we asked whether view-symmetry plays an important role in recognition. To do this, we compared the representation of same identity and different identity faces across different combinations of viewpoint. Our hypothesis was that view-symmetric representations in the DCNN should show a greater effect of identity compared to view-asymmetric representations. Finally, we asked directly whether humans and DCNNs compute view symmetry in similar ways. We compared behavioural responses (perceptual similarity ratings for pairs of images) with the response of different layers in the DCNN. We then compared the response to viewpoint at different layers of DCNN against neural responses measured using fMRI. Our aim was to establish any correspondence between processing in brain regions and different layers of the DCNN.

2. Methods

2.1. Stimuli

Face images were taken from two face image databases: Pointing '04 Image Database (Gourier, 2004) and Radboud Faces Database (Langner et al., 2010). The Pointing images were taken from 15 different subjects. There were two sets of images created from images taken at different times. Participants were instructed to change the angle of head rotation to create changes in yaw (0°, ±15°, ±30°, ±45°, ±60°, ±75°, ±90°) and pitch (0°, ±15°, ±30°). We also measured roll by rotating frontal view faces (0°, ±15°, ±30°, ±45°, ±60°, ±75°, +90°). Exemplars of Pointing '04 face images used in this study database are shown in Supplementary Figs. 1–3. Images from the Radboud database included 5 identities. Changes in yaw were achieved by the position of different cameras at set angles (0°, ±45° and ±90°). Exemplars from the Radboud face database used in this study are shown in Supplementary Fig. 4.

2.2. Deep convolutional neural network

We used the VGG-Face DCNN trained to discriminate facial identity (Parkhi et al., 2015). This network is trained and tested on its ability to detect identity on naturally varying face images including images that vary in pose or viewpoint. The training set for VGG-Face involved 2.6 million face images from 2622 people. The images are naturally varying and, although mostly frontal, contain variation in yaw and pitch. The training of the DCNN was augmented by horizontally flipping the images. We cropped our images to a square bounding box centered on the face and resized to 224 x 224 pixels for input into the DCNN. The DCNN consists of 13 convolutional layers and 3 fully connected (Fc) layers, which were used for the analysis. Each convolutional layer is followed by one or more non-linear layers, such as rectified linear units or max pooling, which were not used in this analysis. The dimensions of the layers are as follows: Conv1 = 224 x 224 x 64 = 3,211,264; Conv2 = 112 x 112 x 128 = 1,605,632; Conv3 = 56 x 56 x 256 = 802,816; Conv4 = 28 x 28 x 512 = 401,408; Conv5 = 14 x 14 x 512 = 100,352; Fc6 = 4096; Fc7 = 4096; Fc8 = 2622.

We measured the pairwise Fisher's z correlations between the feature vectors of all face images within each DCNN layer. Activations from a given layer of the DCNN were flattened into vectors and correlated between image pairs. For each layer of the DCNN, this gave rise to an image-by-image correlation matrix. These correlations were then averaged over same- and different-identity pairings separately for each viewpoint combination. For the Pointing '04 database, we calculated cross-validated similarity matrices by correlating images taken in the first session with images taken in the second session. No cross-validation was used for the Radboud database as only one repetition of each identity and viewpoint combination is available.

To determine whether the output of each layer of the DCNN reflected a view-specific or view-symmetric representation, we performed a series of representational similarity analyses (RSAs; Kriegeskorte et al., 2008) for the same-identity and different-identity pairings separately. A multiple regression analysis was performed for each DCNN layer, entering the corresponding DCNN correlations matrix as the outcome variable and the model view-specific and view-symmetric matrices as predictor variables. The view-specific model has decreasing similarity values with increasing angular difference. Maximum similarity (1.0) is predicted for identical viewpoints and minimum similarity (0.0) is predicted for the largest difference in viewpoint (e.g. −90° and 90° for yaw). All other similarity values in the model were a linear function of the difference in viewpoint. The view-symmetric model added view-symmetric values to the view-specific model by assigning a linear decrease in similarity according to the absolute viewpoint angle. Maximum similarity was assigned to pairs with identical absolute viewpoint angle (e.g., ±90° with ±90° yaw), and minimal similarity to pairs with a 90° difference between the absolute angles (i.e., 0° with ±90° yaw). All variables were converted to z-scores such that regression coefficients are reported in standardized units.

To determine the effect of identity, we compared the correlation matrices for the same-identity and different-identity faces for each layer of the DCNN. For each cell, the mean difference in correlation between same-identity and different-identity matrices was calculated. A larger value indicated a larger difference between same- and different identity faces, which in turn suggested a better emergence of identity information at this specific stage of computation within DCNN.

2.3. Behavioural task

To determine how behavioural judgements of similarity correspond with the response of different layers of the DCNN, we recruited 69 participants (male = 26, female = 43, median age = 21 years, age range = 18–26) for a perceptual similarity task. All participants had normal or corrected to normal vision and were drawn from an opportunity sample of students and staff at the University of York. All participants gave their

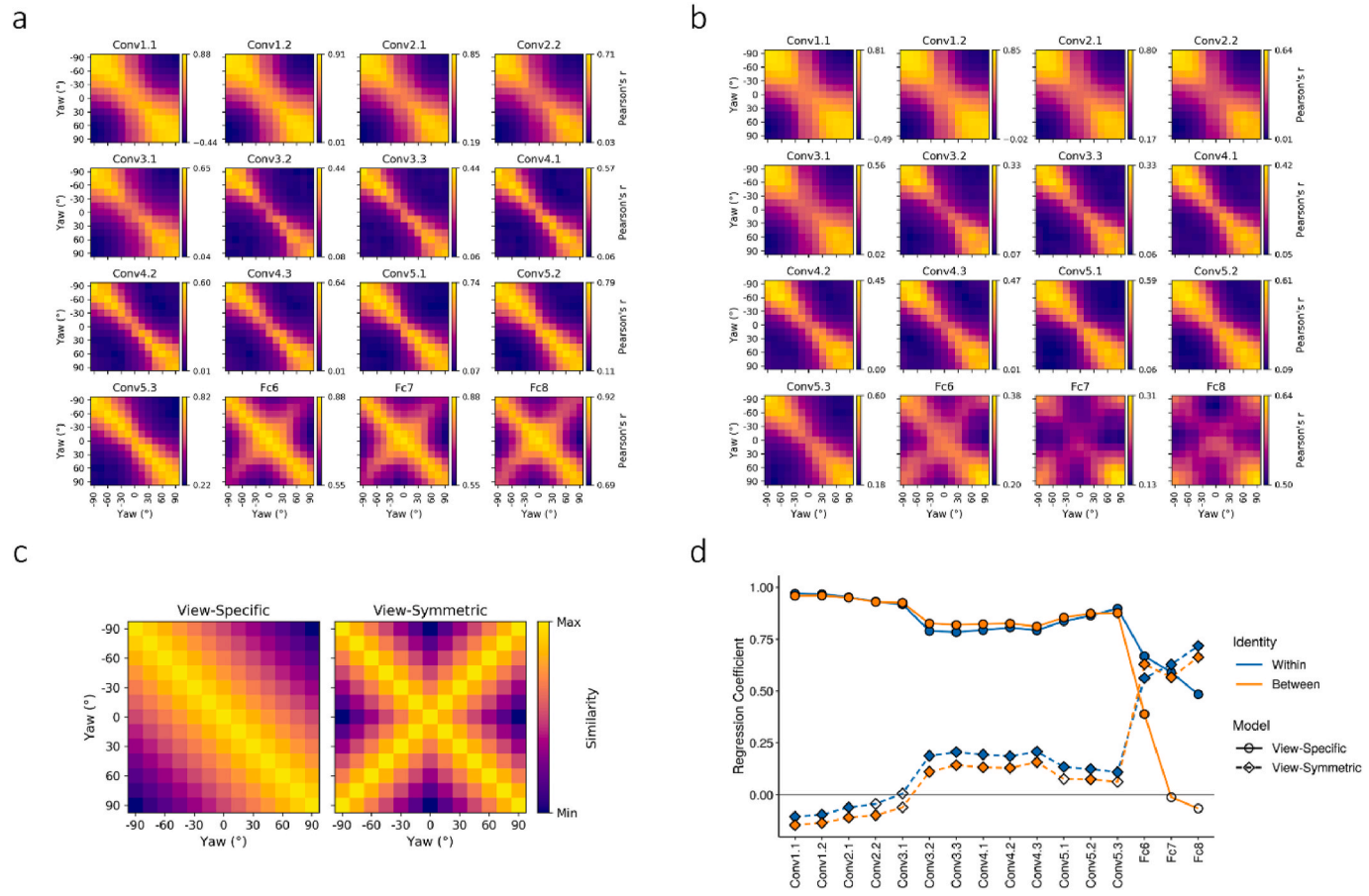


Fig. 1. Response of DCNN to images of faces that result from yaw rotations of the head. Correlation matrices showing the similarity to pairs of (a) within and (b) between identity comparisons at different viewpoint combinations. (c) View-specific and view-symmetric models were used in a regression analysis of the correlation matrices. (d) View-specific responses were evident in the convolutional (Conv) layers of the DCNN. However, view-symmetric responses were more evident in the fully connected (Fc) layers of the DCNN for both within and between identity faces. Filled symbols indicate coefficients showing a significant difference from zero.

written informed consent. The study was approved by the Ethics Committee of the Psychology Department as the University of York.

To determine whether human ratings of perceptual similarity could be used to predict the outputs of a deep neural network, participants made perceptual similarity judgements on the 5 unfamiliar identities from the Radboud dataset. Each was shown at 5 yaw rotations (−90°, −45°, 0°, 45°, 90°) giving rise to 25 unique images. There were a total of 300 unique image pair comparisons. Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version, 2021.1) (Peirce et al., 2019). Each trial began with a white fixation cross superimposed on a grey background for 0.5 s. This was followed by a pair of faces that were presented sequentially each for 2s. Images subtended approximately 8° of visual angle. The order of trials was randomised for each individual participant. Participants were required to respond with a button press indicating how similar they perceived the images to be, on a scale of 1–7 (1 being less similar and 7 being more similar). Participants were instructed to make this judgment as fast and as accurately as possible.

Participants were asked to judge the perceptual similarity of the images and were not instructed to focus on any aspect of the stimulus (identity or viewpoint). This was an important aspect of the design, as explicitly focussing on specific aspect of the stimulus (identity or viewpoint) could have led to participants making more metacognitive judgements. We performed further representational similarity analyses to assess whether human perceptual similarity ratings could be used to predict the outputs of a DCNN. We correlated the average perceptual similarity ratings or response time matrices from human participants with the corresponding correlation matrices from each layer of the DCNN.

2.4. fMRI task

To determine any correspondence between brain regions and different layers of the DCNN, we used a previously collected data set (Rogers and Andrews, 2022).

25 participants (female = 14, male = 11, median age = 21, age range = 18–47) were recruited for the fMRI experiment from an opportunity sample of students and staff at the University of York. All participants had normal or corrected to normal vision and gave their written informed consent. The study was approved by the York Neuroimaging Centre (YNiC) Research Ethics Committee.

The main fMRI experimental scans used a block design with 5 different stimulus conditions each depicting a different yaw rotation (−90°, −45°, 0°, 45°, 90°). Images from the different conditions taken from the Radboud data set were shown in a blocked design. In each block, 5 images from different identities were shown at the same viewpoint. Within each block, each image was presented for 1 s followed by a 200 ms grey screen. A 9 s fixation screen was presented between each block. There were 5 viewpoints and each was shown 6 times during the scan, giving a total of 30 blocks. The order of the blocks was pseudorandomised across the scan. Images subtended a visual angle of approximately 15° and were viewed on a screen at the rear of the scanner via a mirror placed immediately above the participant's head. Participants maintained attention during the scans by detecting occasional changes in the colour of a central fixation cross, responding via a button press.

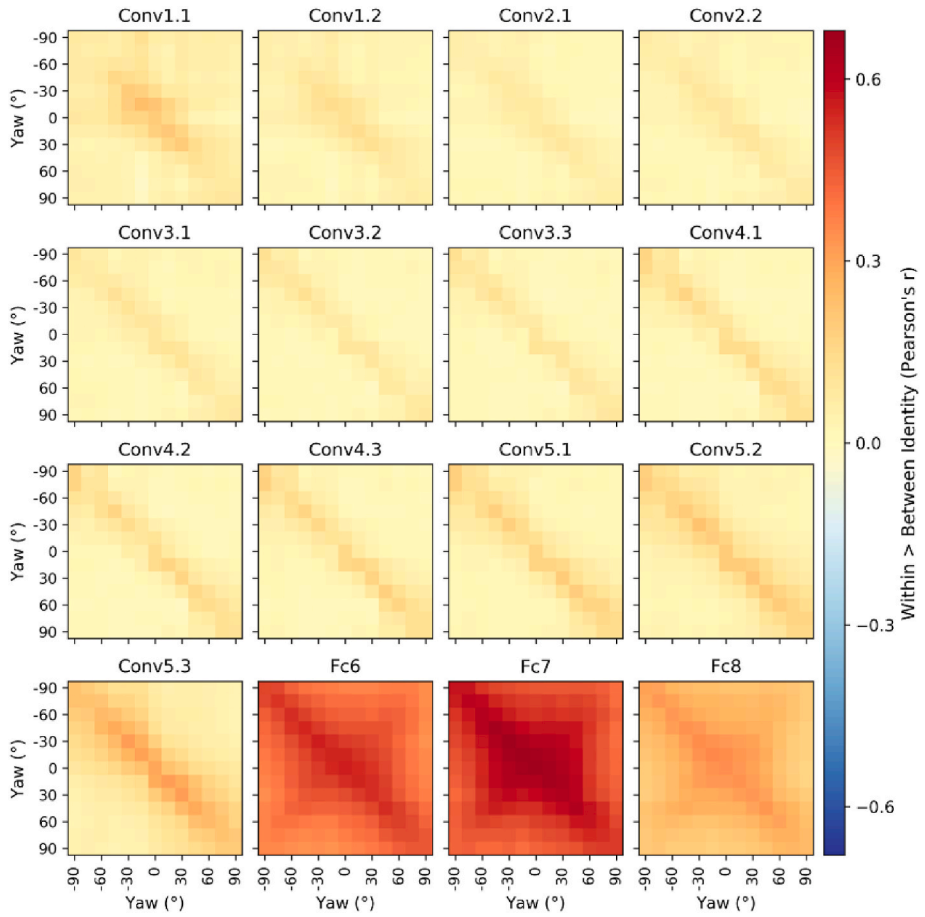


Fig. 2. Difference between within identity and between identity correlation matrices (see Fig. 1a and b) for yaw. Higher correlations for within identity comparisons were evident in the fully connected layers of the DCNN.

All imaging data was collected using a GE 3 T HD Excite MRI system with an eight-channel phased array head coil tuned to 127.4 MHz, at YNiC. We acquired T1-weighted structural images comprising 176 sagittal slices (TR = 7.74 ms, TE = 2.93 s, flip angle = 20°, FOV = 290 × 290 mm, matrix size = 256 × 256, 1 × 1.13 × 1.13 mm voxels). Functional data were acquired from 38 contiguous axial slices in a bottom-up interleaved order via a gradient-echo EPI sequence (TR = 3 s, TE = 32.7 ms, flip angle = 90°, FOV = 288 × 288 mm, matrix size = 128 × 128, 2.25 × 2.25 × 3 mm voxels).

Data were analysed with FEAT version 5.0.9 (http://www.fmrib.ox.ac.uk/fsl; Jenkinson et al., 2012). Data pre-processing included correction for head motion using rigid-body registration (MCFLIRT; Jenkinson et al., 2012), slice timing correction, non-brain removal (BET; Smith, 2002), spatial smoothing with a Gaussian kernel (FWHM = 5 mm), grand-mean intensity normalisation by a single multiplicative factor, and high-pass temporal filtering (σ = 50 s). The BOLD response for each condition was modelled with a boxcar function convolved with a double gamma haemodynamic response function, with head motion covariates added into the GLM (FILM; Woolrich et al., 2001).

Multivariate representational similarity analyses were performed with PyMVPA (version 2.6, Hanke et al., 2009). For each participant, we performed a whole-brain searchlight RSA (Kriegeskorte et al., 2008), using spherical ROIs with a 2 voxel (5.5 mm) radius. Within each sphere, the univariate parameter estimates were correlated between pairwise combinations of viewpoint conditions. The off-diagonal elements of the neural correlation matrix were then correlated against the model matrices (view-specific, view-symmetric, DCNN matrices, behavioural matrices). In each case, the resulting representational similarity correlation was assigned to the central voxel of the sphere, and the process repeated while iterating the sphere over the volume to generate whole-brain correlation maps. For each participant, the correlation maps were then transformed to the MNI 2 mm space. Group-level statistical maps were then generated using sign-flip permutation tests (5000 permutations), implemented with FSL's Randomize tool (Winkler et al., 2014), to contrast the individual correlation maps against zero. We display the group-level t-statistic maps masked by thresholding the corresponding p-statistic maps at $p < 0.001$ (uncorrected).

We visualised the group-level maps on the *fsaverage* cortical surface. To aid this visualisation, we additionally plotted the locations of early visual regions (V1, V2, V3 and hV4) obtained from a retinotopic atlas (Wang et al., 2015), and key face-preferential regions (FFA, OFA, and STS) obtained from an independent category localiser scan (see Noad et al., 2023).

3. Results

3.1. View-symmetric responses to yaw in fully-connected layers of DCNN

Fig. 1 shows the correlation for (a) within identity and (b) between identity face pairs across all combinations of viewpoint following yaw rotations of the head (see Suppl. Fig. 1). The convolutional layers (Conv) of the DCNN show a clear view-specific pattern with face images that have a similar viewpoint having similar patterns of response. These are shown on the diagonal from the top left to the bottom right. Interestingly, the view-specific tuning of the response in these convolutional layers increases for viewpoints away from frontal views which appear to have the tightest tuning. However, in the fully-connected layers (Fc), view-symmetry representations emerge. These are shown on the

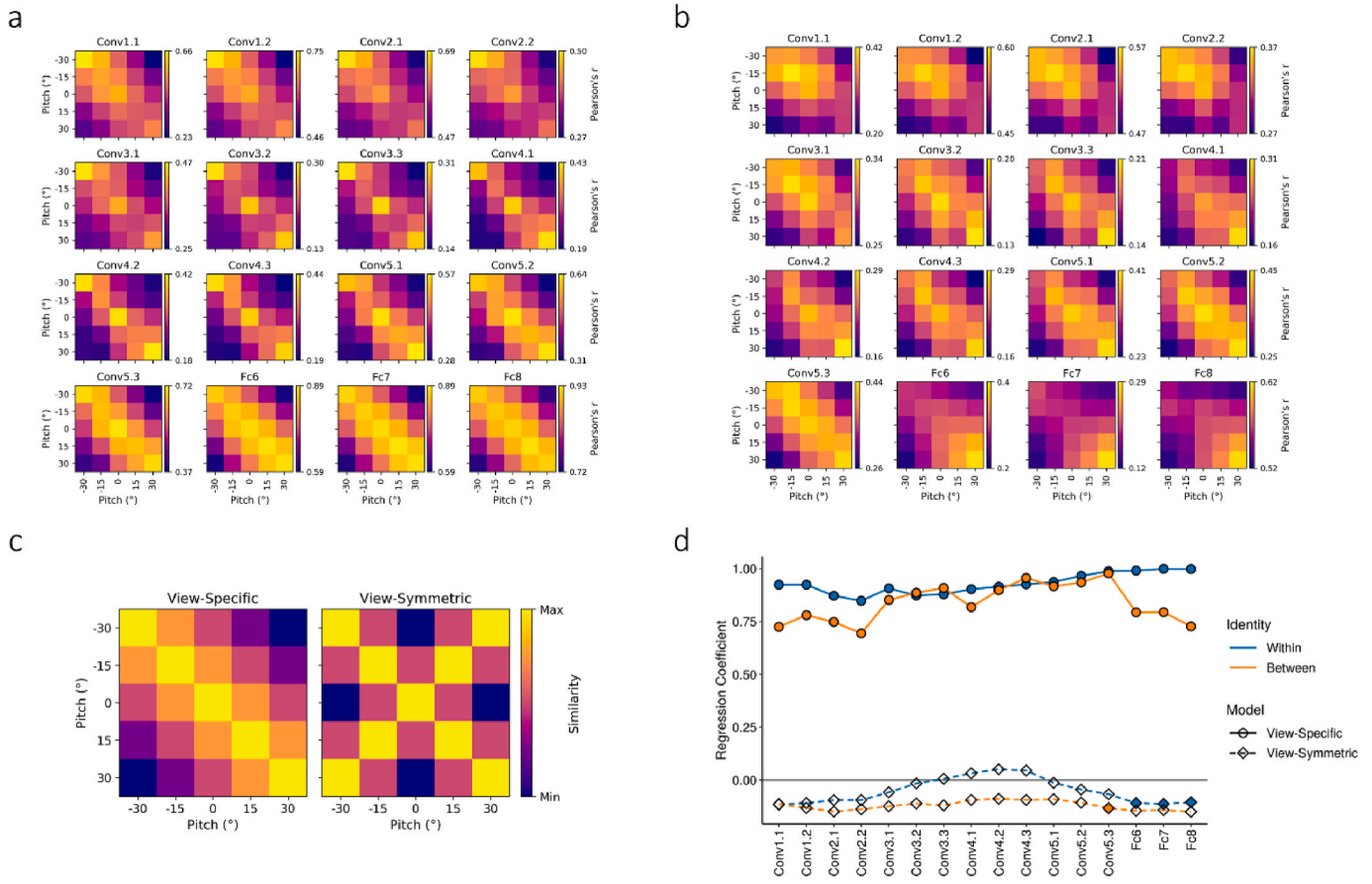


Fig. 3. Response of DCNN to images of faces that result from pitch rotations of the head. Correlation matrices showing the similarity to pairs of (a) within and (b) between identity comparisons at different viewpoints. (c) View-specific and view-symmetric models were used in a regression analysis of the correlation matrices. (d) View-specific responses were evident in both the convolutional (Conv) and fully-connected (Fc) layers of the DCNN. Filled symbols indicate coefficients showing a significant difference from zero.

diagonal from the bottom left to the top right. To quantify this effect, we used a regression-based representational similarity analysis to compare the correlation matrices in different layers of the DCNN against 2 models (Fig. 1c): a view-specific model (in which there was a graded response to changes in viewpoint) and a view-symmetric model (that included similar responses to symmetric views). Fig. 1d shows that early convolutional layers were best fit with a view-specific model for both within and between identity comparisons. However, in the fully-connected layers, representations were better predicted by the view-symmetry model, again for both within and between identity comparisons. Interestingly, view-specific representations were also evident in the fully-connected layers for within but not between identity comparisons.

Next, we compared within identity versus between identity similarity across different viewpoints in each layer of the DCNN. Fig. 2 & Suppl. Fig. 5 show the difference between the within (Fig. 1a) and between (Fig. 1b) identity correlation matrices for each layer of the DCNN. The difference in the convolutional layers was most evident for faces with a similar viewpoint. However, the biggest differences were evident in the fully connected layers, particularly Fc6 and Fc7. The difference was greater for faces with similar viewpoint, but also for faces with symmetric viewpoints. That is, the difference between same identity and different identity values in the DCNN was higher for symmetric compared to non-symmetric combinations in the fully-connected layers.

3.2. *View-specific but no view-symmetric responses to pitch in DCNN*

Fig. 3 shows the correlation matrices for (a) within identity and (b) between identity pairs of faces across all combinations of viewpoint following pitch rotations of the head (see Suppl. Fig. 2). To determine whether the patterns of response corresponded to a view-specific or view-symmetric, a regression analysis was performed. This showed that the responses to pitch in all layers (convolutional and fully-connected) were predicted by the view-specific, but not the view-symmetric model. This shows that the response to pitch in the DCNN is specific to a particular viewpoint and that symmetrical viewpoints do not lead to similar responses.

Next, we compared within identity versus between identity similarity across different viewpoints in each layer of the DCNN. Fig. 4 and Suppl. Fig. 6 show the difference between the within (Fig. 3a) and between (Fig. 3b) identity matrices for each layer of the DCNN. Throughout all layers the biggest difference was evident for faces with the same viewpoint. However, the biggest differences were evident in the fully connected layers, particularly Fc6 and Fc7.

3.3. *View-symmetric responses to roll in both the convolutional and fully-connected layers of DCNN*

Fig. 5 shows the correlation matrix for all combinations of viewpoint for (a) within identity and (b) between identity faces following roll rotations of the head (see Suppl. Fig. 3). To determine whether the patterns of response were view-specific or view-symmetric, a regression analysis was performed. The convolutional and fully-connected layers of the DCNN showed view-specific responses. However, in the later convolutional layers and the fully-connected layers, view-symmetry is more dominant.

Next, we compared within identity versus between identity

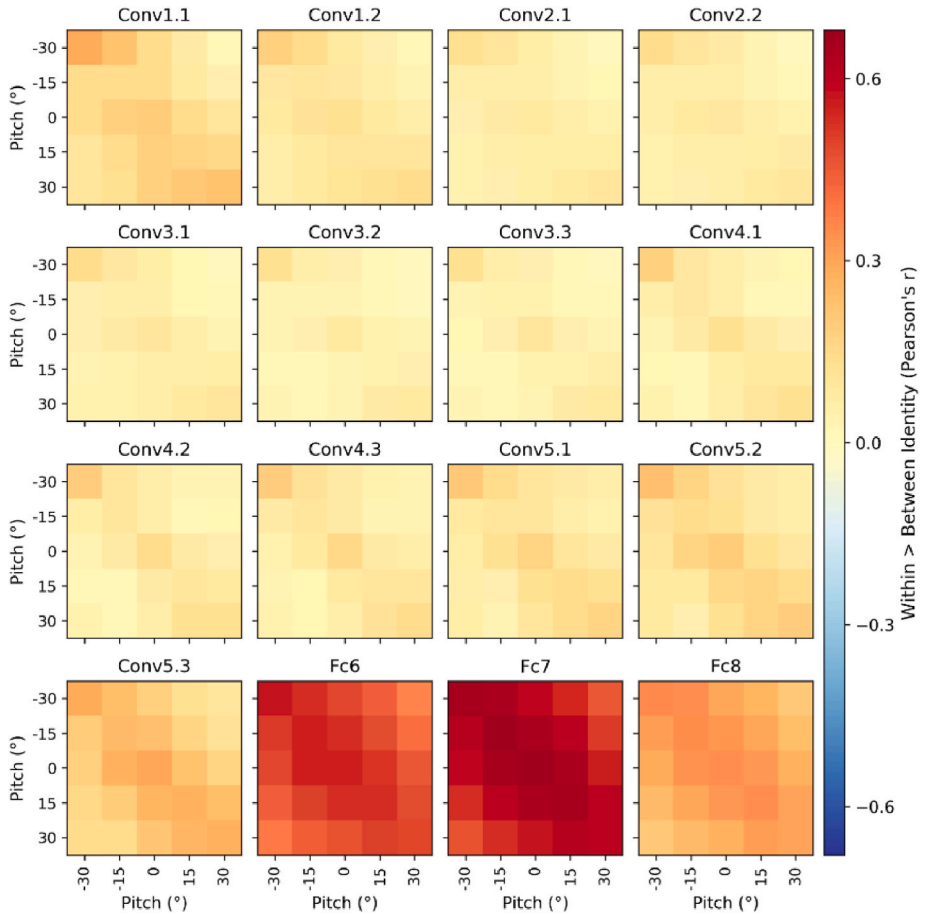


Fig. 4. Difference between within identity and between identity correlation matrices (see Fig. 3a and b) for pitch. Higher correlations for within identity comparisons were evident in the fully connected layers of the DCNN.

similarity across different viewpoints in each layer of the DCNN. Fig. 6 shows the difference between the within and between identity matrices for each layer of the DCNN. This shows that the biggest differences were evident in the fully connected layers, particularly Fc6 and Fc7 (Supplementary Fig. 7), and particularly for smaller rotations between ±30° within the range of plausible rotations for natural head movements.

3.4. *Fully-connected layers of DCNN predict perceptual similarity judgements*

Next, we asked whether view-symmetry for yaw emerged in the fully-connected layers of the DCNN for a different set of faces from the Radboud Faces database (see Suppl. Fig. 4). Supplementary Fig. 8 shows that the response of the convolutional layers of the DCNN was predicted by a view-specific model, whereas the response of the fully-connected layers is predicted by a view-symmetric model. Supplementary Fig. 9 shows that there was greater similarity for within identity compared to between identity face images in the fully-connected layers of the DCNN.

To determine whether behavioural responses in human participants were similar to the output of the DCNN, participants made judgements of perceptual similarity on the same face images. Fig. 7 shows the output of the DCNN and behavioural measures of perceptual similarity and response time for all combinations of viewpoint. A clear representation of identity is evident in the fully-connected layers with diagonal lines showing face pairs with the different viewpoints but the same identity (see Fig. 7a). A similar pattern is shown in Fig. 7b for the perceptual similarity ratings. We then correlated the behavioural matrices with corresponding matrices from each layer of the DCNN separately for perceptual similarity and response time. There was a low correlation between the behavioural measures and the convolutional layers of the DCNN. However, there was higher correlation between the behavioural measures and the fully-connected layers (see Fig. 7c). The negative correlations with response time reflect the fact that shorter response times indicate better performance. We also correlated the behavioural responses with the view-specific and view-symmetric models (see Suppl. Fig. 8c). View-symmetric responses were evident in the behavioural rating data, for both same identity ($r = 0.67$, $p = 0.03$) and different identity ($r = 0.77$, $p < 0.001$). In contrast, view-specific is only found in different identity ($r = 0.80$, $p < 0.01$), and not for same identity ($p > 0.2$). These findings indicate that perceptual similarity ratings given by human can be used to predict the outcomes of a deep neural network, and that this occurs particularly for same identity faces in the fully connected layers.

3.5. *View-symmetric neural responses in higher visual areas*

We next asked how neural representations of changes in yaw correlated with the view-specific and view-symmetric models and human ratings of perceptual similarity (Fig. 8; Table 1). Fig. 8b shows regions in the brain whose response is predicted by either the view-specific or view-symmetric models. View-specific representations were evident in bilateral regions of the occipital lobe that correspond to early visual areas (see Fig. 8a). In contrast, the view-symmetric representations were associated with more lateral and anterior bilateral responses in the occipito-temporal lobe. These responses overlapped with face regions, such as OFA (see Fig. 8a). These findings show a progression from a view-specific representation to a view-symmetric representation

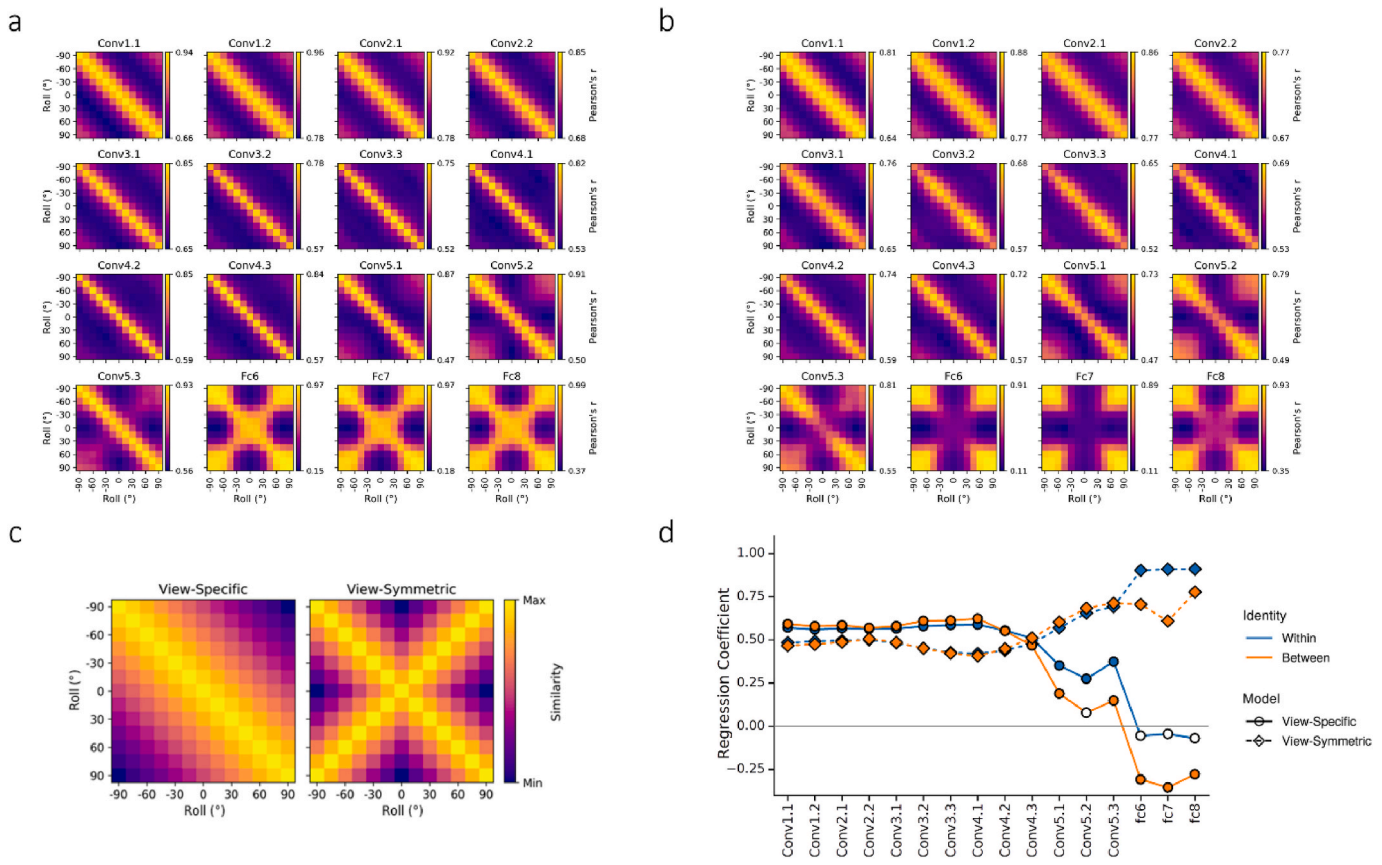


Fig. 5. Response of DCNN to images of faces that result from rotations of the head along the roll axis from the Pointing '04 Image Database. Correlation matrices showing the similarity to pairs of (a) within and (b) between identity faces at different viewpoints. (c) View-specific and view-symmetric models were used in a regression analysis of the correlation matrices. (d) View-specific and view-symmetric responses were evident in the convolutional layers of the DCNN. However, view-symmetric responses were dominant in the fully connected layers of the DCNN. Filled symbols indicate coefficients showing a significant difference from zero.

of faces from early to higher visual areas. Next, we asked how the perceptual similarity ratings correlated with patterns of response across the brain. This showed a more widespread pattern that overlapped with both the early and higher-level visual areas.

We then used a whole-brain search light analysis to compare the similarity of the neural response to different layers of the DCNN (Fig. 9). The output of the convolutional layers correlated with the neural response in early visual regions (see Fig. 8a). However, the output of the fully-connected layers correlated with more lateral and anterior regions that correspond with higher visual areas, such as the OFA (see Fig. 8a). These results show that the change in the neural representation of viewpoint evident in the DCNN corresponds to a change in neural representation along the ventral surface of the occipito-temporal lobes between early and higher visual areas.

4. Discussion

The aim of this study was to investigate how changes in the viewpoint of the face caused by rotations of the head are represented in human and artificial neural networks. We investigated rotations along 3 axes: yaw (left-right), pitch (up-down) and roll (in-plane rotation) in a deep convolutional neural network (DCNN). For yaw, we found that view-specific representations were evident in the convolutional layers of the DCNN, but that view-symmetric representations emerge in the fully-connected layers. The ability to discriminate same identity images from different identity images was most evident in the fully-connected layers and this difference was enhanced for symmetrical compared to non-symmetrical viewpoints. However, a similar progression from view-specific to view-symmetric representations was not evident for pitch or roll. Finally, we asked whether the responses to yaw in a DCNN predicted behavioural and neural responses in humans. We found that the output of the convolutional layers of the DCNN predicted patterns of response in early visual areas, whereas the output of the fully-connected layers predicted behavioural responses and patterns of response in higher visual areas.

In the first part of this study, we compared the representation of face images resulting from different rotations of the head in a DCNN trained to recognise face identity (Parkhi et al., 2015). For rotations along the left-right axis (yaw), we found that the initial representation of the face in the convolutional layers of the DCNN is view-specific. That is, faces with similar viewpoints generated similar patterns of response. However, in the later fully-connected layers of the DCNN, we found that the output of the DCNN showed view-symmetry. That is, face images generated from opposite rotations of the head (e.g. left and right profile) generated similar patterns of response. The DCNN was trained to recognise identity not viewpoint. This suggests that the response to view-symmetry is an emergent feature that is important for recognition of identity. One potential issue in the interpretation of this data is that the training of DCNN was augmented by the flipping the images across the horizontal axis. This may have led to increased matching of images along the yaw axis. Nonetheless, matching faces across different images occurs in natural viewing as a result of changes in head rotation or movement of the observer. Moreover, these findings are consistent with previous studies showing that view-symmetry emerges in artificial neural networks for head rotations along the yaw axis that have not been trained with images flipped across the horizontal axis (Leibo et al., 2017; Yildirim et al., 2020; Farzmahdi et al., 2023).

To determine whether view-symmetry results from the pooling of

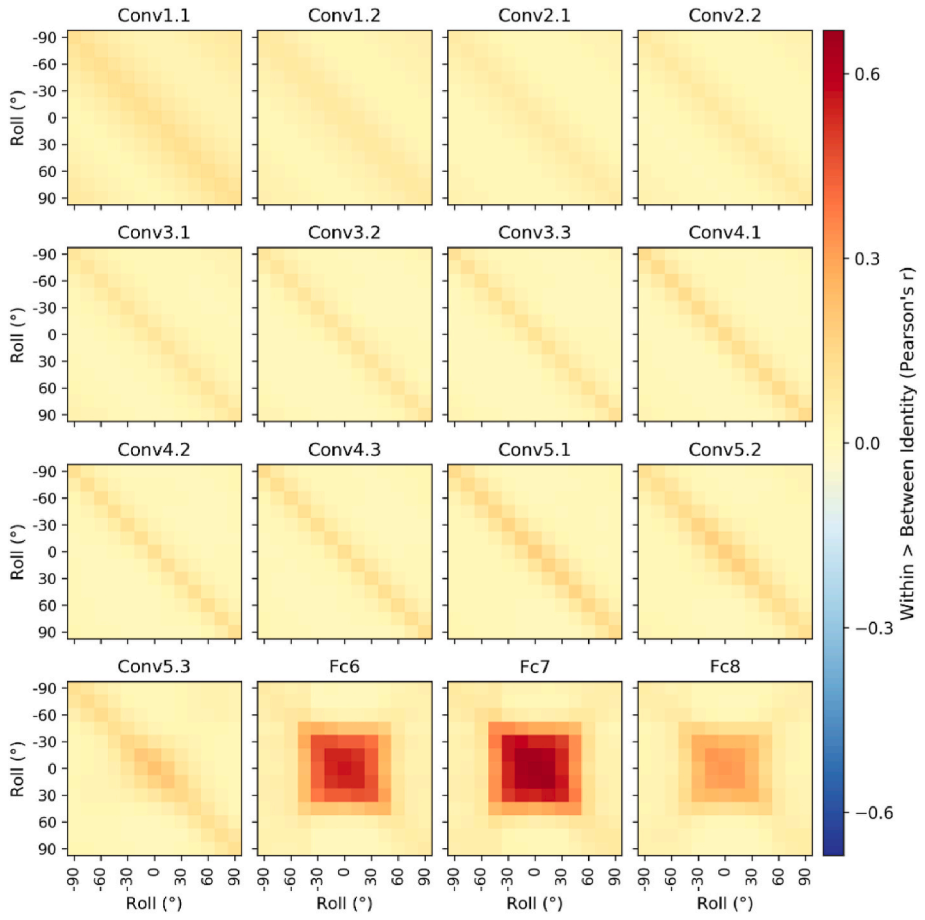


Fig. 6. Difference between within identity and between identity correlation matrices (see Fig. 5a and b) for roll. Higher correlations for within identity comparisons were evident in the fully connected layers of the DCNN.

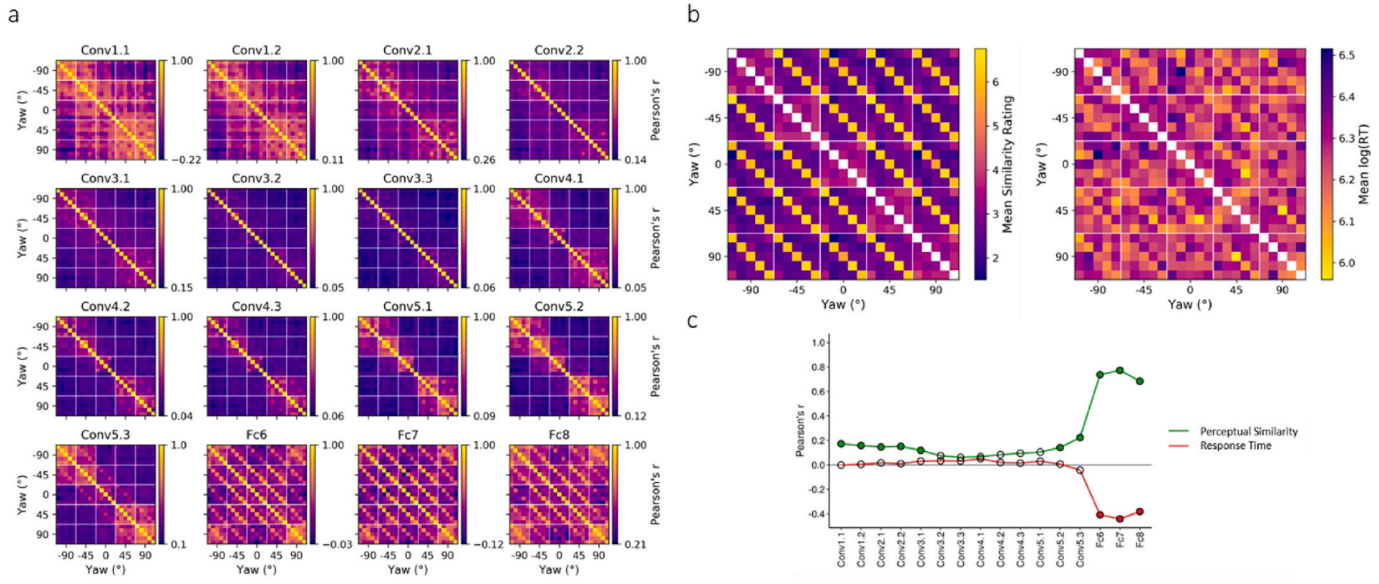


Fig. 7. Similarity between behavioural judgements of yaw in humans and the DCNN. (a) Correlation matrices showing the similarity to individual pairs of faces at different viewpoint combinations in different layers of the DCNN. (b) Perceptual similarity and response time to individual pairs of faces at different viewpoint combinations from behavioural task. (c) Correlation between the DCNN and behavioural measures. The highest correlations between the behavioural measures and the DCNN were evident in the fully connected layers. Negative correlations with response time indicated that lower response time correspond to better performance. Filled symbols indicate correlations showing a significant difference from zero.

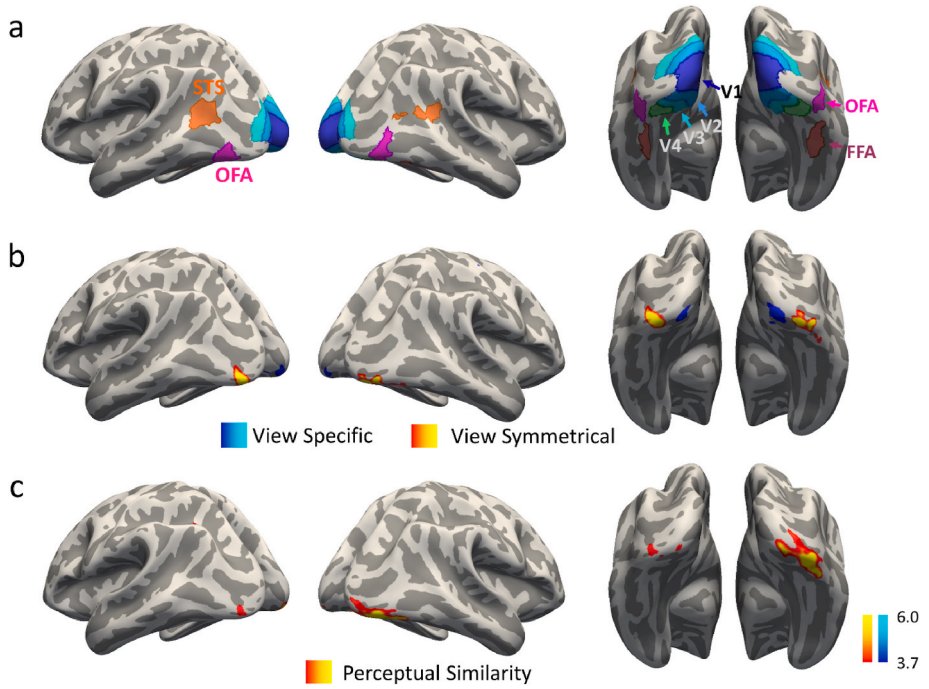


Fig. 8. Searchlight fMRI analysis of models and perceptual judgements. (a) Location of early visual regions and face regions. (b) Regions that show view-specific (blue-light blue) and view-symmetric (red-yellow) responses. (c) Regions that correspond to perceptual similarity judgements. Responses were thresholded at $p < 0.001$ (uncorrected). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Local maxima MNI coordinates and t-values of group average searchlight maps.

Model	Left hemisphere		Right hemisphere	
	MNI coordinates	t	MNI coordinates	t
Conv1	−12, −86, −16	4.9	16, −92, −12	5.3
Conv2	−14, −94, −14	5.0	16, −92, −12	4.9
Conv3	−14, −92, −12	6.0	16, −90, −12	5.4
Conv4	−14, −92, −12	6.4	16, −90, −12	5.6
Conv5	−14, −92, −12	6.5	22, −86, −12	6.2
FC6	−32, −92, −10	6.2	40, −88, −10	7.9
FC7	−32, −92, −10	6.7	40, −88, −10	7.7
FC8	−32, −94, −10	6.3	40, −88, −10	7.1
Perceptual similarity	−22, −88, −18	7.0	34, −80, −16	12.8
View-specific	−14, −92, −10	5.4	16, −92, −12	5.7
View-symmetric	−30, −92, −10	7.6	30, −88, −14	8.1

mirror representations or simply occurs for opposite rotations of the head, we measured the response to pitch. Because the face is not symmetric across the horizontal axis (i.e. the top half of the face is not a mirror image of the bottom half), upward movements of the face will not be mirror images of downward movements of the face. For pitch, view-specific representations were evident in both the convolutional and fully-connected layers of the DCNN, but there was no evidence of a view-symmetric patterns of response. Therefore, in contrast to yaw, a similar progression from view-specific to view-symmetric representation in the DCNN was not evident for face images generated by rotations in pitch. These findings are consistent with behavioural findings in humans showing that there was no advantage for face recognition across symmetrical views caused by rotations in pitch (Favelle and Palmisano, 2018).

View-symmetric representations were evident for in-plane rotations of the head (roll). However, these view-symmetric representations were evident throughout the convolutional layers of the DCNN. These appear to be driven by the similarity of ± 90, which give rise to horizontally oriented faces. Although these are mirror images, it seems unlikely that the convergence of mirror representations of the face is driving the effect at this early stage of the DCNN. Rather, it would seem that the similarity in response reflects the similarity in the orientation of the face image. In the fully-connected layers, view-symmetry dominated view-specific responses. It is more likely that this reflects the actual convergence of mirror representations. These findings fit with neuroimaging findings showing that responses to roll do not emerge from an initial view-specific response (Rogers and Andrews, 2022). This may reflect the fact that these head rotations are less common in natural viewing.

Together these results provide support for the idea that view-symmetry in DCNNs results from the convergence of mirror-symmetric representations (Farzmahdi et al., 2023). Because faces have bilateral symmetry along the vertical axis, left-right rotations (yaw) generate mirror-symmetric images (e.g. left and right ¾ views or left and right profile). These mirror-symmetric images are likely to lead to mirror-reversed representations in the convolutional layers of the DCNN. Because of the spatial organization of the convolutional layers, it is not possible to pool or converge these mirror representations. However, it is possible for these representations to converge in the fully-connected layers of the DCNN and it is here that the view-symmetric responses appear.

To determine whether these symmetric representations confer an advantage for face recognition, we compared the responses to same identity and different identity faces in the DCNN. As VGG-Face is a deep learning model designed to recognise faces (Parkhi et al., 2015), we would predict a higher similarity in the output of face pairs with the same identity compared to face pairs with different identities, particularly in the fully-connected layers of the DCNN. We first compared the difference between same and different identity responses across all combinations of viewpoint following yaw rotations. In the convolutional layers of the DCNN, a significant difference between same identity (within) and different identity (between) faces was only evident when the faces had a similar viewpoint. Presumably, this reflects the fact that same identity faces shown from the same viewpoint have more similar image properties than corresponding images from different identity faces. However, in the fully-connected layers, the difference between same identity and different identity faces was greater across all

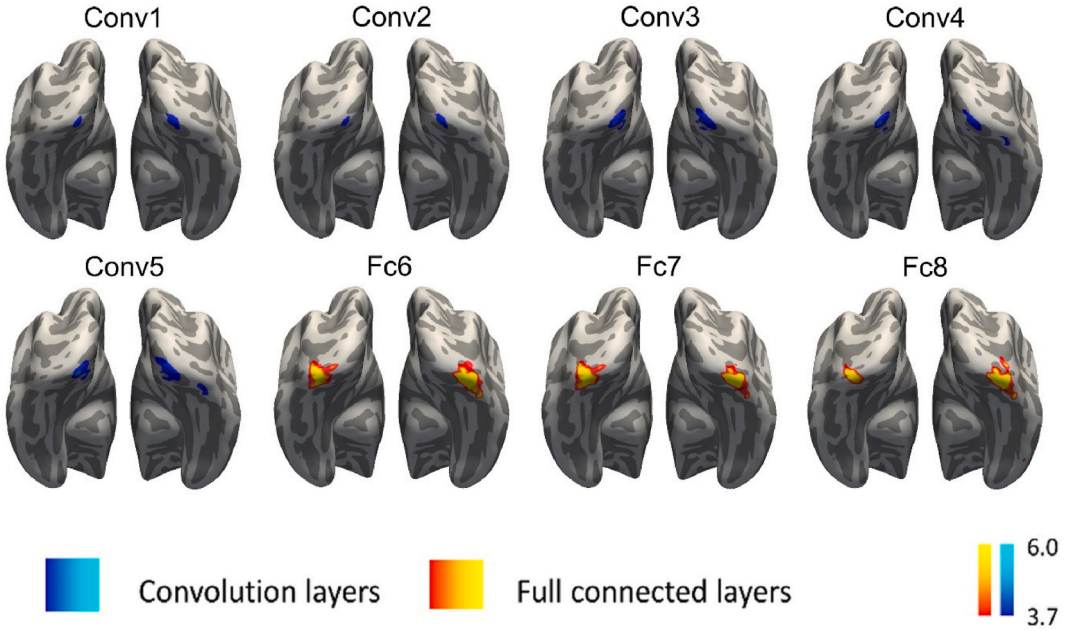


Fig. 9. Searchlight fMRI analysis showing the correspondence with different layers of the DCNN. Convolutional layers of the DCNN correlated with responses in early visual areas. In contrast, the fully connected layers of the DCNN correlated with responses in higher visual areas. Responses were thresholded at $p = 0.001$ (uncorrected).

combinations of viewpoint. Interestingly, for yaw the difference was greatest for the same or symmetrical viewpoints (see Fig. 2 and Suppl. Fig. 5). This provides support for the idea that view-symmetry plays an important role in the recognition of faces in the DCNN. This is consistent with findings from human behaviour showing that the recognition of newly learnt faces is greater when the faces at test are presented with a symmetrical rather than a non-symmetrical viewpoint (Flack et al., 2019). The findings also follow neurophysiological data showing that neurons in face region AL are selective for view-symmetric faces from the same identity (Freiwald and Tsao, 2010).

Previous studies have shown that information about viewpoint continues to be coded alongside identity in the fully-connected layers of the DCNN (Parde et al., 2017; Hill et al., 2019) and that faces from the same identity elicit a more similar response if shown from a similar viewpoint (Abudarham et al., 2021). Our findings extend these findings by showing that this viewpoint advantage is also evident for symmetric views. A similar view-symmetric difference between same-identity and different-identity pairs was not evident for pitch or roll (see Figs. 3 and 5, Suppl. Figs. 6–7). For pitch, there was a bias toward same-identity faces with the same viewpoint (see Fig. 3 and Suppl. Fig. 6). This again suggests that the recognition of faces is sensitive to viewpoint even in the fully-connected layers. For roll, the bias was for rotations around vertical or 0° (see Fig. 3 and Suppl. Fig. 6). This may reflect that, in natural viewing, faces do not typically vary along the roll axis.

In their seminar paper on the neuronal response to facial viewpoint, Freiwald and Tsao (2010) measured responses to faces that varied across yaw, pitch and roll. Consistent with the output of the fully-connected layers of the DCNN, they showed neurons in face region AL had view-symmetric responses to yaw and roll. 75% of the neurons in AL showed view-symmetric responses, but they did not report the relative occurrence of symmetry for yaw and roll. In contrast, they did not report any view-symmetric responses to pitch. In a separate analysis, they recorded the responses to different viewpoints from cells in region AL. The different viewpoints included rotations along yaw (left full profile, left half profile, right half profile, right full profile) and pitch (straight, up, down). They showed that the responses to different viewpoints collapsed into 3 clusters: (1) left and right full profiles, (2) left and right half profiles, and (3) up, down, and straight. This provided further support for a view-symmetry for yaw. Howeer, as they only had 3 viewpoints for pitch it was not possible to determine view-symmetry in this analysis.

In the second part of the study, we compared the output of different layers of the DCNN with the perceptual and neural responses in humans. Given that a clear view-symmetrical pattern was evident for yaw, we restricted our analysis to rotations of the face along this axis. First, we measured the perceptual similarity for face pairs with different combinations of viewpoint. The behavioural pattern of response was then correlated with the output from the different layers of DCNN. This showed that the fully-connected layers of the DCNN predicted perceptual similarity and response time. We then compared patterns of response from the DCNN with neural responses in the brain. Using a searchlight analysis, we measured the similarity in the pattern of response to pairs of faces across all combinations of viewpoint in all voxels of the brain. This similarity matrix was then compared with the corresponding similarity matrix from each layer of the DCNN. The results showed that the output of the convolutional layers predicted responses in early visual areas. However, the output of the fully-connected layers predicted responses in higher visual areas, including face regions. These results are consistent with previous neuroimaging studies showing a change in the representation from view-specific to view-symmetric along the visual hierarchy (Axelrod and Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019; Rogers and Andrews, 2022). This transition in the representational properties may correspond to differences in the functional organization of these regions. The retinotopic organization of early visual regions imposes strong spatial constraints to processing, which may be analogous to the convolutional layers of the DCNN. In contrast, higher visual areas are less constrained by the spatial location and may, like the fully-connected layers, be able to pool information in a way that generates view symmetric responses. This may reflect a qualitative difference in the way that information is processed in early compared to higher visual areas.

The discovery of view symmetry for yaw rotations in human and artificial neural networks supports the idea that this is an intermediate representation from a view-specific to a view-invariant representation (Freiwald and Tsao, 2010). However, given that a view-invariant representation requires the convergence of all viewpoints, it is not

immediately obvious why other (non-symmetrical) combinations of viewpoint are not evident in an intermediate representation. One possible explanation for this is that, if view-symmetric responses reflect the convergence of mirror representations, this may reflect a more efficient (less computationally expensive) way of representing information. Combinations of non-symmetrical viewpoints would not be based on mirror representations and would not confer the same efficiency.

In conclusion, we compared view-symmetry in humans and an artificial neural network trained to recognise faces (VGG-Face). We found that view-symmetry is an emergent property of the DCNN for left-right rotations (yaw). There was an initial view-specific representation in the convolutional layers, but then a view-symmetric representation emerged in the fully-connected layers. The ability to differentiate identity was greater across symmetrical compared to non-symmetrical viewpoints suggesting that this may facilitate recognition. A similar transition from a view-specific to a view-symmetric representation was not evident for either pitch or roll. Next, we compared patterns of response in the DCNN to changes in viewpoint with corresponding behavioural and neural responses in humans. We found that the response of the convolutional layers of the DCNN correlated with early visual areas, whereas the response of the fully-connected layers correlated with higher visual regions. Finally, we found that the fully-connected layers of the DCNN predicted judgements of perceptual similarity. These findings suggest that view-symmetry may reflect the efficient coding of information in the brain and that this confers a recognition advantage in both humans and artificial neural networks.

CRediT authorship contribution statement

Xun Zhu: Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **David M. Watson:** Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Daniel Rogers:** Formal analysis, Investigation. **Timothy J. Andrews:** Conceptualization, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgements

This work was also sponsored in part by the China Scholarship council (CSC) scholarship No. 202108330320. We would like to thank Mike Burton for helpful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2024.109061.

Data availability

Data will be made available on request.

References

Abudarham, N., Grosbard, I., Yovel, G., 2021. Face recognition depends on specialized mechanisms tuned to view-invariant facial features: insights from deep neural networks optimized for face or object recognition. Cognit. Sci. 45 (9), e13031.
Axelrod, V., Yovel, G., 2012. Hierarchical processing of face viewpoint in human visual cortex. J. Neurosci. 32 (7), 2442–2452.
Bruce, V., Young, A., 1986. Understanding face recognition. Br. J. Psychol. 77 (Pt 3), 305–327.
Burton, A.M., Bruce, V., Hancock, P.J.B., 1999. From pixels to people: a model of familiar face recognition. Cognit. Sci. 23 (1), 1–31.
Busey, T.A., Zaki, S.R., 2004. The contribution of symmetry and motion to the recognition of faces at novel orientations. Mem. Cognit. 32 (6), 916–931.
Cichy, R.M., Kaiser, D., 2019. Deep neural networks as scientific models. Trends Cognit. Sci. 23 (4), 305–317.
Farzmahdi, A., Zarco, W., Freiwald, W., Kriegeskorte, N., Golan, T., 2023. Emergence of brain-like mirror-symmetric viewpoint tuning in convolutional neural networks. Elife 13, e90256.
Favelle, S., Palmisano, S., 2018. View specific generalisation effects in face recognition: front and yaw comparison views are better than pitch. PLoS One 13 (12), e0209927.
Flack, T.R., Harris, R.J., Young, A.W., Andrews, T.J., 2019. Symmetrical viewpoint representations in face-selective regions convey an advantage in the perception and recognition of faces. J. Neurosci. 39 (19), 3741–3751.
Freiwald, W.A., Tsao, D.Y., 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330 (6005), 845–851.
Gourier, N., 2004. Estimating face orientation from robust detection of salient facial features. In: Proceedings of Pointing 2004, ICPR. International Workshop on Visual Observation of Deictic Gestures, Cambridge, UK.
Guntupalli, J.S., Wheeler, K.G., Gobbini, M.I., 2017. Disentangling the representation of identity from head view along the human face processing pathway. Cerebr. Cortex 27 (1), 46–53.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7 (1), 37–53.
Hill, M.Q., Parde, C.J., Castillo, C.D., Colon, Y.I., Ranjan, R., Chen, J.C., et al., 2019. Deep convolutional neural networks in the face of caricature. Nat. Mach. Intell. 1 (11), 522–529.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62 (2), 782–790.
Kietzmann, T.C., Swisher, J.D., König, P., Tong, F., 2012. Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. J. Neurosci. 32 (34), 11763–11772.
Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis-connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 249.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A., 2010. Presentation and validation of the Radboud faces database. Cognit. Emot. 24 (8), 1377–1388.
Leibo, J.Z., Liao, Q., Anselmi, F., Freiwald, W.A., Poggio, T., 2017. View-tolerant face recognition and Hebbian learning imply mirror-symmetric neural tuning to head orientation. Curr. Biol. 27 (1), 62–67.
Noad, K., Watson, D., Andrews, T., 2023. A network of regions in the human brain involved in processing familiarity. J. Vis. 23 (9), 4913, 4913.
O'Toole, A.J., Castillo, C.D., 2021. Face recognition by humans and machines: three fundamental advances from deep learning. Annu Rev Vis Sci 7, 543–570.
O'Toole, A.J., Castillo, C.D., Parde, C.J., Hill, M.Q., Chellappa, R., 2018. Face space representations in deep convolutional neural networks. Trends Cognit. Sci. 22 (9), 794–809.
Parde, C.J., Castillo, C., Hill, M.Q., Colon, Y.I., Sankaranarayanan, S., Chen, J.C., O'Toole, A.J., 2017. Face and image representation in deep CNN features. 12th IEEE International Conference on Automatic Face & Gesture Recognition.
Parkhi, O., Vedaldi, A., Zisserman, A., 2015. Deep Face Recognition.
Perrett, D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J., Thomas, S., 1991. Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp. Brain Res. 86 (1), 159–173.
Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., et al., 2018. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. Proc. Natl. Acad. Sci. USA 115 (24), 6171–6176.
Rogers, D., Andrews, T.J., 2022. The emergence of view-symmetric neural responses to familiar and unfamiliar faces. Neuropsychologia 172, 108275.
Rolls, E.T., 2012. Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. Front. Comput. Neurosci. 6, 35.
Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. USA 104 (15), 6424–6429.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
Troje, N.F., Bülthoff, H.H., 1998. How is bilateral symmetry of human faces used for recognition of novel views? Vision Res 38 (1), 79–89.
Wang, L., Mruczek, R.E., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cerebr. Cortex 25 (10), 3911–3931.
Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. Neuroimage 92, 381–397.
Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14, 1370–1386.
Yildirim, I., Belledonne, M., Freiwald, W., Tenenbaum, J., 2020. Efficient inverse graphics in biological face processing. Sci. Adv. 6 (10), eaax5979.
Young, A.W., Burton, A.M., 2017. Recognizing faces. Curr. Dir. Psychol. Sci. 26 (3), 212–217.