University of Tartu
Faculty of Social Sciences
Institute of Psychology

Carolin Lüübek

Comparing Psychedelic Visualizations with Convolutional Neural Networks' Feature Visualizations

Research project

Supervisor: Jaan Aru

Running head: Psychedelic and feature visualizations

Tartu 2023

Comparing Psychedelic Visualizations with Convolutional Neural Networks' Feature Visualizations

Abstract

This research studies the relationship between psychedelic visual experiences and convolutional neural networks' (CNNs') feature visualizations. Subjects who had experiences with serotonergic psychedelic substances were asked to rate how well the feature visualizations of two CNNs, one of which explains the primate visual system better than the other, corresponded to their own psychedelic visualizations. No correlation was found between the overall intensity of the psychedelic experiences and the feature visualizations' ratings. Further implications include that the intensity of psychedelic visualizations is better captured by specific measures, such as the complexity and richness of the visualizations. These measures, as well as the inclusion of higher-level properties, such as fractal shapes and distorted objects in one's psychedelic visualizations, contribute more to the perceived correspondence with feature visualizations, while lower-level properties have less influence.

Keywords: psychedelic visualizations, ventral visual hierarchy, convolutional neural networks, feature visualization

Psühhedeelsete Nägemuste ja Konvolutsiooniliste Närvivõrkude Featuuride Visualisatsioonide Seosed

Kokkuvõte

Käesolevas töös uuritakse psühhedeelsete nägemuste ja konvolutsiooniliste närvivõrkude featuuride visualisatsioonide vahelisi seoseid. Serotoniinergilisi psühhedeelikume kasutanud katseisikud hindasid oma psühhedeelsete nägemuste põhjal kahe tehisnärvivõrgu visualisatsioone, millest üks tehisnärvivõrk on parem primaatide nägemishierahia mudel kui teine. Psühhedeelse kogemuse üldise intensiivsuse ja featuuride visualisatsioonide hinnangute vahel seost ei leitud. Tulemuste põhjal järeldatakse, et täpsemad näitajad nagu nägemuste keerukus ja rikkalikkus tabavad psühhedeelsete nägemuste intensiivsust paremini. Tehisnärvivõrkude visualisatsioonid vastasid psühhedeelsetele nägemustele paremini, kui nägemused olid olnud keerukamad, näiteks sisaldanud fraktaleid või moonutatud objekte.

Märksõnad: psühhedeelsed nägemused, nägemishierarhia, konvolutsioonilised närvivõrgud, featuuride visualisatsioonid

Introduction

This research project aims to enhance our understanding of the psychedelic experience by leveraging the modeling capabilities of artificial neural networks (ANNs). Specifically, the inner activity of certain computer vision models called convolutional neural networks (CNNs) is visualized, and those visualizations, as well as their underlying mechanisms, are compared to the visualizations elicited by and the working mechanisms of serotonergic psychedelic substances. This is done both theoretically - by placing the study in the framework where CNNs serve as models of the primate visual system - and in practice - by conducting an experiment on individuals with previous psychedelic experiences.

Physiological and Phenomenological Effects of Psychedelics

Serotonergic psychedelics refer to a subgroup of hallucinogenic substances whose psychoactive effects are primarily mediated by the agonism at the serotonin 5-HT2A receptors in the brain (Nichols, 2016), the “classical” examples of such substances being LSD, DMT, psilocybin, and mescaline. The 5-HT2A receptors are particularly expressed in the cortical regions of the brain, including the frontal, temporal and occipital (containing the visual cortex) cortices (Beliveau et al., 2017). The time-resolved findings from the multimodal imaging study of Timmermann et al. (2023) suggest that the initial effect of psychedelic substances is at the highest levels of cortical organization, such as at the evolutionarily recent transmodal (nonsensory-specific) regions (Braga et al., 2015), with a myriad of effects following in the networks overlapping with the expression density of serotonin receptors. Of specific interest to this study is the increased connectivity between the visual cortex and the higher cortical regions, which has been hypothesized to play a role in establishing the visual quality characteristic for psychedelic experiences (Timmermann et al., 2023; Carhart-Harris et al., 2016).

This hypothesis is supported by the correlation examined between the aforementioned physiological effects and some specific phenomenological effects induced by psychedelic substances. The increased pairwise functional connectivity between the visual and e.g. frontoparietal and default mode networks correlates with the subjective intensity of the psychedelic experience (Timmermann et al., 2023). To focus specifically on visual phenomenology, the psychedelic-induced increased visual cortex global functional connectivity and cerebral blood flow, as well decreased visual cortex alpha power, thought to be tied to

general inhibitory processes (Jensen et al., 2010), correlate with the magnitude of visual hallucinations (Carhart-Harris et al., 2016).

The increases in the visual areas' activity and connectivity suggest that intrinsic brain activity has a greater influence on visual processing under the influence of psychedelics (Carhart-Harris et al., 2016). This is illustrated by the psychedelic-induced increased activity of parahippocampal and retrosplenial cortices, both involved in the retrieval of episodic memories, during a closed-eye imagery task (de Araújo et al., 2002; de Araújo et al., 2012). The finding could suggest that during a psychedelic experience, there is "an endogenous engagement of mnemonic circuits, possibly feeding visual areas with the content" of the visual experiences (de Araújo et al., 2012, p. 2558). During the same closed-eye imagery task, increased activity of associative as well as primary visual areas was examined, the latter of which is generally concerned with processing external stimuli, suggesting that "the visual cortex behaves "as if" there is external input when there is none" (Carhart-Harris et al., 2016, p. 4856). In a similar vein, the spontaneous internally generated activity of the primary visual cortex's neurons manifests as geometric visual hallucinations (Butler et al., 2011), which may in part arise due to the weakening of cortical inhibition, induced by psychedelic substances (Passie et al., 2008).

These findings point towards the visual phenomenology associated with the influence of psychedelic substances arising due to a myriad of effects in the brain regions with a dense expression of serotonin 5-HT2A receptors, the causal chain between which is still not totally understood, with for example the hierarchical predictive coding account of neuronal processing offering a clearer take on it.

Psychedelics in the Hierarchical Predictive Coding Theory

There is a consensus that the brain is not passively receiving sensory inputs, but is rather caught up in predicting and providing explanations for them (Bastos et al., 2012), one of the most computationally plausible models of it being that of hierarchical predictive coding (Rao et al., 1999; Friston, 2005; Bastos et al., 2012). In the visual domain, the hierarchical predictive processing model suggests that top-down connections from higher visual areas carry predictions of lower-level neural activities, and bottom-up connections carry prediction errors resulting from comparing the aforementioned predictions with actual sensory inputs. The ultimate goal of this relationship is to minimize prediction errors by optimizing the predictions (also referred to as priors or beliefs) (Rao et al., 1999; Friston, 2005; Bastos et al., 2012; Carhart-Harris and Friston,

2019). Under normal conditions, the priors can be rather heavily weighted, meaning that it's harder for the prediction errors to actually modify them, in extreme cases resulting in "entrenched pathological priors" (Carhart-Harris and Friston, 2019, p. 317), possibly being one of the reasons behind a number of mental illnesses.

In this framework, the supposed explanation for the full breadth of subjective phenomena associated with the consumption of psychedelics is that via stimulating the serotonergic deep pyramidal cells in the cortex, where priors are thought to be encoded, the confidence in the predictions is disrupted (Carhart-Harris and Friston, 2019). This results in their sensitization to prediction errors, accompanied by the "liberated bottom-up information flow" (Carhart-Harris and Friston, 2019, p. 317). The physiological effects of psychedelic substances described in the last subchapter can be seen as largely coherent with this, e.g. the decreased visual cortex alpha power corresponding to the relaxed top-down modulation, and the increased primary cortex cerebral blood flow corresponding to the increased signaling from lower-level units. In addition to this, as we shall see in the following subchapters, some intuitive or functional parallels can be drawn between this framework and the technique of feature visualization (and those similar to it, like the Deep Dream algorithm (Mordvintsev et al., 2015)), strengthening the foundation on which psychedelic experiences could be studied with the help of CNNs.

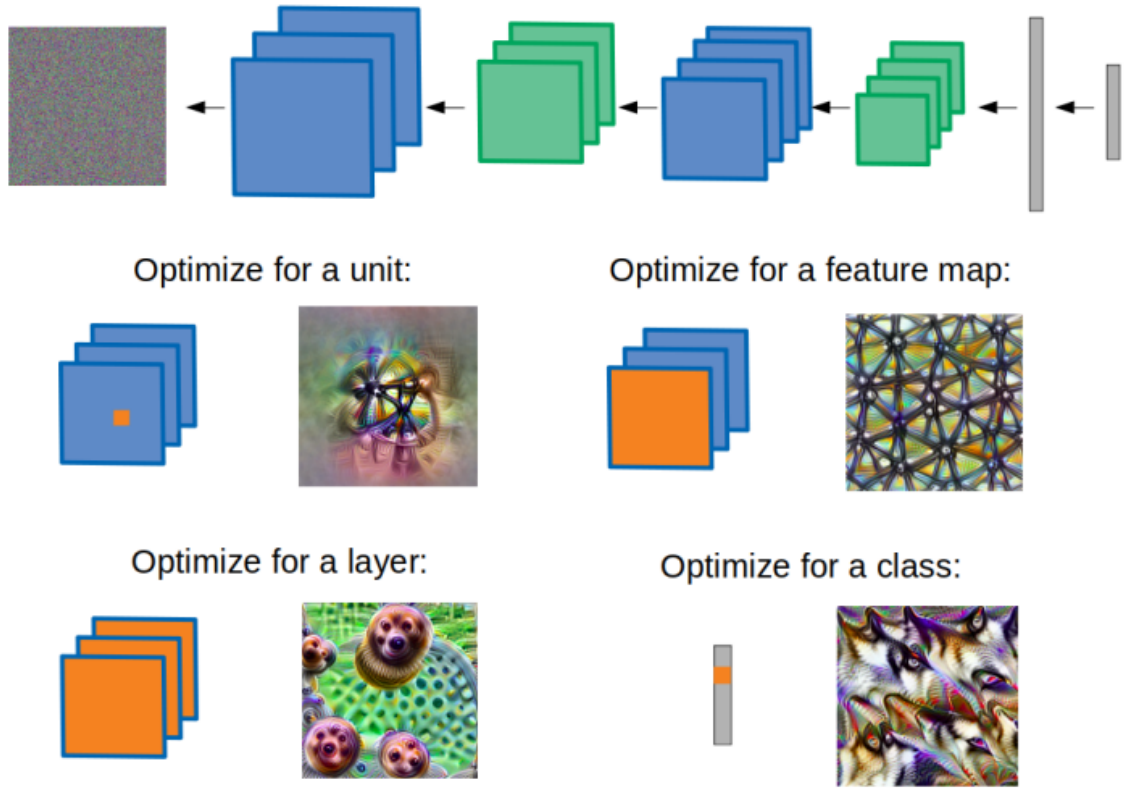
Feature Visualization

This subchapter replicates for the most part the literature review exhibited in Lüübek (2023). Feature visualization is a technique that aims to increase the explainability of artificial neural networks (ANNs) (Olah et al., 2017). In this research, feature visualization implies activation maximization of certain units of convolutional neural networks (CNNs), a type of computer vision models, by input optimization. The potential (explanatory power) of activation maximization rests on the idea that "a pattern to which the unit is responding maximally could be a good first-order representation of what a unit is doing" (Erhan et al., 2009, p. 4). It is mathematically executed by the calculation of gradients - a tool for manipulating the interaction between different network components (Lindsay, 2021), in this case between the input image to the network and the artificial neurons, channels, or layers of the network (Figure 1). During the process of feature visualization, gradients are used to determine how individual pixels of the input image should be changed in order to make the object of optimization activate more (Olah et al., 2017). This means that the final result represents the causal reasons behind the activity of the

unit, helping to differentiate between the features that cause and the ones that merely correlate with its behavior (Olah et al., 2017).

Figure 1

Feature Visualization Process



Note. Figure adapted from Lindsay (2021). The arrows represent the gradients flowing between the object of optimization and the input image.

Feature visualizations can not be taken as representing the entirety of a unit's behavior, as they represent its activity merely under the most optimal conditions (Olah et al., 2020a). Nonetheless, they do offer a human-understandable idea of the inner workings of a model, in addition to the potentiality of offering a new representation or primitive for thinking about the world, illustrated by the high-low frequency detectors forming in the earlier layers of a network called InceptionV1 (OpenVis Conference, 2018; Olah et al., 2020b; Szegedy et al., 2015). The

fascinating aspect about feature visualizations is that as they are initialized from an image of random noise, “the result becomes purely the result of the neural network” (Mordvintsev et al., 2015). Despite the fact that the networks are not asked to do anything but classify natural images, the corresponding representations that form inside the networks turn out to be, at times, incredibly rich and complex (OpenVis Conference, 2018).

Computer Vision Models as In Silico Visual System Models

In addition to the hierarchical architecture of both CNNs and the visual system of primates, which in itself seems to play a crucial role in the emerging similarities between the two (Yamins et al., 2014), the most obvious similarity between CNNs and the visual system of primates might be at the structural level. Namely, the structure of modern CNNs was directly inspired by the findings of Hubel and Wiesel (1962), who discovered that the early visual area V1 consists of simple and complex cells (Fukushima, 1980). Simple cells comprise side-to-side organized receptive fields that allow edge detecting, and complex cells combine those to form orientation-sensitive receptive fields invariant to edges’ spatial phase. CNNs generally repeat analogous operations in their full span, mapping the simple cells with the output of the filter operation and the complex cells with that of the pooling operation, the stacking of which creates larger and larger receptive fields for individual neurons, allowing them to detect more complex features in the downstream layers of the networks (Lindsay, 2021; Xu et al., 2021). This tangible similarity gives rise to the resemblance of the features detected by the lower areas of the two systems - as do the simple and complex cells in the primary visual cortex (V1), the artificial neurons in the first layers of CNNs seem to mainly detect the edges of objects, allowed for by filters called Gabors (Olah et al., 2020a). Moreover, single biological and artificial neurons of the earlier areas of the systems under investigation have been shown to be functionally similar, with the ANNs whose artificial neurons are more similar to the biological neurons being more similar to humans at the behavioral level as well (Marques et al., 2021).

As to the connections between the later areas, a biologically plausible ANN’s intermediate and output layers have been shown to be, respectively, predictive of visual hierarchy’s V4 and IT areas’ neuronal activity (Yamins et al., 2014). Additionally, by using the different layers of CNNs as input to predictive models, it has been shown that the lower layers are better predictors of early visual areas’ neural activity, as are the higher layers of that of later visual areas, suggesting that increasingly complex features are encoded in the downstream areas of the ventral

visual pathway (Eickenberg et al., 2017; Güçlü et al., 2015). By showing that the same findings hold across experimental paradigms, Eickenberg et al. claim (2017, p. 185) that CNNs "capture universal representations of the stimuli that linearly map to and separate cognitive processes".

Regardless of these promising results, the scientific value of ANNs as brain models is still debated. Some emphasize the exploratory value that ANNs yield, particularly due to their computational complexity and the absence of a full understanding of any cognitive function to date (Cichy et al., 2019). Others are leaning towards the redundancy of this relationship, pointing out the seemingly fundamental difference in ways that CNNs and brains represent visual information - for example, CNNs are unable to fully capture the representations of artificial stimuli in any level of the visual hierarchy (Xu et al., 2021). This is troublesome due to the same algorithms supporting the processing of artificial and natural stimuli in primates' brains (Xu et al., 2021). On the other hand, Schrimpf et al. (2020) find the time to be ripe to use the suitable ANNs, which they refer as neurally mechanistic models, for building unified models of entire domains of intelligence, and have created an integrative benchmarking platform called Brain-Score, which for now concentrates on the domain of visual intelligence.

A highly relevant study for the current research is that of Suzuki et al. (2017), which used the Deep Dream algorithm (Mordvintsev et al., 2015) - a technique that visualizes the "understanding" that a particular CNN layer has achieved by enhancing the features encoded in the layer - to manipulate the visuals of a virtual reality program. The goal was to elicit a visual experience similar to that of a psychedelic experience, without the co-occurring pharmacological effects. Its results indicate that the subjective effects elicited by the VR experience were similar to those of a psychedelic experience across multiple dimensions, including those of "intensity" and "pattern". It was suggested in the study that the findings "may shed new light on the neural mechanisms underlying physiologically-induced hallucinogenic states" (Suzuki et al., 2017, p. 6). In addition to the above-described parallels between ANNs and visual systems, this suggestion relies on the intuitive similarity between the Deep Dream algorithm and their take of the top-down modulation occurring within the hierarchical predictive coding theory of perception. Namely, if hallucinations are viewed as the result of an excessively strong weighting of perceptual priors (Teufel et al., 2015), then a functional parallel can be drawn with the algorithm. The top-down signals originating from the higher brain areas are roughly mapped with the use of gradients originating from the specific layer of interest in the ANN, and the

resulting hallucinations with the altered input image. Comparing this approach with the one introduced in Carhart-Harris and Friston (2019), where top-down modulation is thought to be decreased under the influence of psychedelics, we run, at first, into an apparent conflict, demonstrating that it is not straightforward to interpret the common ground between psychedelic experiences and ANNs within this framework.

However, considering that the brain can deal with the lessening of predictions' accuracy and the concomitant increased prediction errors in a multitude of ways, one of them being the upregulation of predictions, a rather clear parallel with the feature visualization and Deep Dream techniques can be drawn (Pink-Hashkes et al., 2017). Maximizing the activity of a certain unit of the network by producing an optimal input for the unit (feature visualization) or manipulating the input so to include more of what is already encoded in the unit (Deep Dream) have functional similarities to upregulating the predictions encoded in a specific layer of the brain and enforcing those on the incoming sensory data in order to deal with the increased prediction errors (Pink-Hashkes et al., 2017). All the aforementioned mechanisms can produce visuals with hallucinatory qualities. The nature of those visuals depends on the level of hierarchy which was focused on, for example the feature visualizations of earlier layers' units contain vivid patterns, which are similar to the simulated predictions of humans' primary visual cortex (V1) and geometric hallucinations experienced under the influence of psychedelics (Bressloff et al., 2001).

The current research, hence, continues with the interpretation offered by Carhart-Harris and Friston (2019). Its supporting evidence already discussed in the previous subchapters includes parallels between the lessened top-down modulation and the increased bottom-up information flow with the physiological effects induced by psychedelic substances, such as the visual cortex's decreased alpha power and the increased cerebral blood flow (Carhart-Harris et al., 2016). In addition to that, the therapeutic benefits associated with the use of psychedelics seem to be mediated by the power of these substances to induce, e.g., insight experiences (Letheby, 2021). The latter is accounted for with the weakening of the priors encoded "at the highest or deepest level of the brain's functional architecture, i.e., the levels that instantiate particularly high-level models such as those related to selfhood, identity, or ego" (Carhart-Harris and Friston, 2019, p. 319).

The Current Research

The preceding sections explored the potential of CNNs to explain the mechanisms underlying visual psychedelic experiences. Some of the discussed evidence includes CNNs' architectural and representational similarities to the primate visual system, and the functional parallels between the effects of psychedelics on the brain and the feature visualization techniques employed in CNNs. With that in mind, this research addresses the overarching question:

- How are psychedelic experiences, particularly in terms of their visual aspects, related to the perception of CNN feature visualizations?

With the specific hypothesis under investigation being:

- There is a positive relationship between the subjective intensity of visual experiences during psychedelic use and the level of correspondence with CNN feature visualizations.

This hypothesis relies on a series of interconnected findings: the intensity of visuals during psychedelic experiences correlates with increased intrinsic brain activity, as indicated by measures like visual cortex's activity and connectivity (Carhart-Harris et al., 2016); CNNs can serve as models for the primate visual cortex (Lindsay et al., 2021); and feature visualizations represent the maximal activity of CNNs, with the technique having functional similarities to certain effects of psychedelic substances in the brain that might manifest as visual hallucinations (Olah et al., 2017; Pink-Hashkes et al., 2017).

The hypothesis is put to test in an experiment where participants have to rate the level of correspondence between the feature visualizations and their own visual experiences during their psychedelic experiences. The specifics of the approach are introduced in the following chapter.

Method

33 self-selected participants were recruited via social media. The only eligibility criterion was having at least one prior experience with a serotonergic hallucinogen.

Feature Visualizations

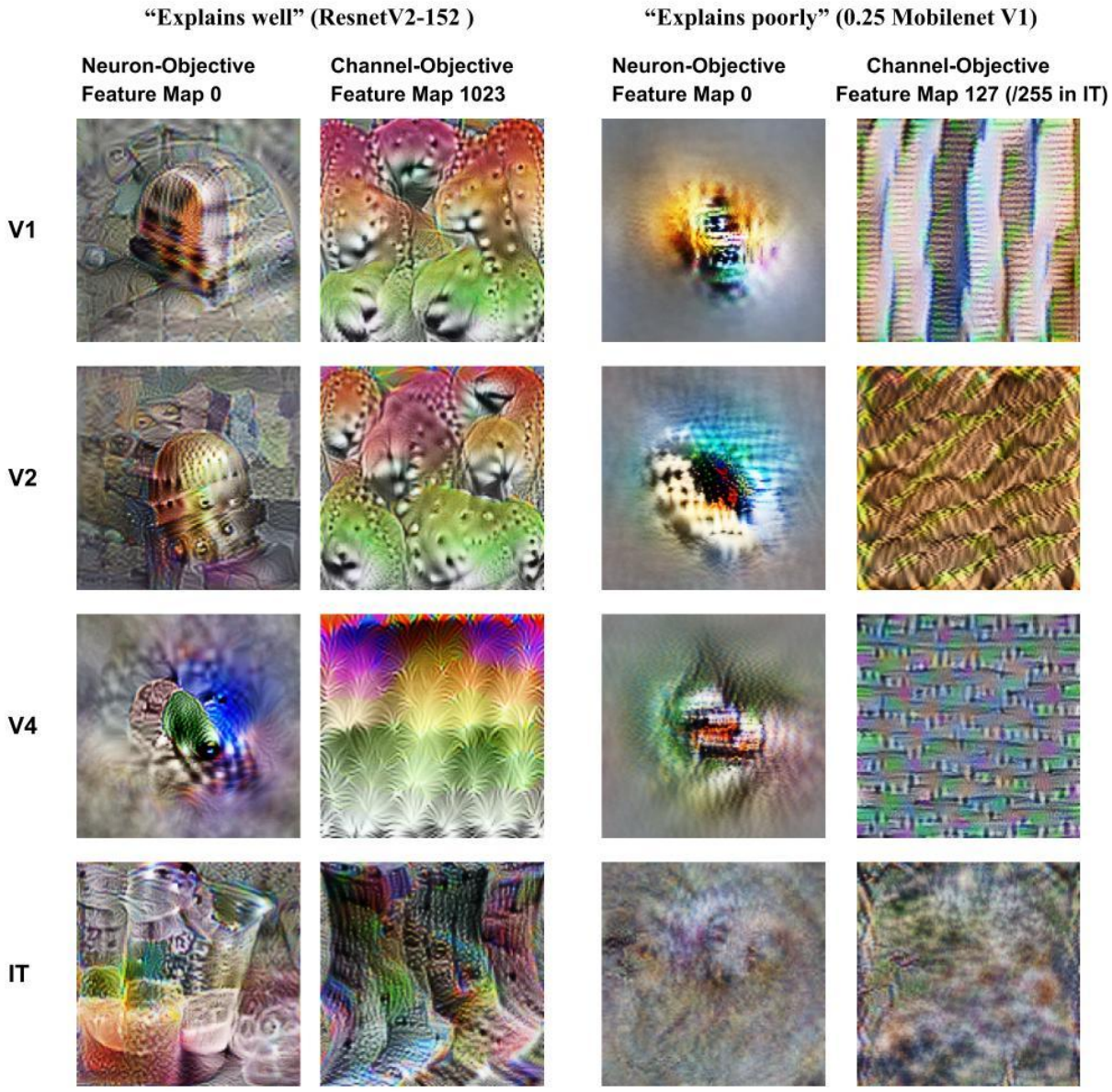
Feature visualizations of two CNNs with contrasting abilities for explaining the responses of the primate visual system were acquired. The two CNNs were identified via the Brain-Score platform[1], where models are scored for their ability to approximate the activity of different parts of the ventral stream as well as human behavior, determined via various benchmarks. The "explains well" brain model was Resnet152 V2 (He et al., 2015), ranked 7th on the Brain-Score leaderboard at the time of the study, with an average vision score of 0.432. The "explains poorly" brain model spanned the 0.25 MobilenetV1 architecture (Howard et al., 2017), ranked between 150th and 168th on the leaderboard at the time of the study, with average vision scores ranging from 0.312 to 0.277. These specific models were chosen because they are implemented in the TensorFlow-Slim image classification model library[2], enabling visualization of their units without the need for separate model training and deployment.

For visualization purposes, only units from the layers that best corresponded to the different brain areas covered by the benchmarks - areas V1, V2, V4, and IT - were selected. The layer mappings provided by the Brain-Score team guided the selection process (see Appendix A). For every chosen layer, four neuron-objective and four channel-objective visualizations were generated, the units across layers being from the same relatively positioned feature maps (Figure 2; see Table B1). The whole layer's visualization was generated with the Deep Dream algorithm (Figure 3). The Lucid library[3] was utilized for visualizing the units, without any additional customizations to the code.

[1] https://www.brain-score.org/
[2] https://github.com/tensorflow/models/tree/master/research/slim
[3] https://github.com/tensorflow/lucid/tree/master

Figure 2

A Selection of the Generated Feature Visualizations



Note. The demonstrated neuron-objective visualizations represent the central unit of the relevant feature map. See Table B1 for detailed information about the visualizations.

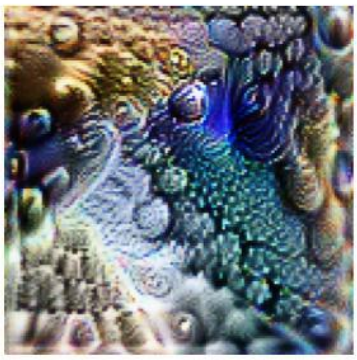
Figure 3

The "Explains well" Model's Layer Visualizations, Obtained with the Deep Dream Algorithm

V1 - resnet_v2_152/block3/unit_32/bottleneck_v2

V2 - resnet_v2_152/block3/unit_30/bottleneck_v2

V4 - resnet_v2_152/block3/unit_3/bottleneck_v2



IT - resnet_v2_152/block4/unit_1/bottleneck_v2

Note. The brain region-CNN layer mapping is brought out under the visualization (see Appendix A for information about the "explains poorly" model).

Questionnaire

The visualizations were utilized to create a two-part questionnaire (https://forms.gle/NMk938QDmLsrwDYR6). In the first part, participants were asked to rate each visualization on a scale from 1 to 10, reflecting on the extent to which it corresponded to their own experiences with visualizations induced by psychedelic substances, the specific question asked being "Based on your previous experiences, how possible is it to experience visuals with similar properties under the influence of psychedelics?". It is important to note that participants were not expected to find a perfect match between the visualizations and their own experiences, but rather reflect on an overall impression of correspondence. Participants were

guided to assess the correspondence across various properties, such as colors, contrasts, edges, lines, and entire objects. After rating all the visualizations, participants were asked to indicate whether they based their ratings on a single psychedelic experience or multiple experiences. Additionally, they were asked to specify whether their ratings primarily relied on open-eye visualizations, closed-eye visualizations, or both equally.

In the second part of the experiment, participants were prompted to provide information about one of their previous psychedelic experiences. They were instructed to choose the most intense experience from the ones they reflected upon in the first part. The intensity of the experience was assessed by participants themselves, instructions guiding them to think about the dosage of the substance consumed and the intensity of the experience's visual aspects. Several questions were asked about the chosen experience, including the substance, the approximate dosage, and its overall intensity.

Furthermore, participants were asked to rate the vividness, complexity, richness, and immersiveness of the visuals during the experience. These dimensions were selected from the subjective experiences questionnaire employed by Timmermann et al. (2023). Questions regarding the prominence of different properties were also included, drawing from the typical feature families of CNNs (Olah et al., 2020b), as well as the qualitative findings of Lüübek (2023), where participants listed the features or qualities of the visualizations that caught their attention the most.

Analysis

One participant was excluded from the analysis due to inappropriate substance selection and another due to invalid ratings, leaving 31 participants' results for the analysis. For most of the analysis, the mean ratings per participant across the feature visualizations were used, with different groupings of visualizations created for specific purposes. The mean ratings of the participants were opted for because of the absence of a direct correspondence between the two models' visualizations. The normality of the data distributions was assessed using D'Agostino and Pearson's normality test, with the results affirming the normality of the participant-based mean ratings of the two models. Before hypothesis testing, the difference between the two models' ratings was assessed with a paired *t* test, to evaluate the meaningfulness of using both models' visualizations' ratings for the subsequent analysis.

Ethics

Ethical standards were upheld throughout the study, including thorough and accurate reporting of hypotheses and methods, valid data analysis methods, honest reporting of results, avoiding selective exclusion of opposing results, protecting participant confidentiality by not collecting any personally identifiable information, and adhering to standards of plagiarism and self-plagiarism. Generative AI was used for the wording revision of several paragraphs, with the specific prompt being "How to make this better: [*text*]?", with the answer thoroughly reviewed and evaluated before making use of (parts of) the rephrased text (OpenAI, 2023).

Results

First, we used two CNN models, one of which explained the primate visual activity well (ResnetV2-152) and the other which explained it poorly (0.25 MobilenetV1), and we studied the differences in ratings of how well their feature visualizations correspond to psychedelic visualizations. The results of the paired *t* test indicated a significant difference in mean ratings between the two models ($t(30) = 3.03$, $p = .005$). Further analysis revealed that participants generally rated the visualizations of the "explains well" model ($M = 5.3$, $SD = 1.9$) slightly higher than those of the "explains poorly" model ($M = 4.5$, $SD = 1.6$), suggesting the former's relatively better correspondence to psychedelic visualizations (Figure 4, 5).

Figure 4

The Mean Ratings of the Two Models' Visualizations across Participants

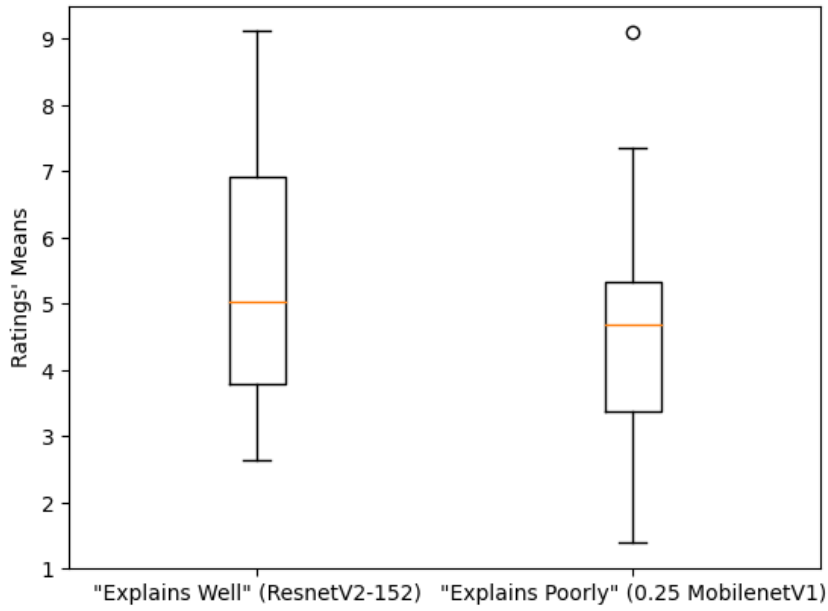
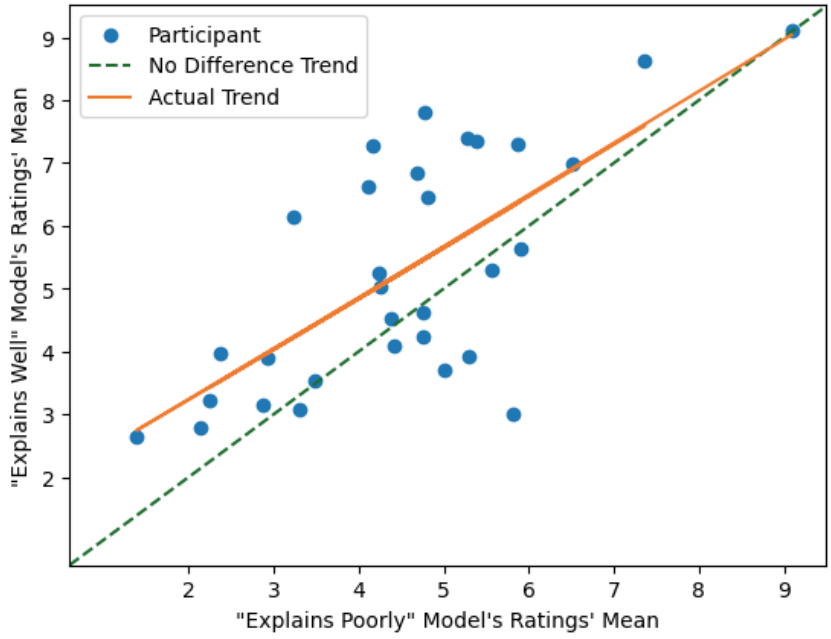


Figure 5

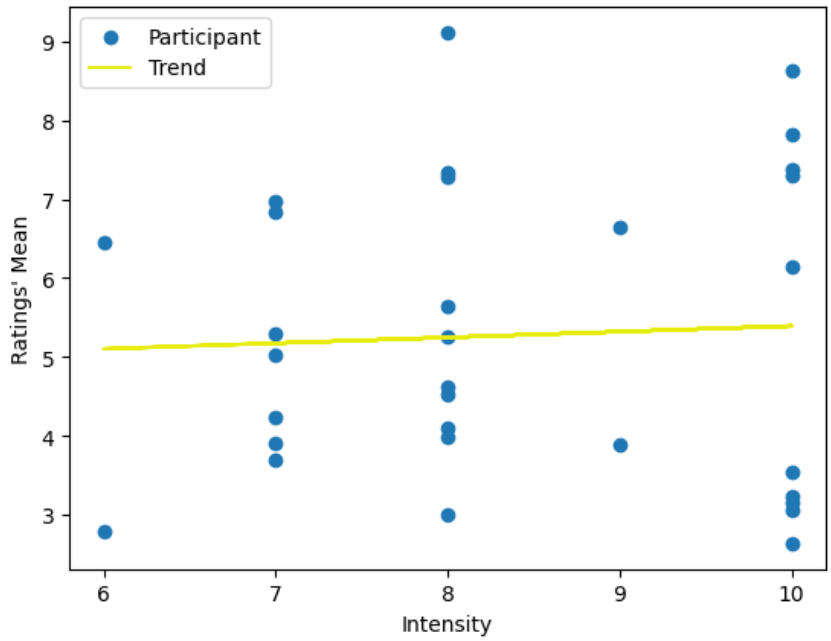
The Relationship Between the Mean Ratings of Both Models per Participant



Therefore, in the next step, to study the relationship between the subjective measures of psychedelic visual experiences and the experiences' level of correspondence with CNN feature visualizations, only the data on the "explains well" model was used. To test the hypothesis that more intense prior experiences lead to higher correspondence with the feature visualizations, the Spearman rank-order correlation coefficient was used to first measure the relationship between the reported overall intensity of a previous psychedelic experience and the mean ratings of the "explains well" model's visualizations. The results indicated that there is no correlation between the two (r (29) = .03, p = .888) (Figure 6).

Figure 6

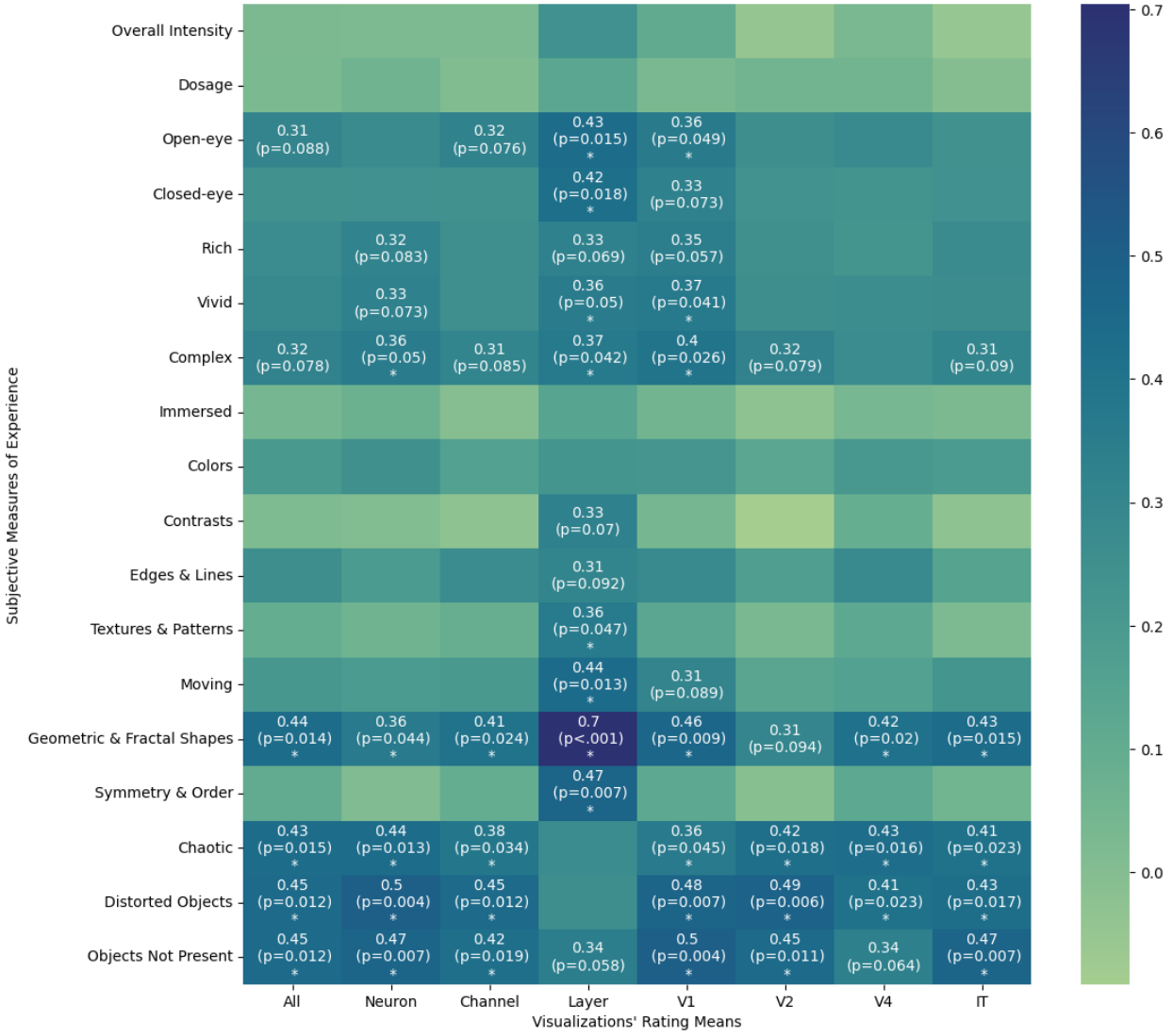
The Relationship Between the Overall Intensity of a Prior Psychedelic Experience and the Mean Ratings of the "Explains Well" Model's Visualizations



As the individual visualizations differed notably in their properties, additional groupings were made before measuring their relationship with the different subjective measures of a prior psychedelic visual experience. Visualizations were grouped based on their optimization objective and their layer of origin. Moderate to high correlations were found between a number of measures and the visualizations' ratings (Figure 7).

Figure 7

Spearman Correlations Between Subjective Measures of a Psychedelic Visual Experience and Visualizations' Ratings



Note. Only moderate and strong correlations ($r > .3$) are annotated. '*' indicates a p value of <= .05.

Grouping the visualizations by their optimization objective yielded three groups - neuron-objective, channel-objective and layer (Deep Dream) visualizations - from which the most positive correlations with subjective measures were observed in the case of layer

visualizations (Figure 3). Significant moderate to high correlations were found between the mean ratings of the correspondence of the layer visualizations to psychedelic visualizations and the extent of open-eye visualizations ($r(29) = .43$, $p = .015$), closed-eye visualizations ($r(29) = .42$, $p = .018$), the vividness ($r(29) = .36$, $p = .05$) and complexity ($r(29) = .37$, $p = .042$) of the visualizations, the prominence of textures or patterns ($r(29) = .36$, $p = .047$), geometric or fractal shapes ($r(29) = .7$, $p < .001$), and symmetry or order ($r(29) = .47$, $p = .007$) in the visualizations, and the extent to which the visuals appeared to move ($r(29) = .44$, $p = .013$) during a prior psychedelic experience (Figure 7).

Neuron- and channel-objective visualizations' (Figure 2) ratings saw significant moderate to high correlations with the prominence of geometric or fractal shapes as well ($r(29) = .36$, $p = .044$; $r(29) = .41$, $p = .024$), but contrary to the layer visualizations' ratings, significant correlation was also observed with the appearance of distorted objects ($r(29) = .50$, $p = .004$; $r(29) = .45$, $p = .012$) and objects not actually present ($r(29) = .47$, $p = .007$; $r(29) = .42$, $p = .019$), and the chaotic quality of the psychedelic visualizations ($r(29) = .44$, $p = .013$; $r(29) = .38$, $p = .034$) (Figure 7).

As for the visualizations' grouping based on their layer of origin, which yielded four groups - visualizations from the CNN's layers best corresponding to the ventral stream's V1, V2, V4 or IT - the most positive correlations with subjective measures were observed in the case of the V1 group, although all the groups had significant moderate to large correlations with the appearance of distorted objects and the chaotic quality of the psychedelic visualizations. In addition to the former, the V1 group's visualizations saw significant correlations with the extent of open-eye visualizations ($r(29) = .36$, $p = .049$), and the vividness ($r(29) = .37$, $p = .041$) and complexity ($r(29) = .40$, $p = .026$) of psychedelic visualizations. The V1, V4 and IT groups' visualizations had a significant correlation with the prominence of geometric or fractal shapes in psychedelic visualizations ($r(29) = .46$, $p = .009$; $r(29) = .42$, $p = .02$; $r(29) = .43$, $p = .015$), and V1, V2 and IT groups with the appearance of objects not actually present in the visualizations ($r(29) = .50$, $p = .004$; $r(29) = .45$, $p = .011$; $r(29) = .47$, $p = .007$) (Figure 7).

Discussion

The current research studied the relationship between psychedelic visual experiences and CNNs' feature visualizations with the question of how the nature of psychedelic visual experiences relates to the perceived correspondence of feature visualizations to psychedelic visualizations. The theoretical framework led us to hypothesize that the correspondence is higher in the case of more intense psychedelic visualizations, as feature visualizations represent the maximal or optimal activity of CNNs' units, and CNNs can be used to explain neural activity (Erhan et al., 2009; Olah et al., 2017; Lindsay, 2021).

The feature visualizations of a CNN that explains the primate visual activity better corresponded better to psychedelic visualizations than those of a CNN that explains the visual activity worse (Figure 4, 5). This suggests that feature visualization can capture certain aspects of psychedelic vision, as the visualizations of the inner activity of CNNs seem to get more psychedelic the better the models can account for brain activity. This is supported by a number of parallels between the two systems. After psychedelics' intake, the activity of the visual cortex, as well as its connectivity with higher brain areas, increases (Carhart-Harris et al., 2016; Timmermann et al., 2023). The visual cortex seems to process inputs from intrinsic brain areas, as illustrated by the heightened connectivity between the visual cortex and brain's mnemonic circuits during a closed-eye imagery task following psychedelics' intake (de Araújo et al., 2012). Feature visualizations are created by forcing the representations encoded in the CNN units onto an input of pure noise, which creates an input that drives the unit to fire maximally (Erhan et al., 2009; Olah et al., 2017; Lindsay, 2021). Seeing a relationship between psychedelic perception and the optimal CNN inputs suggests that there could be potential for further research to unravel some still unknown mechanistic or computational aspects of the visual system with the help of CNNs.

The absent correlation between the overall intensity of psychedelic experiences and the feature visualizations' ratings could be attributed to the specific question in the questionnaire about the intensity of the prior experience being too general ("How intense was that experience? Please rate the overall intensity of the experience, including any visual, auditory, or emotional effects that you may have experienced."). The moderate to high correlations between the numerous more specific subjective experience measures and the feature visualizations' ratings suggest that the intensity of psychedelic visualizations could be better captured by the more tailored questions, instead of that directly about the intensity of the psychedelic experience. It could as

well suggest that the psychedelic experience's intensity does not necessarily bring along psychedelic visualizations, with the latter further depending on, e.g., the set and setting of the experience.

The correlations between specific subjective experience measures and psychedelic experiences cannot be taken at face value, as the correlations were not corrected for multiple comparisons. However, it is still informative to look into the results in an exploratory manner.

An interesting observation is that solely the layer (Deep Dream) visualizations significantly correlate with the extent of closed-eye visualizations, the prominence of textures or patterns and symmetry or order in the psychedelic visualizations, as well as having the strongest correlation with the prominence of fractal and geometric shapes among all the other visualization groupings (Figure 7). This could in part be due to the CNN layers combining information from all of their feature maps, yielding a more complete, overarching "understanding" of an input than that of an artificial neuron detecting only a single feature. This, rather clearly, suggests that to mimic the more "established" or pronounced psychedelic visualizations, such as those containing emphasized patterns, geometric shapes, or symmetry, an encoding of combined information, such as that of a CNN layer, is needed. Fittingly, out of all the visualization groupings, the layer visualizations correlate the least with the chaotic or distorted properties of psychedelic visualizations, whereas the neuron-objective visualizations have the strongest positive correlations with them. This, again rather clearly, suggests that for mimicking less pronounced or realistic visualizations, the encoding of individual features could suffice, such as that of a single artificial neuron. In essence, this finding could as well imply that the dysregulation of a combination of neurons is needed for the perception of more pronounced visualizations, whereas the dysregulation of lesser neurons could suffice for the mere distortion of perception. These implications rely on the functional similarities between artificial and biological neurons, as well as CNNs' layers ability to explain neural activity (Marques et al., 2021; Yamins et al., 2014). For any definitive implications, though, further research is needed, to study the possible causal relationships behind the observed correlations.

No significant differences were observed among the correlations between the subjective measures and the ratings of the feature visualizations separated by their layer of origin. The chosen CNN layers were ones corresponding best to several ventral visual hierarchy's regions. The absence of differences suggests that the representations in the examined layers may not vary

significantly, leading to similar relationships with the subjective measures. The similarity of representations is illustrated by the appearance of the feature visualizations (see Table B1) as well as the layer mappings (see Appendix A). For instance, in the case of the CNN layer that best corresponds to area V4, the corresponding layer was a lower CNN layer compared to those best corresponding to areas V1 and V2. This seems odd at first, as previous studies suggest that higher CNN layers explain the neural activity of later visual system areas better (Eickenberg et al., 2017; Güçlü et al., 2015). Some possible explanations could be that the current layer mapping is done poorly, with a better one consequently leading to a better brain model, or that the representations in the different ventral visual hierarchy's regions are just not that different from each other (M. Schrimpf, personal communication, 14.05.2023). Interestingly, even though a higher CNN layer best corresponded to IT, a later visual area, the relationship between the layer's visualizations and the subjective measures was nonetheless not significantly different from that of the other CNN layers.

In summary, the results indicate that for measuring the intensity of psychedelic visualizations, the overall intensity nor dosage of the psychedelic experience is not enough. More direct measures, such as the complexity or richness of the visualizations, as well as the extent of open-eye or closed-eye visualizations experienced, are more fitting. These measures also moderately correlate with the perception of CNNs' feature visualizations as corresponding to psychedelic visualizations. There is a stronger correlation between the presence of relatively higher-level properties in participants' psychedelic visual experiences and their perception of the feature visualizations' correspondence with the psychedelic visualizations. These higher-level properties include, for example, geometric and fractal shapes, distorted objects, and objects not actually present. The prominence of lower-level properties in the psychedelic visualizations, such as colors, contrasts, and edges, did not seem to suffice for perceiving similarities with the feature visualizations. This could suggest that the higher-level, unrealistic, or chaotic properties of the feature visualizations may have a stronger impact on the assessment, overriding the perception of the lower-level properties.

Future research could adopt an approach similar to Suzuki et al. (2017), using the Deep Dream algorithm or the preferred features of different CNN units to manipulate real-world images instead of pure noise. This approach would leverage the rich and diverse representations offered by CNNs while maintaining the realism of visualizations.

References

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron, 76*(4), 695-711. https://doi.org/10.1016/j.neuron.2012.10.038

Beliveau, V., Ganz, M., Feng, L., Ozenne, B., Højgaard, L., Fisher, P. M., Svarer, C., Greve, D. N., & Knudsen, G. M. (2017). A High-Resolution *In Vivo* Atlas of the Human Brain's Serotonin System. *Journal of Neuroscience*, *37*(1), 120–128. https://doi.org/10.1523/JNEUROSCI.2830-16.2017

Braga, R. M., & Leech, R. (2015). Echoes of the Brain: Local-Scale Representation of Whole-Brain Functional Networks within Transmodal Cortex. *Neuroscientist*, *21*(5), 540-551. https://doi.org/10.1177/1073858415585730

Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2001). Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society of London*, *356*(1407), 299–330. http://doi.org/10.1098/rstb.2000.0769

Butler, T. C., Benayoun, M., Wallace, E., van Drongelen, W., Goldenfeld, N., & Cowan, J. (2012). Evolutionary constraints on visual cortex architecture from the dynamics of hallucinations. *Proceedings of the National Academy of Sciences*, *109*(2), 606-609. https://doi.org/10.1073/pnas.1118672109

Carhart-Harris, R. L., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., Tagliazucchi, E., Schenberg, E. E., Nest, T., Orban, C., Leech, R., Williams, L. T., Williams, T. M., Bolstridge, M., Sessa, B., McGonigle, J., Sereno, M. I., Nichols, D., Hellyer, P. J., Hobden, P., Evans, J., Singh, K. D., Wise, R. G., Curran, H. V., Feilding, A., & Nutt, D. J. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences*, *113*(17), 4853-4858. https://doi.org/10.1073/pnas.1518377113

Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics. *Pharmacological Reviews, 71*(3), 316-344. https://doi.org/10.1124/pr.118.017160

Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences, 23*(4), 305-317. https://doi.org/10.1016/j.tics.2019.01.009

de Araújo, D. B., Baffa, O., & Wakai, R. T.. (2002). Theta Oscillations and Human Navigation: A Magnetoencephalography Study. *Journal of Cognitive Neuroscience*, *14*(1), 70-78. https://doi.org/10.1162/089892902317205339

de Araújo, D. B., Ribeiro, S., Cecchi, G. A., Carvalho, F. M., Sanchez, T. A., Pinto, J. P., de Martinis, B. S., Crippa, J. A., Hallak, J. E. C., & Santos, A. C. (2012). Seeing With the Eyes Shut: Neural Basis of Enhanced Imagery Following Ayahuasca Ingestion. *Human Brain Mapping*, *33*(11), 2550-2560. https://doi.org/10.1002/hbm.21381

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184-194. https://doi.org/10.1016/j.neuroimage.2016.10.001

Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). *Visualizing Higher-Layer Features of a Deep Network* (Report No. 1341). Univeristé de Montréal. https://www.researchgate.net/profile/Aaron-Courville/publication/265022827_Visualizing_Higher-Layer_Features_of_a_Deep_Network/links/53ff82b00cf24c81027da530/Visualizing-Higher-Layer-Features-of-a-Deep-Network.pdf

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202. https://doi.org/10.1007/BF00344251

Güçlü U., & van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience, 35*(27), 10005-10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. arXiv, Article 1512.03385. https://doi.org/10.48550/arXiv.1512.03385

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv, Article 1704.04861. https://doi.org/10.48550/arXiv.1704.04861

Hubel, D.H., & Wiesel T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition . *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00186

Letheby, C. (2021). *Philosophy of Psychedelics*. Oxford University Press. https://doi.org/10.1093/med/9780198843122.001.0001

Lindsay, G. W. (2021). Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *Journal of Cognitive Neuroscience*, *33*(10), 2017-2031. https://doi.org/10.1162/jocn_a_01544

Lüübek, C. (2023, January 16). The Psychedelicism of Feature Visualizations in InceptionV1. *Medium.* https://medium.com/@lyybek.carolin/the-psychedelicism-of-feature-visualizations-in-inceptionv1-9e82fcba6c9b

Marques, T., Schrimpf, M., & DiCarlo, J. J. (2021). *Multi-scale hierarchical neural network models that bridge from single neurons in the primate primary visual cortex to object recognition behavior*. bioRxiv. https://doi.org/10.1101/2021.03.01.433495

Mordvintsev, A., Olah, C. & Tyka, M. (2015, June 18). Inceptionism: Going Deeper into Neural Networks. *Google Research Blog*. https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020a, April 1). An Overview of Early Vision in InceptionV1. *Distill*. https://distill.pub/2020/circuits/early-vision/

Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020b, March 10). Zoom In: An Introduction to Circuits. *Distill*. https://distill.pub/2020/circuits/zoom-in/

Olah C., Mordvintsev A., & Schubert L. (2017, November 7). Feature Visualization. *Distill*. https://distill.pub/2017/feature-visualization/

OpenVis Conference. (2018, July 31). *SHAN CARTER - OpenVisConf 2018* [Video]. Youtube. https://www.youtube.com/watch?v=jlZsgUZaIyY

Passie, T., Halpern, J.H., Stichtenoth, D.O., Emrich, H.M. & Hintzen, A. (2008), The Pharmacology of Lysergic Acid Diethylamide: A Review. *CNS Neuroscience & Therapeutics*, *14*(4), 295-314. https://doi.org/10.1111/j.1755-5949.2008.00059.x

Pink-Hashkes. S., van Rooij, I. J. E. I., & Kwisthout, J. H. P. (2017). Perception is in the details: a predictive coding account of the psychedelic phenomenon. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society, United Kingdom*, 2907–2912. https://www.noisebridge.net/images/e/ef/Perception_is_in_the_Details12.pdf

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*, 79-87. https://doi.org/10.1038/4580

Schrimpf, M., Kubilius, J., Lee, M. J., Apurva Ratan Murty, N., Ajemian, R., & DiCarlo, J. J. (2020). Integrative Benchmarking to Advance Neurally Mechanistic Models of Human Intelligence. *Neuron, 108*(3), 413-423. https://doi.org/10.1016/j.neuron.2020.07.040

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA*, 1-9. https://ieeexplore.ieee.org/document/7298594

Suzuki, K., Roseboom, W., Schwartzman, D.J., & Seth, A. K. (2017). A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. *Scientific Reports*, *7*, Article 15982. https://doi.org/10.1038/s41598-017-16316-2

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyear, I. M., & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, *112*(43), 13401-13406. https://doi.org/10.1073/pnas.1503916112

Timmermann, C., Roseman, L., Haridas, S., Rosas, F. E., Luan, L., Kettner, H., Martell, J., Erritzoe, D., Tagliazucchi, E., Pallavicini, C., Girn, M., Alamia, A., Leech, R., Nutt, D. J., & Carhart-Harris, R. L. (2023). Human brain effects of DMT assessed via EEG-fMRI. *Proceedings of the National Academy of Sciences*, *120*(13). https://doi.org/10.1073/pnas.2218949120

Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, *12*, Article 2065. https://doi.org/10.1038/s41467-021-22244-7

Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., & DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619-8624. https://doi.org/10.1073/pnas.1403112111

Appendix A

Fragment from the Layer Mappings File Provided by the Brain-Score Team


```
{
  "model": "benchmarks.ModelMeta",
  "fields": {
    "model": "resnet-152_v2",
    "key": "IT_layer",
    "value": "resnet_v2_152/block4/unit_1/bottleneck_v2"
  }
},
{
  "model": "benchmarks.ModelMeta",
  "fields": {
    "model": "resnet-152_v2",
    "key": "V1_layer",
    "value": "resnet_v2_152/block3/unit_32/bottleneck_v2"
  }
},
{
  "model": "benchmarks.ModelMeta",
  "fields": {
    "model": "resnet-152_v2",
    "key": "V4_layer",
    "value": "resnet_v2_152/block3/unit_3/bottleneck_v2"
  }
},
{
  "model": "benchmarks.ModelMeta",
  "fields": {
    "model": "resnet-152_v2",
    "key": "V2_layer",
    "value": "resnet_v2_152/block3/unit_30/bottleneck_v2"
  }
}
{
```

```
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_224",
      "key": "IT_layer",
      "value": "Conv2d_13_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_224",
      "key": "V1_layer",
      "value": "Conv2d_7_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_224",
      "key": "V4_layer",
      "value": "Conv2d_7_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_224",
      "key": "V2_layer",
      "value": "Conv2d_6_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_192",
      "key": "IT_layer",
```

```
      "value": "Conv2d_13_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_192",
      "key": "V1_layer",
      "value": "Conv2d_6_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_192",
      "key": "V4_layer",
      "value": "Conv2d_8_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_192",
      "key": "V2_layer",
      "value": "Conv2d_8_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_160",
      "key": "IT_layer",
      "value": "Conv2d_12_depthwise"
    }
  },
  {
```

```
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_160",
      "key": "V1_layer",
      "value": "Conv2d_6_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_160",
      "key": "V4_layer",
      "value": "Conv2d_7_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_160",
      "key": "V2_layer",
      "value": "Conv2d_6_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_128",
      "key": "IT_layer",
      "value": "Conv2d_12_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_128",
      "key": "V1_layer",
```

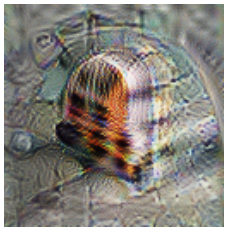
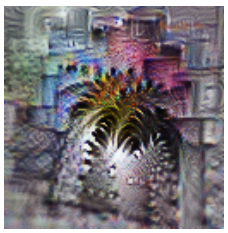
```
      "value": "Conv2d_7_pointwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_128",
      "key": "V4_layer",
      "value": "Conv2d_5_depthwise"
    }
  },
  {
    "model": "benchmarks.ModelMeta",
    "fields": {
      "model": "mobilenet_v1_0.25_128",
      "key": "V2_layer",
      "value": "Conv2d_7_depthwise"
    }
  }
```

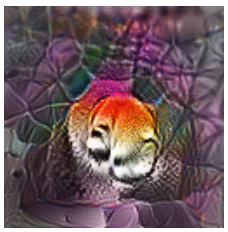
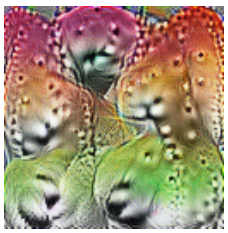
Appendix B

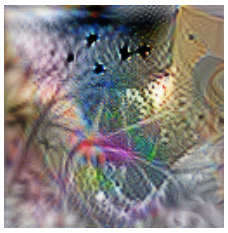
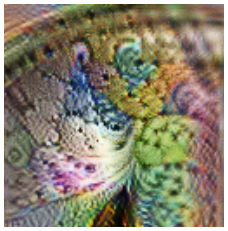
Feature Visualizations' Information

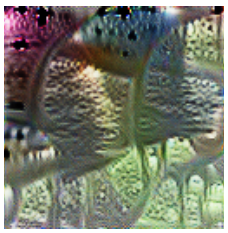
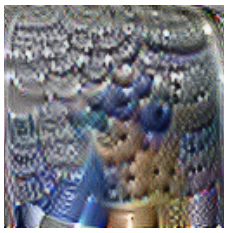
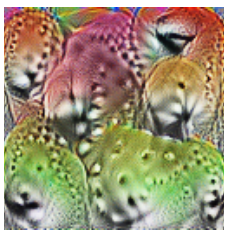
Table B1

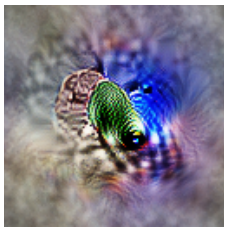
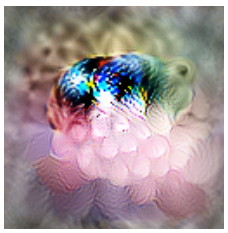
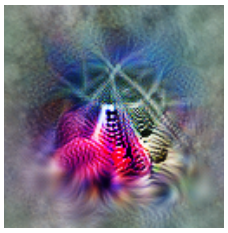
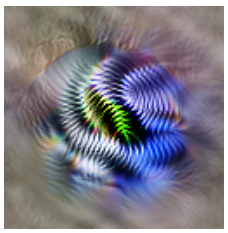
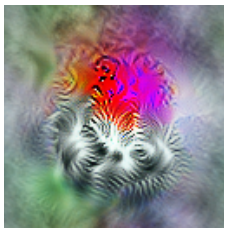
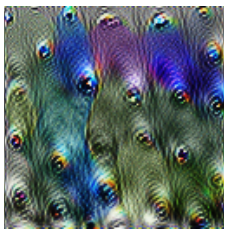
Detailed Information about the Feature Visualizations Included in the Questionnaire

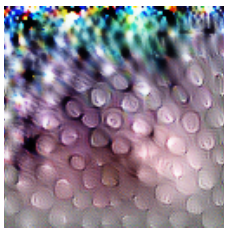
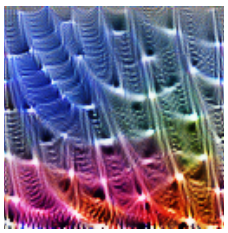
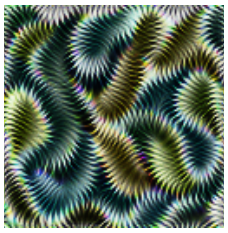
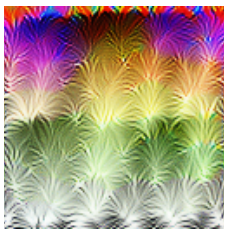
Model (*as implemented in the TensorFlow Slim library (https://github.com/tensorflow/models/tree/master/research/slim))	**Layer**	**Objective**	**Unit**	**Visualization**
Resnet152-V2*	resnet_v2_152/block3/unit_32/bottleneck_v2/add	neuron	0	
Resnet152-V2*	resnet_v2_152/block3/unit_32/bottleneck_v2/add	neuron	256	
Resnet152-V2*	resnet_v2_152/block3/unit_32/bottleneck_v2/add	neuron	512	
Resnet152-V2*	resnet_v2_152/block3/unit_32/bottleneck_v2/add	neuron	768	

Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	1023	
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	0	
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	256	
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	512	
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	768	
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	1023	

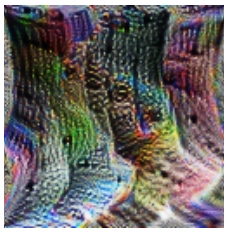
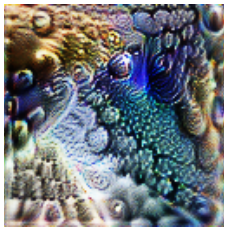
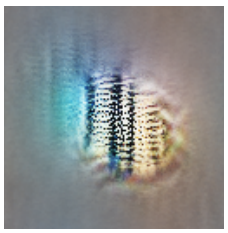
Resnet152-V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	layer	-	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	0	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	256	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	512	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	768	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	1023	

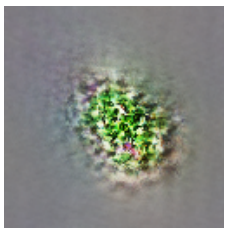
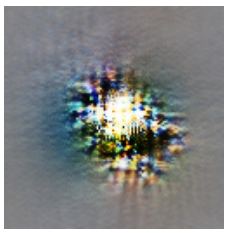
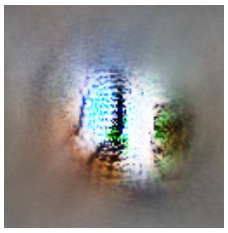
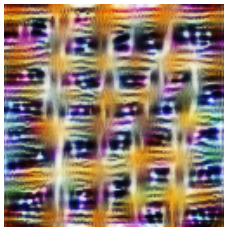
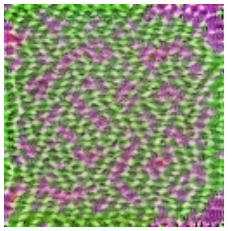
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	0	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	256	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	512	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	768	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	1023	
Resnet152-V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	layer	-	

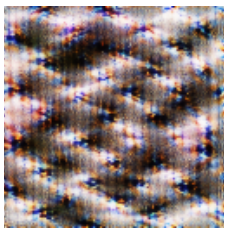
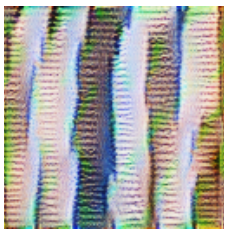
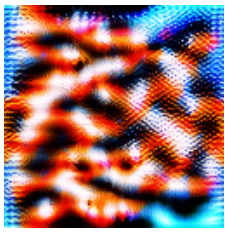
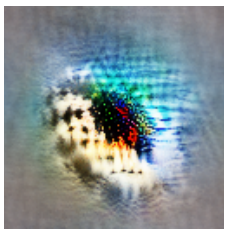
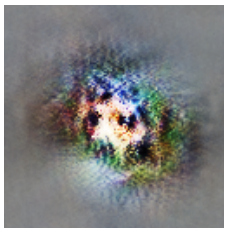
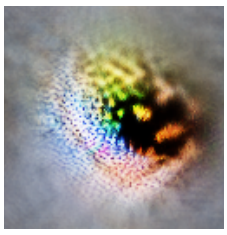
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	neuron	0	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	neuron	256	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	neuron	512	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	neuron	768	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	neuron	1023	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	channel	0	

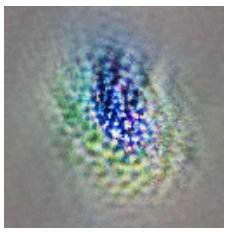
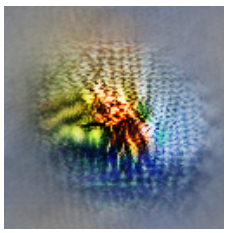
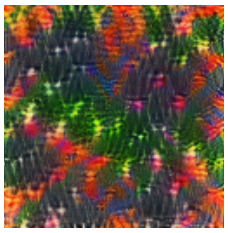
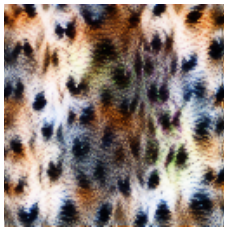
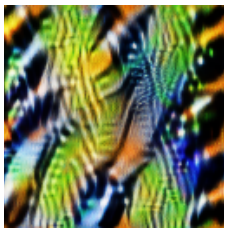
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	channel	256	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	channel	512	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	channel	768	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	channel	1023	
Resnet152-V2*	resnet_v2_152/block3/unit_3/bottleneck_v2/add	layer	-	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	neuron	0	

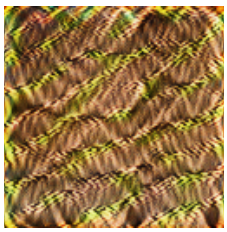
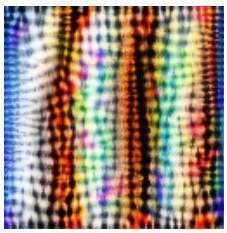
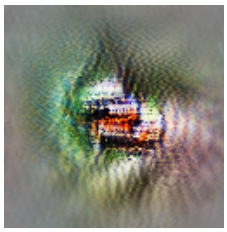
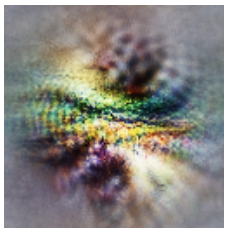
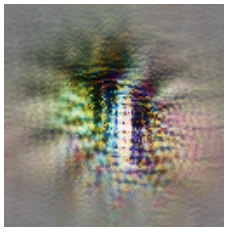
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	neuron	256	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	neuron	512	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	neuron	768	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	neuron	1023	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	channel	0	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	channel	256	

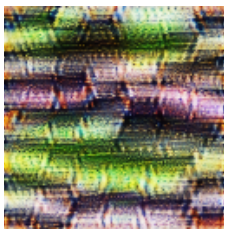
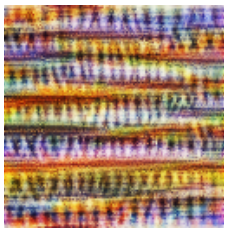
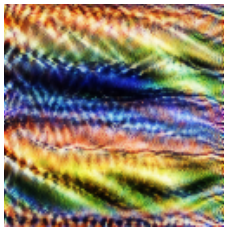
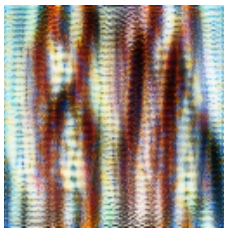
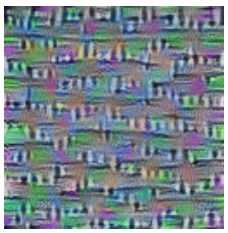
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	channel	512	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	channel	768	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	channel	1023	
Resnet152-V2*	resnet_v2_152/block4/unit_1/bottleneck_v2/add	layer	-	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	neuron	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	neuron	32	

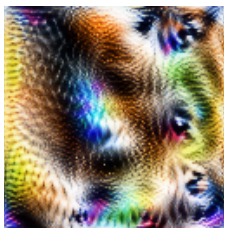
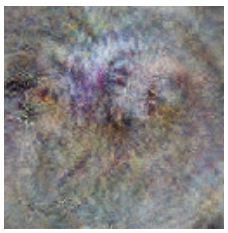
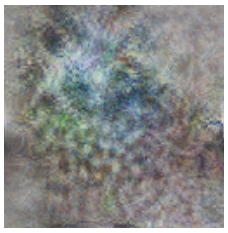
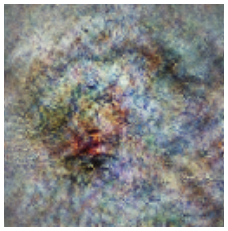
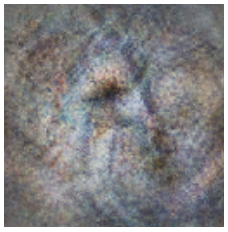
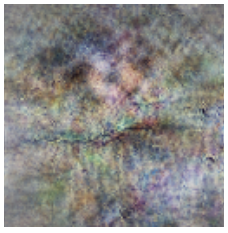
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	neuron	64	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	neuron	96	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	neuron	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	channel	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	channel	32	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	channel	64	

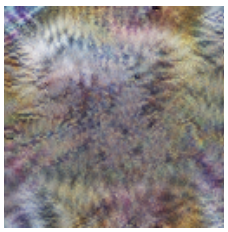
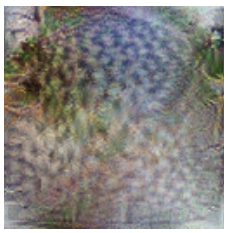
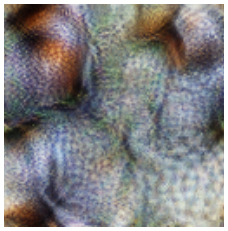
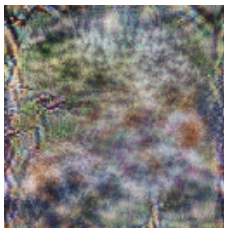
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	channel	96	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	channel	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_6_pointwise/Relu6	layer	-	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	neuron	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	neuron	32	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	neuron	64	

25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	neuron	96	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	neuron	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	channel	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	channel	32	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	channel	64	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	channel	96	

25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	channel	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_7_pointwise/Relu6	layer	-	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	neuron	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	neuron	256	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	neuron	512	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	neuron	768	

25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	neuron	1023	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	channel	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	channel	256	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	channel	512	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	channel	768	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	channel	1023	

25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_8_pointwise/Relu6	layer	-	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	neuron	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	neuron	64	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	neuron	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	neuron	192	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	neuron	256	

25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	channel	0	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	channel	64	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	channel	128	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	channel	192	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	channel	256	
25 MobilenetV1	MobilenetV1/MobilenetV1/Conv2d_13_pointwise/Relu6	layer	-	

Note. Feature visualizations are generated with the Lucid library (https://github.com/tensorflow/lucid).

Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.
Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis DSpace.
Carolin Lüübek