

UNIVERSITY OF TARTU  
Institute of Computer Science  
Computer Science Curriculum

**Carolín Lüübek**

**Investigating Psychedelic Imagery through  
Convolutional Neural Networks' Feature Visualization**

**Bachelor's Thesis (9 ECTS)**

Supervisor: Jaan Aru

Tartu 2023

# **Investigating Psychedelic Imagery through Convolutional Neural Networks'**

## **Feature Visualization**

### **Abstract:**

Convolutional neural networks serve as primate visual system models and their feature visualization has intuitive similarities to psychedelic imagery, as well as the effect of features being similar in both artificial and biological systems. It is explored in this thesis whether the "psychedelicism" inherent in feature visualizations might signify deeper correspondences between convolutional neural networks and visual processing, particularly within the realm of psychedelic imagery.

The main research question of the thesis examines the variance in simulating psychedelic imagery between effective and ineffective visual system models. The results clarify that the feature visualizations of the CNN that approximates the visual system effectively are more accurate. This is speculated to point towards some of the computational mechanisms of effective visual system models being suitable for explaining the neural mechanisms underlying psychedelic imagery. Furthermore, the results hint at the heightened influence of endogenous activity from the primary visual area during psychedelic perception, as well as a pronounced alignment between artificial and biological systems at the neuronal level for early processing stages and at a more abstract level for later processing stages.

### **Keywords:**

Convolutional neural networks, feature visualization, primate visual system, psychedelic imagery

**CERCS:** P170, S260

# **Psühhedeelsete nägemuste uurimine konvolutsiooniliste närvivõrkude üksuste visualiseerimise kaudu**

## **Lühikokkuvõte:**

Konvolutsioonilised närvivõrgud mudeldavad primaatide nägemissüsteemi ning nende üksuste visualiseerimisel on sarnaseid jooni psühhedeelsete nägemustega, kusjuures visualisatsioonide mõju on analoogne nii tehislikes kui bioloogilistes nägemissüsteemides. Käesolevas bakalaureusetöös uuritakse, kas konvolutsiooniliste närvivõrkude üksuste visualisatsioonide psühhedeelsus võib tehnilike ja bioloogiliste nägemissüsteemide vahelisele sügavamale sarnasusele viidata.

Bakalaureusetöö põhiline uurimisküsimus on, kas efektiivsete ja ebaefektiivsete nägemissüsteemi mudelite üksuste visualisatsioonide psühhedeelsuse vahel on erinevus. Tulemuste põhjal on efektiivsema nägemissüsteemi mudeli visualisatsioonid psühhedeelsete nägemustega sarnasemad, mida loetakse sellele viitavat, et efektiivsemate mudelite arvutuslikud mehhanismid võivad psühhedeelsete nägemuste tekkemehhanisme selgitada. Lisaks osutavad tulemused sellele, et psühhedeelsete nägemuste kujunemisel mängib rolli esmase nägemisala aktiivsus, ning tehnilike ja bioloogiliste nägemissüsteemide vaheline analoogia on varasemas töötlustapis tugevaim neuronite tasandil, kuid hilisemas töötlustapis kõrgemal hierarhiatasandil.

## **Võtmesõnad:**

Konvolutsioonilised närvivõrgud, üksuste visualiseerimine, primaatide nägemissüsteem, psühhedeelsed nägemused

**CERCS:** P170, S260

## Table of Contents

Introduction.....	5
1 Background.....	7
1.1 CNNs as Primate Visual System Models.....	7
1.2 Feature Visualization.....	9
1.3 The Neural Mechanisms of Psychedelic Imagery.....	13
2 Method.....	18
2.1 Models.....	18
2.1.1 Brain Score.....	18
2.1.2 ResNet 152 V2.....	19
2.1.3 MobileNet V1 0.25.....	20
2.2 Feature Visualizations.....	20
2.3 Experiment.....	22
3 Results.....	23
4 Discussion.....	27
Conclusion.....	30
References.....	31
Appendix.....	39
I. Layer Commitment Data.....	39
II. Experimental Materials.....	41
III. Licence.....	56

## Introduction

Situated at the crossroads of computer science and psychology, this Bachelor's thesis aims to contribute to our understanding of the neural mechanisms underlying the psychedelic state, more specifically psychedelic imagery. Psychedelic research has seen a renaissance in recent years, mainly due to the fact that psychedelic therapy has proven to be unexpectedly efficient for complex mental disorders like depression (Carhart-Harris et al., 2018; Palhano-Fontes et al., 2019; Kaup et al., 2023) and addiction (Bogenschutz et al., 2015; Garcia-Romeu et al., 2019). However, the neural mechanisms underlying the psychedelic-induced altered states of consciousness are still debated.

To strive towards the aim, convolutional neural networks (CNNs), a type of computer vision models, are leveraged. CNNs can be used for modeling the primate ventral visual stream – the brain regions predominantly responsible for object recognition (Mishkin et al., 1983). As a result, they emerge as invaluable tools for formulating and testing hypotheses about the neural mechanisms underlying psychedelic imagery. Their computations, while sometimes dubbed "black boxes", offer a relatively comprehensible framework compared to the intricate workings of the brain. The significance of psychedelic research is highlighted by its potential therapeutic applications in mitigating symptoms of various mental disorders, as mentioned before, and its broader contributions to the study of consciousness (Carhart-Harris et al., 2014).

The topic is approached both theoretically and practically – first, a theoretical framework is introduced, under which psychedelic imagery could be studied with the help of CNNs. Second, an experiment is conducted, where the information encoded at different CNN levels is visualized and compared with the imagery experienced under the influence of psychedelic substances.

The main research question of the thesis is:

- How does the capacity of CNNs to simulate psychedelic imagery differ between models that effectively or ineffectively represent primate visual system activity?

With the hypothesis put to test being:

- The visualizations of the CNN that effectively models the primate visual system activity exhibit a higher degree of similarity to psychedelic imagery.

It is speculated that if the hypothesis turns out to be true, it is meaningful to explain (some of) the neural mechanisms underlying psychedelic imagery by leveraging CNNs.

In addition to that, it is aimed to shed light on which level of CNNs' hierarchy – artificial neurons, channels or layers – is most meaningful to use as a basis of comparison with psychedelic imagery, as it is not straightforward to interpret the mapping between CNN- and brain units.

The “Background” chapter introduces why CNNs can serve as visual system's models, the technique of visualizing the information encoded in different CNN units, and how the aforementioned could be relevant for understanding the neural mechanisms underlying the psychedelic state. The “Method” chapter introduces the CNNs whose units were visualized, the software that was used, and gives a detailed overview of the experiment that was conducted. The “Results” chapter presents the results of the experiment and the “Discussion” chapter hosts a discussion over their meaning. The materials that were used to put together the experiment are added as appendices.

# 1 Background

The aim of the background chapter is to convince the reader that it's meaningful to make use of CNNs and, more specifically, their feature visualizations, to explain the neural mechanisms underlying psychedelic imagery. First, the framework of using CNNs to model the primate visual system is introduced. Then, the technique of feature visualization is delved into. Finally, some previous studies and possible interpretations for explaining psychedelic imagery with feature visualizations are presented.

## 1.1 CNNs as Primate Visual System Models

The usage of CNNs as primate visual system models relies on multiple analogies and similarities, and should be approached cautiously. The most fundamental similarities go back to the structural level – an artificial neuron is a simplification of the biological neuron. At its core, an artificial neuron, labeled at its “birth” a perceptron, linearly combines incoming data to output a response, if it reaches a certain threshold (Rosenblatt, 1957). This process takes inspiration from a biological neuron, which roughly operates by gathering signals via its dendrites and passing them on via its axon, if the signals surpass the threshold potential (Delétang, 2020).

Going a step further, the predecessor of modern computer vision models, a network titled “neocognitron”, had a structure directly inspired by that of the early visual system (Fukushima, 1980). Namely, the simple cells of the early visual system, which have the orientation-sensitive ability to detect the edges of objects, and the complex cells of the early visual system, which combine the responses of the simple cells, making it possible to detect edges regardless of their spatial location, are functionally replicated in the “neocognitron” (Hubel and Wiesel, 1962). In other words, the early areas of both biological and artificial visual systems detect the low-level properties of objects, such as edges and lines (Hubel and Wiesel, 1962; Olah et al., 2020b).

As in the biological brain, the above-described single units are combined into networks in the artificial systems, allowing for the ability to solve complex visual tasks. A fundamental difference between these networks, though, is that CNNs have a strictly feedforward and hierarchical architecture, with the representations encoded in higher layers forming directly based on those preceding them. The biological visual system, on the other hand, exhibits recurrent processing, which has been found to be critical for solving more complex image

recognition tasks (Kar et al., 2019). Importantly, this recurrent activity is poorly accounted for by the typical feedforward CNNs, but has been found to be somewhat compensated by making these networks deeper (Kar et al., 2019). The hierarchical formation of representations in the visual system is challenged by, for example, the neural activity in the visual areas being better predicted by deep neural networks that do not learn representations in a hierarchical manner (St-Yves et al., 2022). The absence of strict hierarchical processing in the biological visual areas could also be reflected in its sub-parts performing distinct (and at times overlapping) functions (Dwivedi et al., 2021). This would allow us to “treat hierarchy as an emergent property rather than a requirement for successful vision” (St-Yves et al., 2022, p. 20; Konkle, 2021), and illustrates the computational complexity and multifaceted relationships in the brain.

Regardless of that, there are several “hierarchical” parallels between CNNs and the visual system, which hints towards at least somewhat hierarchical processing in the ventral visual stream. In the spatial domain, the representations encoded in the lower CNN layers are more similar to the response of earlier visual areas, and the representations encoded in the higher CNN layers are more similar to the response of later visual areas (Yamins et al., 2014; Cadieu et al., 2014; Güçlü and van Gerven, 2015; Eickenberg et al., 2017; Dupré la Tour et al., 2021). In the temporal domain, the earlier visual system response correlates more with the activity of the lower CNN layers, as does the later visual system response with the activity of higher CNN layers (Cichy et al., 2016). In the frequency domain, the gamma activity - thought to reflect the feedforward communication between earlier and later visual areas (van Kerkoerle et al., 2014) - of the earlier visual areas correlates more with the activity of lower CNN layers, while that of the later visual areas correlates more with the activity of higher CNN layers (Kuzovkin et al., 2018). These correspondences between the CNN layers and the visual system areas are usually estimated by linear regression models predicting the neural responses from the CNN layer activations or by representational similarity analysis of the neural responses and the CNN layer activations (Kuzovkin et al., 2018).

One of the strengths of CNNs is their ability to actually solve the same tasks that the visual system solves. This allows us to directly compare the behavior of the models and the subjects. Performance-wise, several CNNs have approached human-level accuracy on image classification (Russakovsky et al., 2015), as well as nearing human-level accuracy on classifying out-of-distribution data (Geirhos et al., 2021). It has also been shown that the more models



resemble biological vision systems on a single neuron level, for example by the function of neurons in the early visual areas, the more they resemble humans on a behavioral level (Marques et al., 2021). Many researchers have taken an alternative approach, though, and instead compare the errors that models and humans make. This has led to a discovery that, at a lower, object-level resolution, models and humans make similar mistakes, but at a higher, single-image-level resolution, the error behavior of CNNs and humans is distinctly different, with the behavior of other primates resembling human behavior more than that of CNNs (Rajalingham et al., 2018; Geirhos et al., 2021). Moreover, a key finding is that CNNs tend to recognize objects based on their texture, whereas humans recognize objects more readily based on their shape or outline, a finding that makes sense in the light of the inductive bias of locality of CNNs (Baker et al., 2018). These findings have been taken to imply that some characteristics of CNNs, such as their architecture or optimization process, are fundamentally limiting for capturing primate behavior, and point towards an overall crucial difference between the two systems (Rajalingham et al., 2018; Baker et al., 2018).

Taking into account the above-described limitations, deep neural networks for computer vision are still “the closest human engineered system to biological visual system we know” (Willeke et al., 2023, p. 10). One of the most relevant similarity dimensions to the current study is CNNs’ ability to create maximally exciting images to neurons in specific visual areas and the remarkable similarity of the groupings of these stimuli to the feature families that evolve in artificial neural networks (Willeke et al., 2023, Olah et al., 2020b). As the feature visualizations of CNNs are at the core of the experiment carried out in this study, the knowledge that similar features can maximally activate both artificial and biological units is important to keep in mind. The process of creating those maximally activating stimuli for CNNs is described in the following subchapter.

## **1.2 Feature Visualization**

Feature visualization is a tool for increasing neural network interpretability by taking an approach of zooming in to individual neurons, channels or layers of the network (Olah et al., 2020a). More specifically, feature visualization implies activation maximization of neural network units by input optimization. The process involves performing gradient ascent in the input space (Erhan et al., 2009). Put simply, by calculating the gradients, it can be determined

how individual pixels of an input image should be changed in order to make a chosen optimization objective activate more (Lindsay, 2021). The process is visualized in Figure 1.

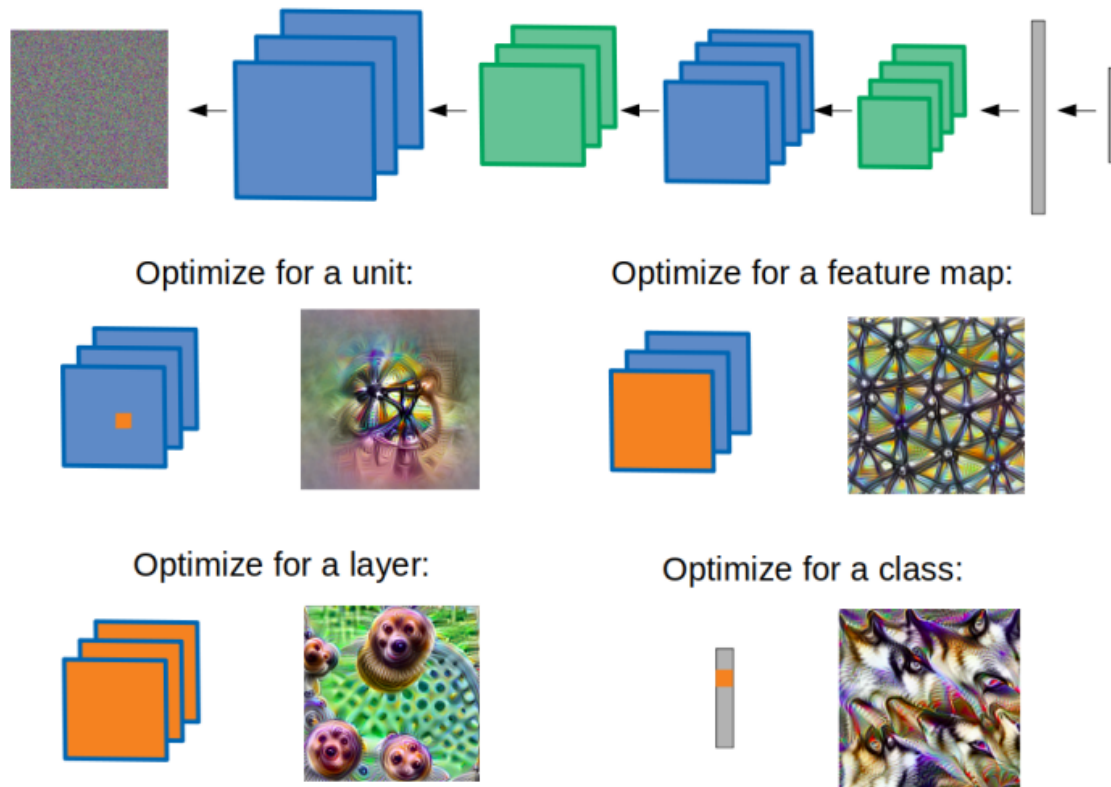


Figure 1. The process of feature visualization - arrows represent the gradient calculation between the object of optimization (depicted in orange) and the input image (Lindsay, 2021).

The input can be random noise or an arbitrary image. In the case of random noise, the result becomes “purely the result of the neural network”, with the visualization depicting what information exists in the model (Mordvintsev et al., 2015; OpenVis Conference, 2018). In the case of an arbitrary image, on the other hand, the result contains what the unit “sees” in the image, with the input biasing the model towards certain interpretations, for example, a horizon line could get amplified with tower-like features (Mordvintsev et al., 2015; OpenVis Conference, 2018). Generally, optimizing the input to activate units from lower CNN layers results in visualizations of lines, color contrasts, edges, and textures, and optimizing for units from higher CNN layers results in visualizations that could contain entire objects (Figure 2).

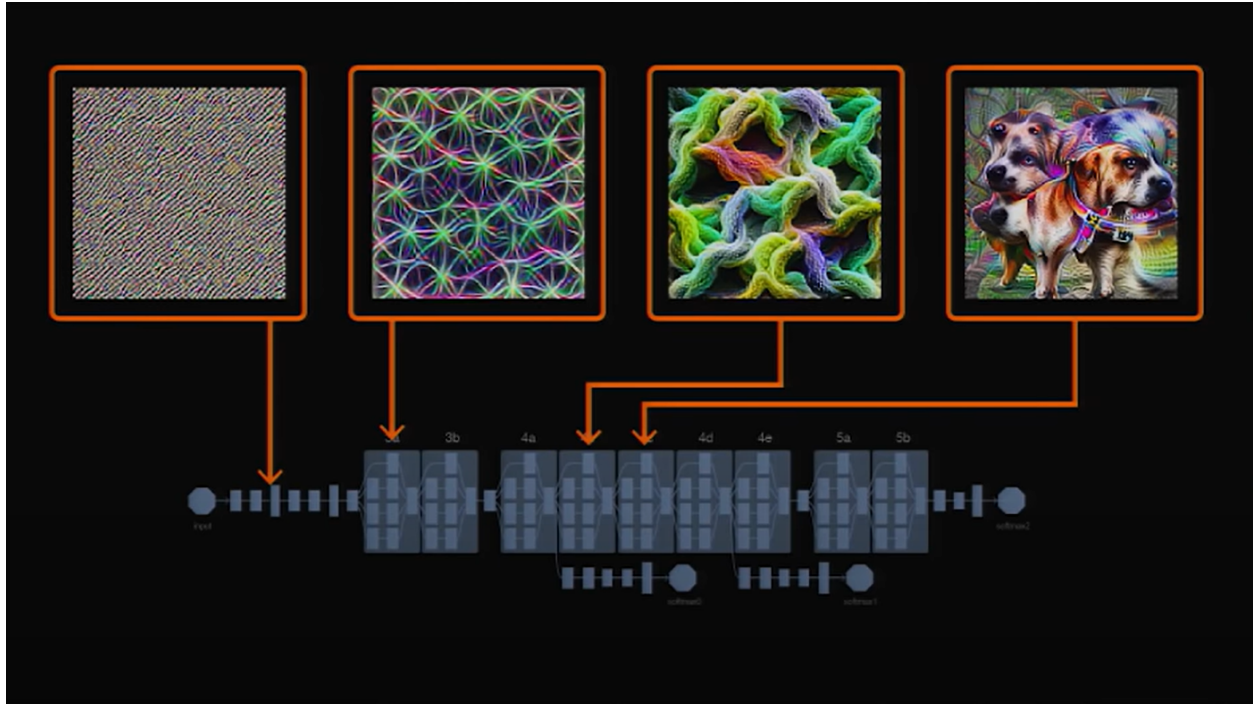


Figure 2. An example of the evolution of features in the layers of InceptionV1 (Szegedy et al., 2015; OpenVis Conference, 2018).

Feature visualization is informative, because “a pattern to which the unit is responding maximally could be a good first-order representation of what a unit is doing”, meaning it reveals for which kinds of patterns or objects a unit has learned to “look for” during, in our case, image classification (Erhan et al., 2009, p. 4). It’s also more informative than looking at dataset examples of what activates a unit the most, because feature visualization only emphasizes the cause of the activation, leaving out the correlated, yet unimportant information (Olah et al., 2017). It should be noted that on its own, feature visualization most probably produces uninformative results, as several “shortcuts” can be found for maximizing the activity of units, for example by emphasizing high frequency patterns in images (Yosinski et al., 2015; Olah et al., 2017). This can be alleviated, though, by using regularization techniques that force the visualizations to resemble natural images, for example by making neighboring pixels obtain similar values (Yosinski et al., 2015; Olah et al., 2017). Another limitation of feature visualization is that it only reveals one facet of a unit – that of maximal activation. This problem can be alleviated by, for example, enforcing diversity between successive visualization iterations (Olah et al., 2017).

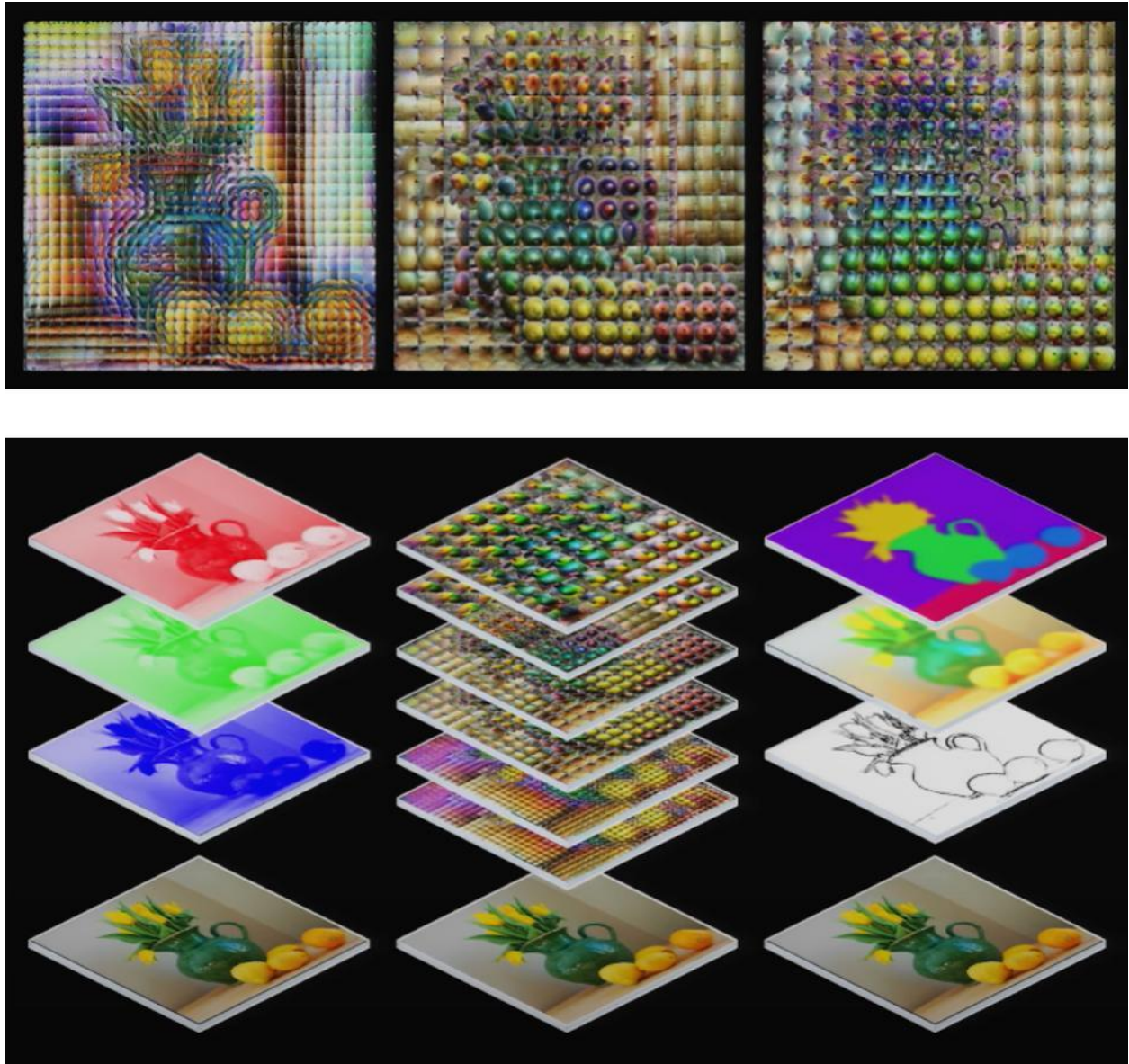


Figure 3. Top - an example of maximally activated features at each position of an input image in 3 distinct layers. Bottom - a comparison of primitives for thinking about the world: an RGB representation of an image, a feature-based representation of an image, and a more human-familiar representation of an image (OpenVis Conference, 2018).

Features can be understood as neural networks' attempts to make sense of the natural world, as they emerge in the course of these networks learning how to classify natural world images (OpenVis Conference, 2018). With this in mind, it is particularly interesting to refer back to a theme already mentioned in the previous subchapter – the same features that maximally activate

CNN units also maximally activate neurons in specific regions of the primate visual system (Willeke et al., 2023). Features represent a novel primitive for thinking about the world (OpenVis Conference, 2018), perhaps also having a unique artistic value (Figure 3). However, considering the impact of features on biological neurons, there is also potential to elucidate currently ambiguous neural mechanisms, such as the neural coding of information in the ventral visual stream, using these features.

As feature visualizations are a human interpretable way of representing what neural network units compute (Erhan et al., 2009), they allow us to compare artificial neural networks and biological systems on a more intuitive level. For example, by many, feature visualizations are deemed to resemble the visualizations elicited by psychedelic substances (LaFrance, 2015). Turns out that this similarity might not only lie on a subjective level, but be revealing of certain neural mechanisms underlying the psychedelic state. This possibility is further explored in the following subchapter.

### **1.3 The Neural Mechanisms of Psychedelic Imagery**

Over a significant portion of the past century, psychedelic research remained controversial for societal and political factors, but has emerged, in recent years, as a profoundly promising avenue of inquiry. This potential is exemplified by the therapeutic applications attributed to psychedelics, alongside their pivotal role in advancing our comprehension of consciousness (Carhart-Harris et al., 2018; Palhano-Fontes et al., 2019; Kaup et al., 2023; Bogenschutz et al., 2015; Garcia-Romeu et al., 2019; Carhart-Harris et al., 2014). It is already established that CNNs can effectively model certain neural mechanisms of visual perception, and intriguingly, the feature visualizations, which maximally activate CNN units, also elicit maximal activation in biological counterparts while being subjectively perceived as having psychedelic characteristics (Lindsay, 2021; Willeke et al., 2023; LaFrance, 2015). This prompts the exploration of whether the "psychedelicism" inherent in feature visualizations might signify deeper correspondences between CNNs and visual processing, particularly within the realm of psychedelic imagery. This subchapter delves into the introduction of pertinent neural mechanisms underlying psychedelic imagery, alongside highlighting previous achievements in investigating the psychedelic state through the application of feature visualization techniques.

There is a consensus that instead of the brain passively receiving sensory inputs, perception is an active process (Rao et al., 1999; Friston, 2005; Bastos et al., 2012). The hierarchical predictive coding theory, a prominent framework for "active" perception, suggests a dynamic interplay between bottom-up and top-down processing. Bottom-up processing integrates sensory information and prediction errors arising from the contrast between sensory inputs and "top-down" prior beliefs (Rao et al., 1999; Friston, 2005; Bastos et al., 2012). Prior beliefs can be understood as world models encoded in higher brain regions, which help to predict and make sense of the sensory inputs.

There is evidence that the trade-off between bottom-up and top-down information could be mediated by the serotonergic system, which is the primary locus of action for serotonergic psychedelic substances, such as LSD, psilocybin, DMT, and mescaline (Nichols, 2016; Azimi et al., 2020). A possible mechanism underlying psychedelics' impact on this balance is their interaction with serotonergic receptors on pyramidal cells in the cortex, where prior beliefs are thought to be encoded (Carhart-Harris and Friston, 2019). This interaction reduces confidence in prior beliefs, which means that externally or internally sourced bottom-up information wields greater influence on perception, potentially reshaping top-down priors. In the case of pathological or limiting priors, this mechanism could underpin psychedelics' therapeutic effects (Carhart-Harris and Friston, 2019).

Psychedelics are known for inducing various visual effects, including both open- and closed-eye hallucinations. Physiologically, increased connectivity between the visual cortex and the frontal and mnemonic areas of the brain, as well as increased activity of the visual cortex itself following the administration of psychedelics, has been observed (Timmermann et al., 2023; Carhart-Harris et al., 2016; de Araújo et al., 2012). The connectivity and activity of the visual cortex are positively correlated with the magnitude of visual hallucinations, as well as subjective reports indicating an increase in the richness and ability to manifest closed-eye imagery after the intake of psychedelics (Carhart-Harris et al., 2016; de Araújo et al., 2012). In addition to that, a decrease in the alpha power is observed, which is thought to serve a general inhibitory function, being representative of the top-down communication between brain areas (Carhart-Harris et al., 2016; van Kerkoerle et al., 2014). These physiological findings could be indicative of the increased influence of exogenously or endogenously stemming bottom-up information on



perception, which, without the constraining or normalizing effect of top-down processing, could manifest as visual hallucinations (Carhart-Harris and Friston, 2019; Bressloff et al., 2001).

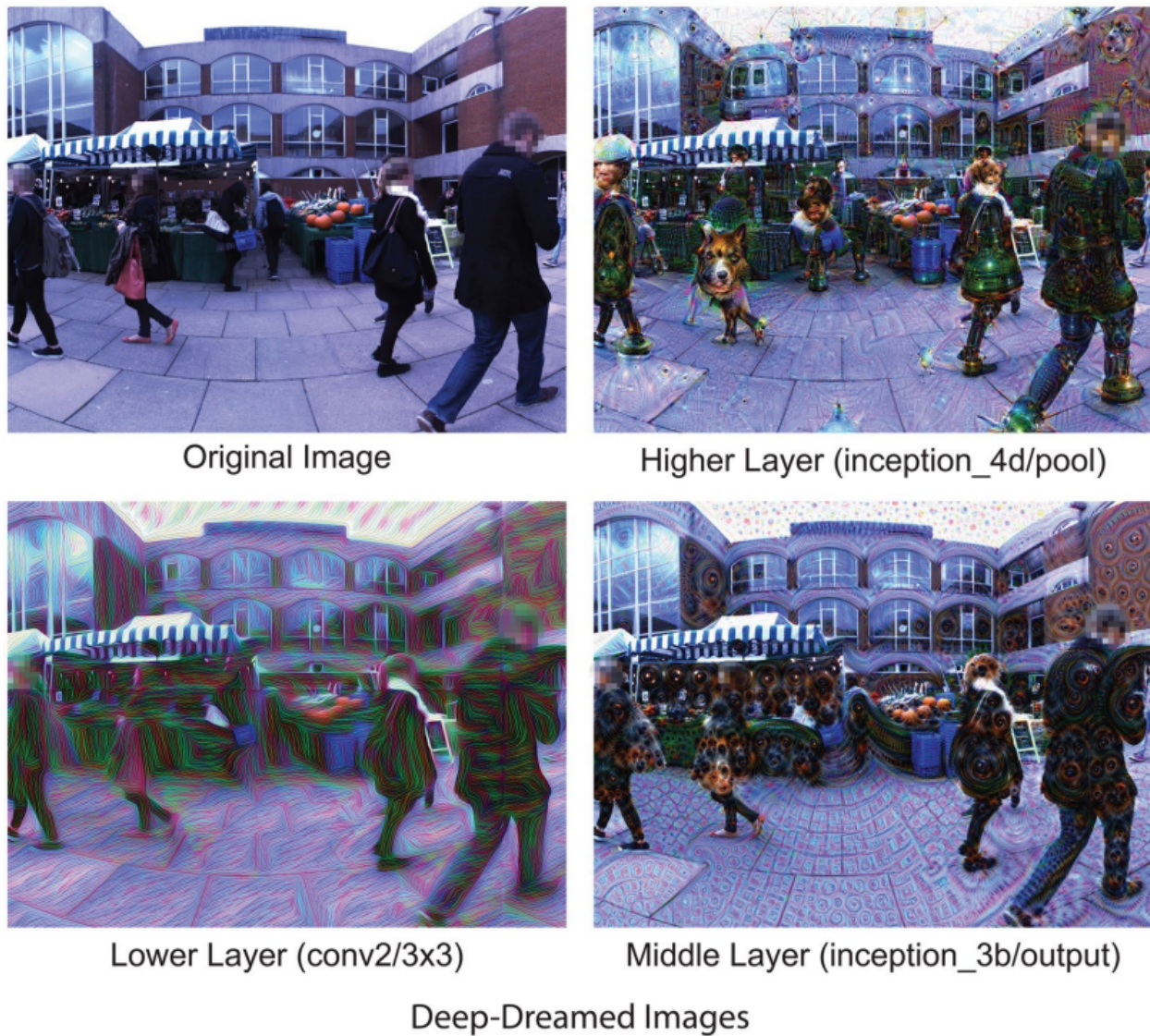


Figure 4. A natural world image modified with the Deep Dream algorithm (Mordvintsev et al., 2015); the pictures are acquired by maximizing the activation of different InceptionV1 (Szegedy et al., 2015) layers by manipulating the original image (Suzuki et al., 2017).

There are several studies aiming to explain the mechanisms underlying psychedelic perception with the help of CNNs. Studying the effect of a virtual reality program modified with the Deep Dream algorithm, a feature visualization technique that manipulates an arbitrary image so as to maximize the activation of a particular layer (Mordvintsev et al., 2015; Greco et al., 2021),

revealed that the subjective effects elicited by being immersed in that program are broadly similar to the subjective effects elicited by psychedelic substances (Suzuki et al., 2017). Moreover, measuring the brain activity of those watching a video that was modified with the aforementioned algorithm, revealed similar changes in brain activity to those actually occurring under the influence of psychedelic substances, while also providing additional insights, such as the connectivity between different brain areas increasing exclusively in the gamma band, which is thought to reflect bottom-up processing (Greco et al., 2021, van Kerkoerle et al., 2014). An example of a real world image modified with the Deep Dream algorithm can be seen in Figure 4. Interpreting the results of studies of this kind is not straightforward. Considering CNNs' capacity to model some of the processing of the visual system and the intrinsic nature of feature visualization techniques to modify input while maintaining network integrity, one could conceptualize the resemblance between feature visualizations and psychedelic imagery as the latter arising from the imposition of prior beliefs on incoming sensory information (Suzuki et al., 2017). This interpretation is consistent with the findings about pathological hallucinations, but not with what we know about the psychedelic mechanisms described in the last paragraphs, such as the weakening of top-down modulation represented by the decreased alpha power (Teufel et al., 2015; Carhart-Harris et al., 2016; van Kerkoerle et al., 2014). A possible workaround to this contradiction could be that the brain can deal with the perturbation of top-down control by upregulating the priors or predictions represented by the top-down processing (Pink-Hashkes et al., 2017). The upregulation of priors would then be paralleled with the imposition of the information encoded in the CNN units to the input image – the underlying mechanism of feature visualization.

An alternative interpretation would be that psychedelics disturb the bottom-up processing of exogenous (environmental) information, while inducing the bottom-up processing of endogenous (internally stemming) information (Schartner and Timmermann, 2020). This means that the spontaneous neural activity in the visual areas has a stronger effect on perception, illustrated by, for example, psychedelics increasing the activity of the visual areas during eyes-closed imagery to comparable levels with the activity of the visual system which actually receives sensory inputs (de Araújo et al., 2012). The spontaneous activity of the neurons in the primary visual cortex has been shown to manifest as geometrical hallucinations (Butler et al., 2012), having similarities with the feature visualizations of units from the lower CNN layers. This interpretation is



consistent with the assumption that bottom-up neural activity increases after the consumption of psychedelics, having a greater effect on perception, as well as being supported by the abundance of serotonergic cells in the visual system (Carhart-Harris and Friston, 2019; Beliveau et al., 2017). In this case, the bottom-up endogenous activity is paralleled with the imposition of the information encoded in the CNN units to the input image.

The similarities between psychedelic imagery and feature visualizations are put to test in an experiment, the process and results of which are introduced in the following subchapters.

## 2 Method

The practical part of this Bachelor's thesis put the similarities between feature visualizations and psychedelic imagery to test. The aim was to strengthen the ground for understanding and explaining the neural mechanisms underlying psychedelic imagery, making use of the computational mechanisms underlying feature visualizations. The main purpose of the experiment was to study how the capacity of CNNs to simulate psychedelic imagery differs between models that effectively or ineffectively represent primate visual system activity. It was speculated that if the visualizations of the CNN that effectively models the primate visual system activity exhibit a higher degree of similarity to psychedelic imagery, it is meaningful to explain the psychedelic mechanisms of the visual system with the help of CNNs' feature visualization. The results of the same experiment were analyzed further in Lüübek (2023a), where the main question addressed was different from the focus of the current thesis, being mainly centered on the more psychological topic of how one's previous psychedelic experiences influence their perception of CNNs' feature visualizations. Generative artificial intelligence (AI) was used to improve the readability of a handful of paragraphs of the thesis, with the specific prompt being: "How to make this better: [text]?", the result being a combination of the original text and certain AI-enhanced sentences or phrases (OpenAI, 2023).

### 2.1 Models

The first step was to choose two CNNs that differ in their capacity to represent the primate ventral visual stream activity. For this, the Brain Score platform<sup>1</sup> was leveraged.

#### 2.1.1 Brain Score

Brain Score is an integrative benchmarking platform, the goal of which is to incentivize the progress of neurally mechanistic visual intelligence models (Schrimpf et al., 2020). Neurally mechanistic models refer to models built solely of networks of artificial neurons and integrative benchmarks refer to various neural and behavioral measurements of primates that the models are being evaluated on. The main idea behind the platform is that by testing the models on benchmarks that encompass the full span of a domain of intelligence, in this case visual

---

<sup>1</sup> <https://www.brain-score.org/>

intelligence, integrative models of that domain of intelligence can be reached, and better models can be built by combining the right components of different models. Those neurally mechanistic models then represent hypotheses of how the brain could accomplish that domain of intelligence, which can be tested, improved, as well as falsified.

The benchmarks are more than mere brain- or behavioral data – they contain the means for testing the models on the same experiments that subjects took part in. A few examples of the experimental data constituting the benchmarks include neuronal responses to various stimuli in different ventral visual stream areas and behavioral signatures of primates in object discrimination tasks (Marques et al., 2021; Rajalingham et al., 2018). In order to be comparable with the subjects, models need to fulfill a few requirements, such as having their layers mapped with certain ventral visual stream areas and their (artificial neuron) activations mapped with neural responses (such as spike rates). By fulfilling the requirements models become *in silico* brain models that can be experimented on, just like an experimental subject can be experimented on.

The brain models then receive a score for each of the benchmarks included on the platform, and a leaderboard of current best brain models emerges. This leaderboard was used to guide the choice of models for the current study. Namely, the best and the worst brain model, which were concurrently 1) with an underlying CNN structure and 2) available as pre-trained models in the feature visualization library introduced in one of the following subchapters, were chosen.

### **2.1.2 ResNet 152 V2**

The “effective visual system model” chosen for visualization purposes turned out to be a model titled ResNet 152 V2, a residual neural network engineered to support a very deep architecture (He et al., 2015). Very broadly, it’s a convolutional neural network that includes additional connections between its non-successive layers, which help to improve its image classification performance, while also adding more layers to the network. Its top-1 accuracy on the ImageNet Large Scale Visual Recognition Challenge – the most widely used benchmark for computer vision models – is 77.8%, while its top-5 accuracy is 94.1% (Russakovsky et al., 2015). The version of the model that was used for visualization was the pre-trained one of the TensorFlow-Slim image classification model library<sup>2</sup>. The model placed 9th on the Brain Score

---

<sup>2</sup> <https://github.com/tensorflow/models/tree/master/research/slim>

leaderboard at the time of writing, with an average score of 0.432, calculated based on its performance on all the benchmarks of the platform.

### **2.1.3 MobileNet V1 0.25**

The “ineffective visual system model” chosen for visualization purposes turned out to be a version of the MobileNet V1 network (Howard et al., 2017). The main goal of MobileNets is to provide computationally effective, yet sufficiently accurate computer vision models. The computational cost of MobileNets can be modified by two hyperparameters – a width multiplier, which reduces the number of channels in each layer of the model, and a resolution multiplier, which modifies the resolution of the input image and the internal representations of the model. The TensorFlow-Slim implementation of MobileNet V1 was used for visualization, with the width multiplier hyperparameter set to 0.25. Its top-1 accuracy on the ImageNet Large Scale Visual Recognition Challenge is 41.5%, while its top-5 accuracy is 66.3%. The various versions of MobileNet V1 0.25, differing by their resolution multiplier, ranked 160th-178th on the Brain Score leaderboard at the time of writing, with their average scores ranging from 0.312 to 0.277, calculated based on their performance on all the benchmarks of the platform.

## **2.2 Feature Visualizations**

The second step was to visualize the units of the chosen models. This was done by leveraging the Lucid library<sup>3</sup> – an open-source software for neural network interpretability research. It was decided to only visualize units from the CNN layers that corresponded to the ventral visual stream areas V1, V2, V4, and IT. The layer commitment data provided by the Brain Score team guided the selection process (see Appendix I). A complication emerging here was that the TensorFlow-Slim implementation of the “ineffective visual system model” did not contain all the exact layers that corresponded to the ventral visual stream areas. The researcher had to thus decide on, at their best discretion, which of the implementation’s layers to use. All the layers which units were used for visualization are listed in Appendix II.

To address one of the research questions of the study – which level of CNNs’ structural hierarchy is most meaningful to use as a basis of comparison with psychedelic imagery – visualizations were obtained from all the possible levels of hierarchy. This means that visualizations that most

---

<sup>3</sup> <https://github.com/tensorflow/lucid>

excited a single artificial neuron, a collection of neurons – a channel, and a collection of channels – a layer, were obtained. The specific units were chosen so that they spanned the breadth of the layers and remained at the same relative positions across the layers. It was speculated that this would enable a more meaningful analysis of the differences between the visualizations from individual layers, thus potentially allowing for assumptions about the mechanisms underlying psychedelic visualizations at a greater, single-visual-area resolution.

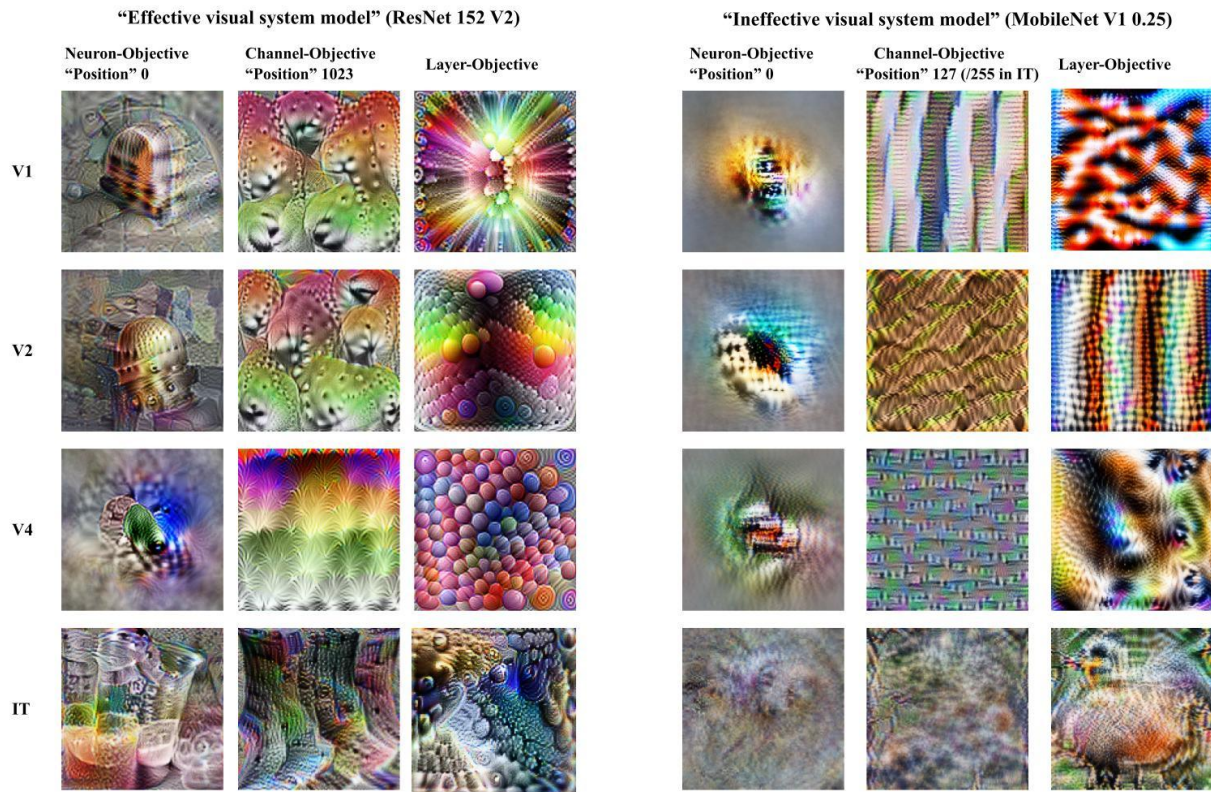


Figure 5. A selection of the generated feature visualizations.

The process resulted in a dataset of 88 visualizations – 5 neuron- and 5 channel-objective visualizations, as well as the whole layer's visualization, were created for both of the models' every layer that was focused on. A selection of the visualizations is presented in Figure 5. All the visualizations can be seen in Appendix II.

## 2.3 Experiment

The feature visualizations of the effective and the ineffective visual system models were used to put together a two-part questionnaire<sup>4</sup>. Participants were recruited via social media, with the only eligibility criterion being having at least one experience with a serotonergic psychedelic substance. 35 self-selected participants took part in the experiment, with 32's results included in the analysis. Three participants' results were eliminated because of one's invalid ratings and the others' invalid substance-selection.

In the first part of the experiment, participants were asked to rate each feature visualization based on how possible they thought it was to experience psychedelic imagery with similar properties. By properties the overall characteristics, complexity, or intensity of the colors, contrasts, edges, textures, shapes, and objects in the feature visualizations was meant. The visualizations could be rated on a scale from 1 to 10, with 1 being impossible and 10 being very possible. Participants were directed to only base their ratings on their own experiences, not to dwell on an idea of psychedelic imagery they could have obtained via external mediums.

In the second part of the experiment, participants were asked to provide information about one of their psychedelic experiences, preferably the most intense one from the ones they relied on while rating the visualizations. By intensity, mostly that of the imagery was favored, but participants were free to choose the experience they thought was most relevant for the cause. Questions about the substance, dosage, overall intensity, and memorability of the experience were asked. The remaining questions focused more on the visual aspects of the experience, with the ones about the extent of open- or closed-eye visualizations and their vividness, complexity, richness, and immersiveness chosen by the example of Timmermann et al. (2023) and Carhart-Harris et al. (2016). Finally, questions about the prominence or extent of different properties – such as colors, patterns, and distorted objects – were asked, with the specific properties chosen by the example of Lüübek (2023b) and Olah et al. (2020b).

---

<sup>4</sup> <https://forms.gle/3PHJtde1coEP82Z38>

### 3 Results

This chapter presents the results obtained from the experiment. The main goal of the experiment was to find out how the capacity of CNNs to simulate psychedelic imagery differs between models that effectively or ineffectively represent primate visual system activity. It was hypothesized that if the visualizations of an effective visual system model are perceived to be more similar to psychedelic imagery, feature visualization could serve as a meaningful tool for studying the neural mechanisms underlying psychedelic imagery. Participants who had experiences with serotonergic hallucinogens rated the feature visualizations of two CNNs, ResNet 152 V2 – the effective visual system model, and MobileNet V1 0.25 – the ineffective visual system model, having no information about the origin of the individual feature visualizations nor about the exact purpose of the experiment. Visualizations were rated on a scale of 1-10, based on how possible participants deemed it to be to experience psychedelic imagery similar to the feature visualizations, 1 being impossible and 10 being very possible.

To assess the significance of the difference between the ratings of the two models, a participant-based analysis was opted for. The reasons for that include the absence of a direct correspondence between the feature visualizations of the two models and the violation of data independence due to the same participants rating the visualizations of both models, rendering the *t*-tests that could have been performed in a visualization-based analysis irrelevant or invalid. This means that the mean rating of each participant across visualizations was compared between conditions. The approximate normality of the data was assessed using the D'Agostino and Pearson's normality test. Its results justified the use of a paired *t*-test, the results of which indicate a significant difference in the mean ratings of individual participants between the two models ( $t(31) = 3.14, p = .004$ ). With a few exceptions, participants found the visualizations of the effective visual system model more resemblant of psychedelic imagery (Figure 6).



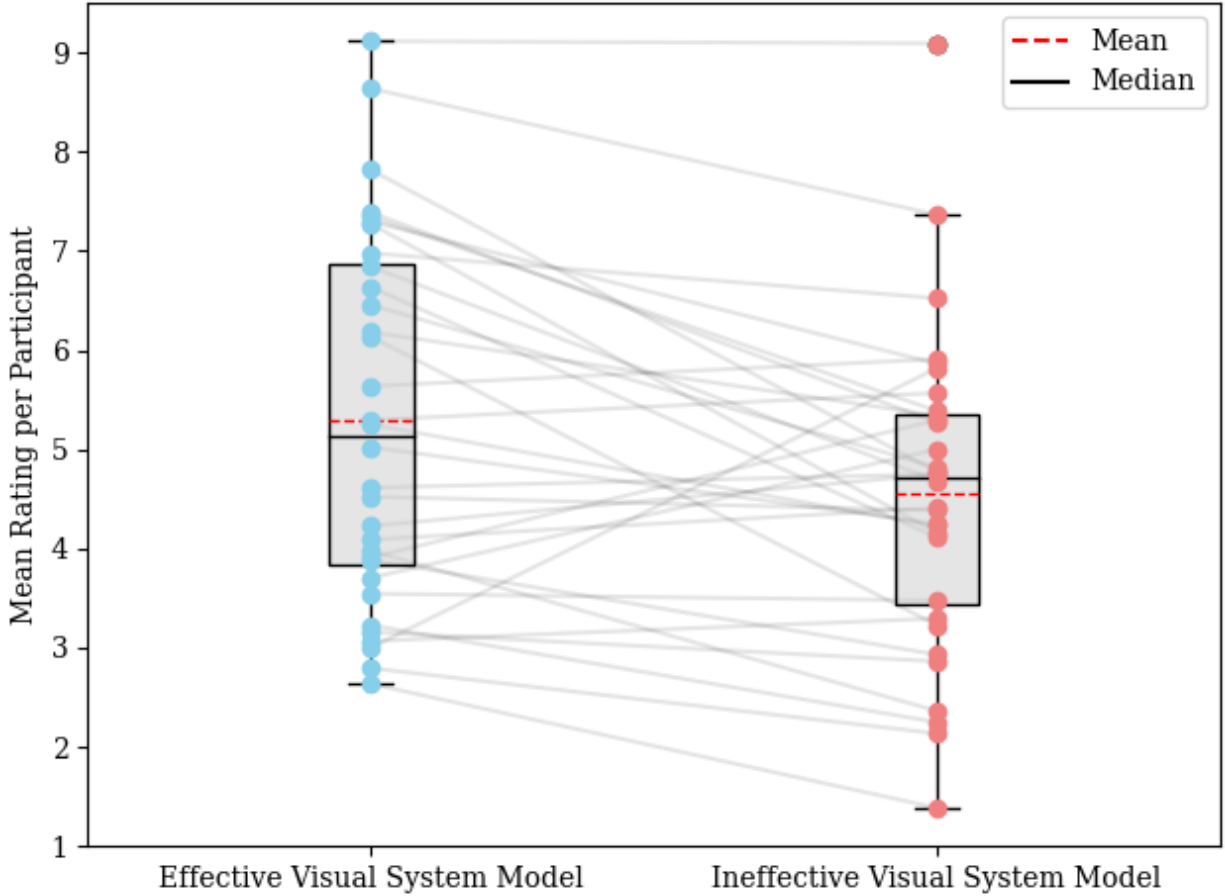


Figure 6. A comparison of the mean ratings per participant and the ratings’ variance between the two models.

As the individual feature visualizations differed rather notably in their properties, it was deemed meaningful to conduct some visualization-based analysis as well. Only the visualizations of the effective visual system model were focused on, as the participant-based analysis indicated that those are more similar to psychedelic imagery. The mean rating of each visualization across participants was calculated. No comparison tests were carried out due to reasons pointed out in the previous paragraph. To address the second research question of the study – which level of CNNs’ hierarchy is most meaningful to use as a basis of comparison with psychedelic visualizations – visualizations were separated by their optimization objectives and layers of origin (Figure 7). The results pointed towards no notable difference between the average ratings of the visualizations with different optimization objectives, with the most variance in the case of neuron-objective visualizations and the least in the case of layer-objective visualizations



( $M(\text{neuron}) = 5.31$ ,  $SD(\text{neuron}) = 0.72$ ;  $M(\text{channel}) = 5.30$ ,  $SD(\text{channel}) = 0.66$ ;  $M(\text{layer}) = 5.30$ ,  $SD(\text{layer}) = 0.59$ ). However, while additionally considering the visualizations' layers of origin, the neuron-objective visualizations from the layer corresponding to V1 ( $M = 6.05$ ,  $SD = 0.72$ ), the channel-objective visualizations from the layer corresponding to V4 ( $M = 5.98$ ,  $SD = 0.73$ ), and the layer-objective visualization from the layer corresponding to IT ( $M = 6.06$ ) received the highest average ratings.

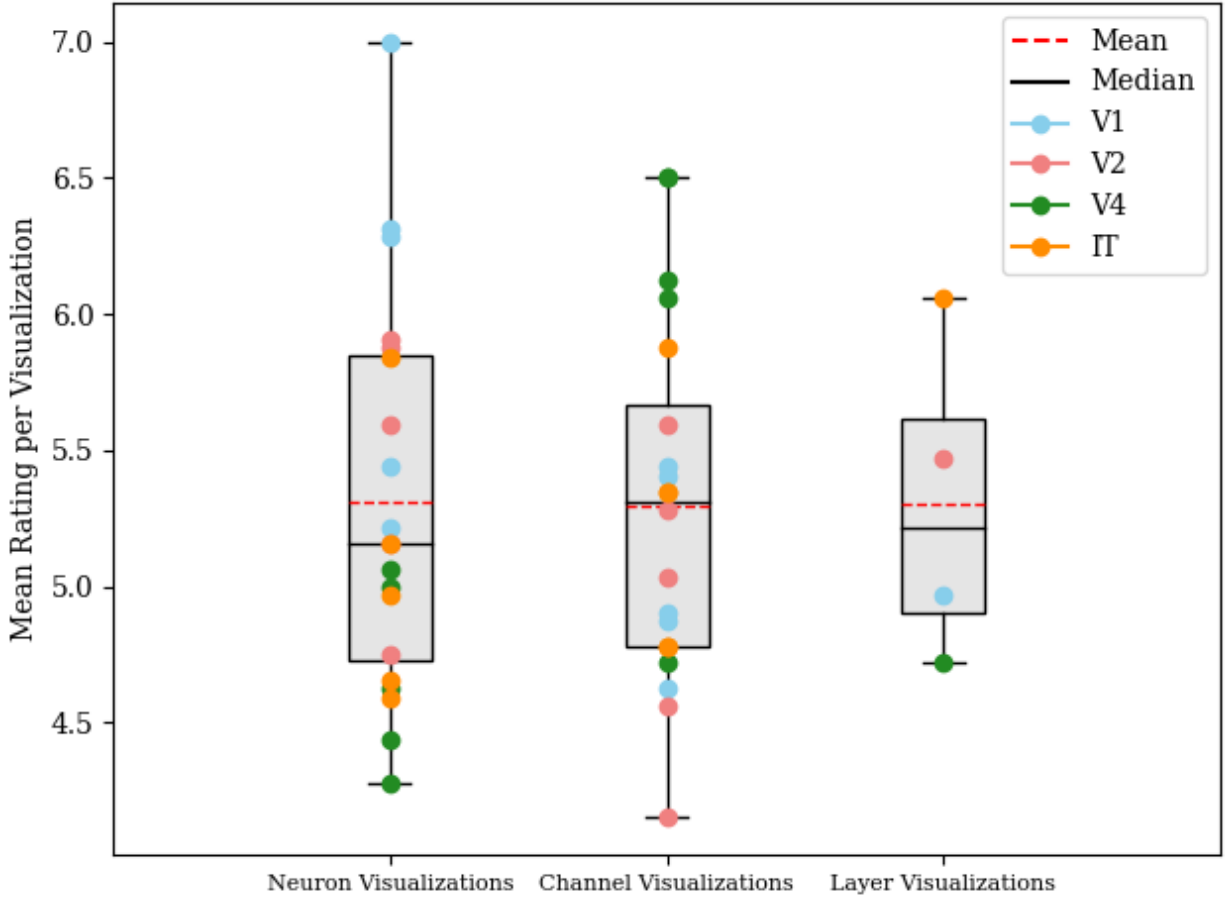


Figure 7. A comparison of the mean ratings per visualization and the ratings' variance between the three optimization objectives. The layers of origin are indicated by the different colors.

Overall, there seemed to be no remarkable differences between the average ratings of the visualizations from different layers of origin, with the mean of the “V1 ratings” being slightly higher than of those from the other layers of origin, and the variance of the “V4 ratings” being the highest ( $M(V1) = 5.50$ ,  $SD(V1) = 0.74$ ;  $M(V2) = 5.22$ ,  $SD(V2) = 0.55$ ;  $M(V4) = 5.28$ ,  $SD(V4) = 0.85$ ;  $M(IT) = 5.22$ ,  $SD(IT) = 0.52$ ) (Figure 8).

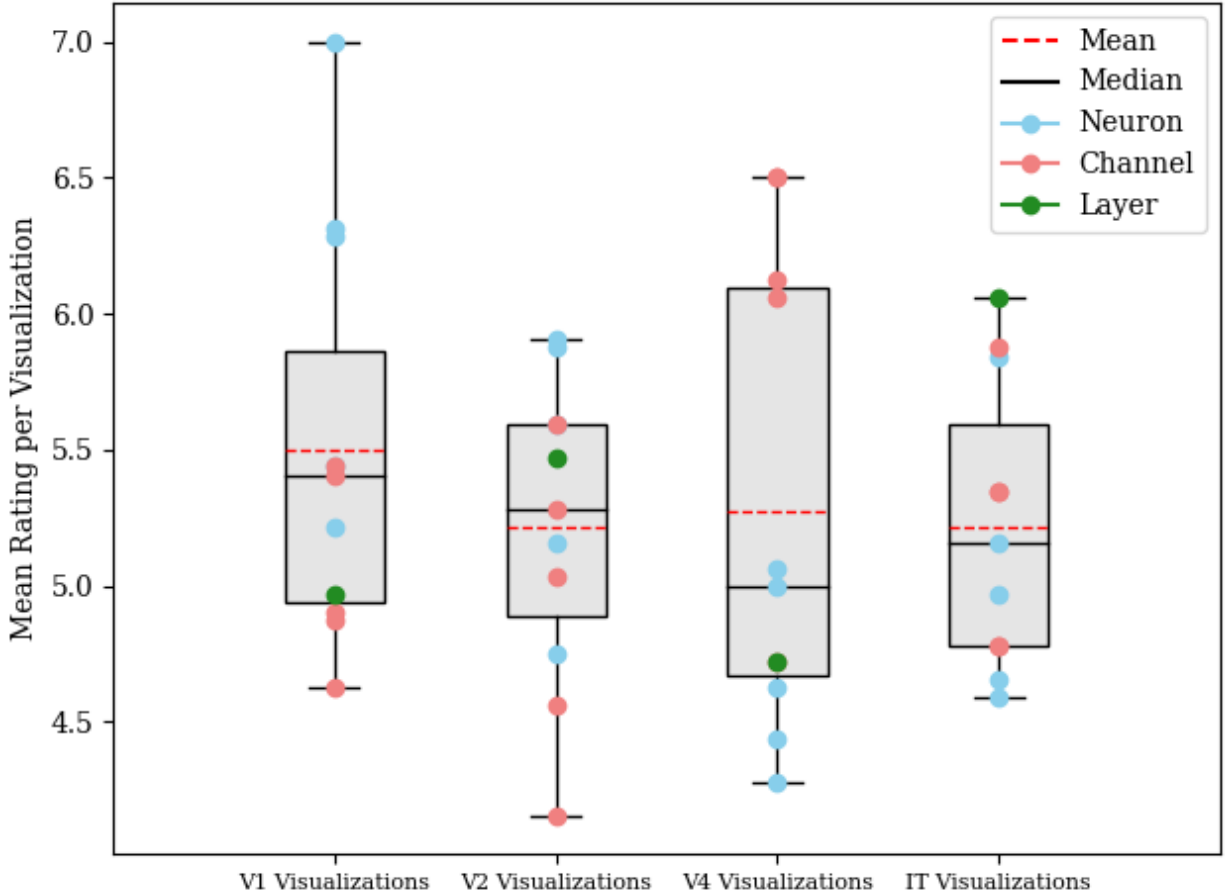


Figure 8. A comparison of the mean ratings per visualization and the ratings' variance between the four layers of origin. The optimization objectives are indicated by the different colors.

Data from the second part of the experiment was analyzed in Lüübek (2023a), where correlations between participant-based mean ratings of the visualizations and the different subjective measures were investigated. As in Lüübek (2023a) it was deemed necessary to take into account the results of the current study, specifically how the ratings differ between the visualizations of the two models, the current results were touched on in that paper as well. It should be mentioned, however, that the main research question of Lüübek (2023a) was more psychological, namely how are one's previous psychedelic experiences related to the perception of CNN feature visualizations, with the main hypothesis put to test being that the more intense psychedelic imagery one has experienced, the higher their ratings of the feature visualizations. Thus, the current thesis is complementary to Lüübek (2023a). The possible implications of the current results are discussed in the following chapter.

## 4 Discussion

The practical part of this thesis aimed to explore whether the widespread opinion that CNNs' feature visualizations resemble psychedelic imagery could signify deeper correspondences between CNNs and visual processing (LaFrance, 2015). It was speculated that if the visualizations of the effective visual system model resemble psychedelic imagery more, there is potential to investigate the neural mechanisms underlying psychedelic imagery through the application of feature visualization techniques.

The aim was approached by asking subjects with psychedelic experiences to rate the feature visualizations of an effective and an ineffective visual system model, based on how possible they thought it was to experience psychedelic imagery with similar qualities. The results indicate that on average, participants perceived the visualizations of the effective visual system model to be more resemblant of psychedelic imagery than those of the ineffective visual system model. It is known from previous research that during psychedelic imagery, the activity and connectivity of the visual system are higher than during normal conditions (Carhart-Harris et al., 2016; de Araújo et al., 2012). Besides feature visualizations resembling psychedelic imagery on a subjective level, they are created by maximizing the activation of CNN units, having a parallel with the aforementioned heightened activity of the visual system during psychedelic imagery (Erhan et al., 2009). Moreover, the stimuli that maximize the activity of CNN units also maximize the activity of biological units (Willeke et al., 2023). The current results clarify that not all feature visualizations resemble psychedelic imagery to the same extent, with those of the CNN that approximates the visual system effectively being more accurate. This could point towards some of the computational mechanisms of effective visual system models being suitable for explaining the neural mechanisms underlying psychedelic imagery.

A defining characteristic of the effective visual system model used in this study was its very deep architecture, which is thought to somewhat approximate the recurrent processing in the brain (Kar et al., 2019). This could be a possible reason for its visualizations being more resemblant to psychedelic imagery. Additionally, it must be acknowledged that the effective visual system model was concurrently a more powerful computer vision model, with its top-1 classification accuracy on the ImageNet Large Scale Visual Recognition Challenge being almost twice as high as that of the ineffective visual system model – 77.8% compared to 41.5%. A recommended

approach for forthcoming research endeavors is to center their analyses on feature visualizations extracted from CNNs that share more analogous computational or architectural characteristics, while diverging in their approximation of the visual system. This strategy aims to mitigate potential confounding influences stemming from varying computational and architectural attributes.

As explained in “The Neural Mechanisms of Psychedelic Imagery” chapter, a possible interpretation for explaining the neural mechanisms underlying psychedelic imagery through the computational mechanisms of feature visualization is that psychedelic imagery stems from the greater influence of internally sourced bottom-up information on perception (Schartner and Timmermann, 2020). This is supported by, for example, the spontaneous activity of the primary visual area V1 manifesting as geometrical hallucinations, which resemble the feature visualizations of earlier CNN layers (Butler et al., 2012). In this study, feature visualizations were obtained from the CNN layers that best corresponded to 4 visual system areas – V1, V2, V4, and IT. While statistical comparison tests between visualizations from different layers of origin weren't feasible due to the experimental design, the available outcomes show no substantial discrepancies among them. Nonetheless, the present findings hint that visualizations stemming from the layer corresponding to the primary visual area V1 might be perceived as most akin to psychedelic imagery. This appears consistent with the finding that, under the influence of psychedelics, the activity of the primary visual area during closed-eye imagery parallels its response to actual sensory inputs (de Araújo et al., 2012). It is suggested to further investigate the differences between the “psychedelicism” of the feature visualizations originating from CNN layers tied to different visual system regions, with the current results hinting at the potential greater influence of endogenous activity from the primary visual area during psychedelic imagery.

A secondary goal of this study aimed to illuminate the alignment between CNN architecture and the primate visual system. While existing research has established functional similarities between artificial and biological neurons (Marques et al., 2021), the task of identifying biological counterparts for artificial channels or layers is more intricate. One plausible perspective suggests that artificial activations embody conceptual representations within the neural network, wherein channel and layer activations encapsulate more abstract notions than those inherent in neurons (T. Khajuria, personal communication, 20.04.2023). In tackling this issue, visualizations were

obtained that maximized the activity of artificial neurons, channels and layers. As mentioned before, no statistical comparison tests could have been performed, and the available results suggest no remarkable differences between the mean ratings of the visualizations with different optimization objectives. However, when additionally considering the visualizations' layers of origin, neuron-objective visualizations from the layer corresponding to the primary visual area, channel-objective visualizations from the layer corresponding to the area V4, and the layer-objective visualization from the layer corresponding to the area IT received the highest ratings. These outcomes may hint at the alignment between artificial and biological systems being most pronounced at the neuronal level for the earlier stages of both systems, while manifesting at a more abstract level for the later stages.

CNNs are generally considered to approximate only the initial phases of neural processing in the primate visual system, with their efficacy diminishing as recurrent processing stages emerge, which are not adequately captured by CNNs (Lindsay, 2021). Primarily designed for image classification, the development of features could be conceptualized as the result of CNNs trying to make sense of the natural world (OpenVis Conference, 2018). While correspondences between CNNs and the visual system have prompted the notion that "capturing the rich statistics in images of the world that surrounds us must be a driving principle of the structure of visual cortex" (Eickenberg et al., 2017, p. 192; Olshausen & Field, 1996), it's crucial to acknowledge that the visual system's scope extends beyond mere image or scene classification, encompassing, for example, visual reasoning and navigation (Lindsay, 2021). These limitations of CNNs raise inquiries about the degree to which feature visualization can approximate psychedelic imagery. Intuitively, it seems that feature visualization applied to random noise would better capture closed-eye imagery, as in both cases, the end result depicts the information that is already encoded in the systems (Mordvintsev et al., 2015). However, the inherent limitation of CNNs to approximate only the initial phases of visual processing implies that closed-eye imagery could prove elusive for feature visualization. A pertinent suggestion for future investigations is to delineate between closed-eye and open-eye visualizations, which would shed light on whether feature visualization can capture distinct forms of psychedelic imagery.

## Conclusion

The primary objectives of this Bachelor's thesis were twofold: first, to establish a framework for employing CNN feature visualization techniques to unravel the neural underpinnings of psychedelic imagery; and second, to practically investigate this phenomenon. This inquiry is, predominantly, motivated by the interplay between CNNs and the primate visual system, ranging from neurons to behavior, the capacity of feature visualizations to maximally activate both artificial and biological units, and the intuitive parallels between the physiological effects of psychedelic substances and the computational mechanisms underlying feature visualization.

The experiment conducted involved individuals with prior psychedelic experiences rating the feature visualizations of two CNNs, based on how possible they estimated it to be to experience psychedelic imagery with similar qualities. The results confirmed the hypothesis that visualizations from an effective visual system model exhibit a significantly higher resemblance to psychedelic imagery than those from an ineffective visual system model. This finding suggests that the computational mechanisms within effective visual system models could offer insight into the neural processes that drive psychedelic imagery. The correspondence seemed the greatest in the case of visualizations stemming from the layer corresponding to the primary visual area, speculated to hint at the greater role of that visual area in constructing perception during psychedelic imagery. Furthermore, the alignment between artificial and biological systems seemed the greatest at the neuronal level during the early processing stages of the two systems, and at a higher abstraction level during the later processing stages of the two systems.

In terms of future investigations, it is suggested to refine the experimental design by visualizing the units of more computationally similar CNNs, incorporating non-overlapping rating groups, and introducing control groups. Additionally, applying feature visualization both on random noise and natural world images could enhance the ecological validity of the results, potentially providing a more comprehensive understanding of the differences between the neural mechanisms underlying open-eye and closed-eye visualizations.

## References

- Azimi, Z., Barzan, R., Spoida, K., Surdin, T., Wollenweber, P., Mark, M. D., Herlitze, S., & Jancke, D. (2020). Separable gain control of ongoing and evoked activity in the visual cortex by serotonergic input. *Elife*, 9, Article e53552. <https://doi.org/10.7554/eLife.53552>
- Baker, N., Lu, H., Erlikhman, G., & Kellmann, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS Computational Biology*, 14(12), Article e1006613. <https://doi.org/10.1371/journal.pcbi.1006613>
- 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>
- Bogenschutz, M. P., Forcehimes, A. A., Pommy, J. A., Wilcox, C. E., Barbosa, P. C., & Strassman, R. J. (2015). Psilocybin-assisted treatment for alcohol dependence: a proof-of-concept study. *Journal of Psychopharmacology*, 29(3), 289-299. <https://doi.org/10.1177/0269881114565144>
- 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., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., & Nutt, D. (2014). The entropic brain: a theory of conscious states

- informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, 8, 20. <https://doi.org/10.3389/fnhum.2014.00020>
- 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., ... & 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., Bolstridge, M., Day, C. M., Rucker, J., Watts, R., Erritzoe, D. E., Kaelen, M., Giribaldi, B., Pilling, S., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Curran, H. V., & Nutt, D. J. (2018). Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. *Psychopharmacology*, 235, 399-408. <https://link.springer.com/article/10.1007/s00213-017-4771-x>
- 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., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6, Article 27755. <https://doi.org/10.1038/srep27755>
- 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>
- Delétang, G. (2020, September 20). Neurophysiology and Artificial Neural Networks. *Medium*. <https://medium.com/swlh/neurophysiology-and-artificial-neural-networks-7084aff45eed>
- Dupré la Tour, T., Lu, M., Eickenberg, M., & Gallant, J. L. (2021). *A finer mapping of convolutional neural network layers to the visual cortex*. SVRHM 2021 Workshop @ NeurIPS. <https://openreview.net/forum?id=EcoKpq43UI8>



- Dwivedi, K., Bonner, M. F., Cichy, R. M., & Roig, G. (2021). Unveiling functions of the visual cortex using task-specific deep neural networks. *PLoS Computational Biology*, 17(8), Article e1009267. <https://doi.org/10.1371/journal.pcbi.1009267>
- 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). Université 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](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>
- Garcia-Romeu, A., Davis, A. K., Erowid, F., Erowid, E., Griffiths, R. R., & Johnson, M. W. (2019). Cessation and reduction in alcohol consumption and misuse after psychedelic use. *Journal of Psychopharmacology*, 33(9), 1088-1101. <https://doi.org/10.1177/0269881119845793>
- Geirhos, R., Narayanappa, K., Mitzkus, B., Thieringer, T., Bethge, M., Wichmann, F. A., & Brendel, W. (2021). Partial success in closing the gap between human and machine vision. *Advances in Neural Information Processing Systems*, 34, 23885-23899. <https://openreview.net/forum?id=QkljT4mrfs>
- Greco, A., Gallitto, G., D'Alessandro, M., & Rastelli, C. (2021). Increased entropic brain dynamics during DeepDream-induced altered perceptual phenomenology. *Entropy*, 23(7), 839. <https://doi.org/10.3390/e23070839>
- 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>
- Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, 22, 974-983. <https://doi.org/10.1038/s41593-019-0392-5>
- Kaup, K. K., Vasser, M., Tulver, K., Munk, M., Pikamäe, J., & Aru, J. (2023). Psychedelic replications in virtual reality and their potential as a therapeutic instrument: an open-label feasibility study. *Frontiers in Psychiatry*, 14, <https://doi.org/10.3389/fpsyt.2023.1088896>
- Konkle, T. (2021). *Emergent organization of multiple visuotopic maps without a feature hierarchy*. bioRxiv. <https://doi.org/10.1101/2021.01.05.425426>
- Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.-P., Baciú, M., Kahane, P., Rheims, S., Vidal, J. R., & Aru, J. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications Biology*, 1, Article 107. <https://doi.org/10.1038/s42003-018-0110-y>
- LaFrance, A. (2015, September 3). When Robots Hallucinate. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2015/09/robots-hallucinate-dream/403498/>
- Li, Y., Yosinski, J., Clune, J., Lipson, H., & Hopcroft, J. (2015). *Convergent learning: Do different neural networks learn the same representations?* arXiv. <https://doi.org/10.48550/arXiv.1511.07543>
- 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](https://doi.org/10.1162/jocn_a_01544)

- Lüübek, C. (2023a). *Comparing Psychedelic Visualizations with Convolutional Neural Networks' Feature Visualizations* [Unpublished research paper]. University of Tartu.
- Lüübek, C. (2023b, January 16). The Psychedelism 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>
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414-417.  
[https://doi.org/10.1016/0166-2236\(83\)90190-X](https://doi.org/10.1016/0166-2236(83)90190-X)
- 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>
- Nichols, D. E. (2016). Psychedelics. *Pharmacological Reviews*, 68(2), 264-355.  
<https://doi.org/10.1124/pr.115.011478>
- Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020a, March 10). Zoom In: An Introduction to Circuits. *Distill*. <https://distill.pub/2020/circuits/zoom-in/>
- Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020b, April 1). An Overview of Early Vision in InceptionV1. *Distill*.  
<https://distill.pub/2020/circuits/early-vision/>
- Olah C., Mordvintsev A., & Schubert L. (2017, November 7). Feature Visualization. *Distill*.  
<https://distill.pub/2017/feature-visualization/>
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607-609.
- OpenAI. (2023). *ChatGPT* (August 3 version) [Large language model].  
<https://chat.openai.com/chat>
- OpenVis Conference. (2018, July 31). *SHAN CARTER - OpenVisConf 2018* [Video]. Youtube.  
<https://www.youtube.com/watch?v=jlZsgUZaIyY>

- Palhano-Fontes, F., Barreto, D., Onias, H., Andrade, K. C., Novaes, M. M., Pessoa, J. A., Mota-Rolim, S. A., Osório, F. L., Sanches, R., dos Santos, R. G., Tófoli, L. F., Silveira, G. de O., Yonamine, M., Riba, J., Santos, F. R., Silva-Junior, A. A., Alchieri, J. C., Galvão-Coelho, N., L., Lobão-Soares, B., ... & Araújo, D. B. (2019). Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychological medicine*, 49(4), 655-663. <https://doi.org/10.1017/S0033291718001356>
- 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](https://www.noisebridge.net/images/e/ef/Perception_is_in_the_Details12.pdf)
- Rajalingham, R., Issa, E. B., Bashivan P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks. *Journal of Neuroscience*, 38(33), 7255-7269. <https://doi.org/10.1523/JNEUROSCI.0388-18.2018>
- 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>
- Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115, 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- Schartner, M. M., & Timmermann, C. (2020) Neural network models for DMT-induced visual hallucinations. *Neuroscience of Consciousness*, 2020(1). <https://doi.org/10.1093/nc/niaa024>
- 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. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA*, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- St-Yves, G., Allen, E. J., Wu, Y., Kay, K., & Naselaris, T. (2022). *Brain-optimized neural networks learn non-hierarchical models of representation in human visual cortex*. bioRxiv. <https://doi.org/10.1101/2022.01.21.477293>
- 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>
- van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*, 111(40), 14332-14341. <https://doi.org/10.1073/pnas.1402773111>
- Willeke, K. F., Restivo, K., Franke, K., Nix, A. F., Cadena, S. A., Shinn, T., Nealley, C., Rodriguez, G., Patel, S., Ecker, A. S., Sinz, F. H., & Tolias, A. S. (2023). *Deep learning-driven characterization of single cell tuning in primate visual area V4 unveils topological organization*. bioRxiv. <https://doi.org/10.1101/2023.05.12.540591>
- 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>

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). *Understanding neural networks through deep visualization*. arXiv. <https://doi.org/10.48550/arXiv.1506.06579>

## Appendix

### I. Layer Commitment Data<sup>5</sup>

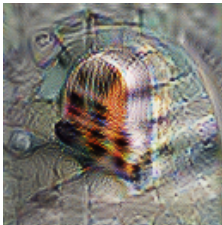


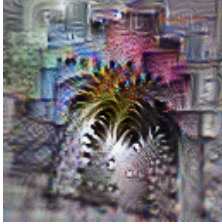
Model	Brain Area	CNN Layer
ResNet 152 V2		
	V1	resnet_v2_152/block3/unit_32/bottleneck_v2
	V2	resnet_v2_152/block3/unit_30/bottleneck_v2
	V4	resnet_v2_152/block3/unit_3/bottleneck_v2
	IT	resnet_v2_152/block4/unit_1/bottleneck_v2
MobileNet V1 0.25 128		
	V1	Conv2d_7_pointwise
	V2	Conv2d_7_depthwise
	V4	Conv2d_5_depthwise
	IT	Conv2d_12_depthwise
MobileNet V1 0.25 160		
	V1	Conv2d_6_pointwise
	V2	Conv2d_6_pointwise
	V4	Conv2d_7_depthwise
	IT	Conv2d_12_depthwise
MobileNet V1 0.25 192		
	V1	Conv2d_6_pointwise
	V2	Conv2d_8_depthwise
	V4	Conv2d_8_pointwise

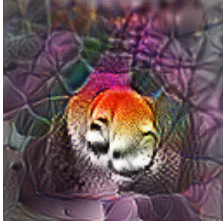
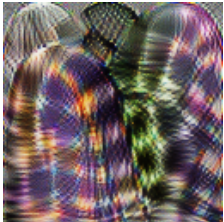
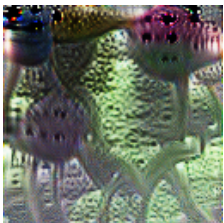
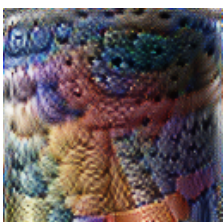
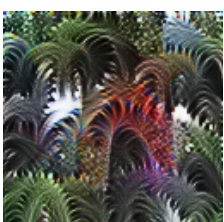
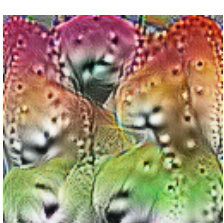
<sup>5</sup> M. Schrimpf, personal communication, 26.04.2023



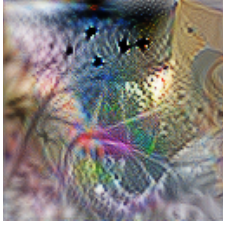



	IT	Conv2d_13_depthwise
MobileNet V1 0.25 224		
	V1	Conv2d_7_depthwise
	V2	Conv2d_6_depthwise
	V4	Conv2d_7_pointwise
	IT	Conv2d_13_depthwise


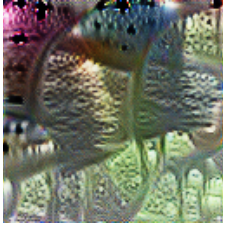
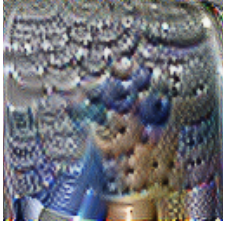
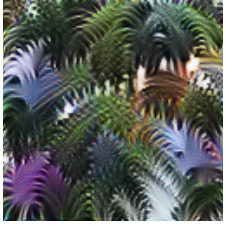
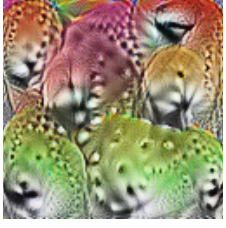



## II. Experimental Materials

Model (*as implemented in the TensorFlow Slim library ( <a href="https://github.com/tensorflow/models/tree/master/research/slim">https://github.com/tensorflow/models/tree/master/research/slim</a> ))	Layer	Objective	Unit	Visualization
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	0	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	768	

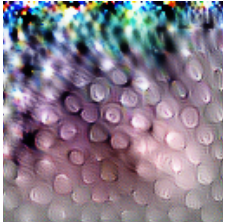
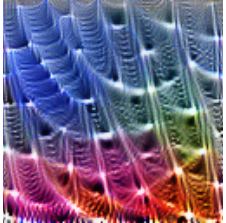
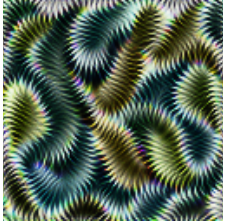



ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	neuron	1023	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	0	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	768	
ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	channel	1023	




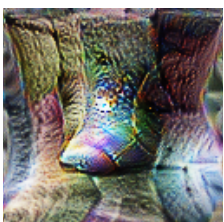


ResNet 152 V2*	resnet_v2_152/block3/unit_32 /bottleneck_v2/add	layer	-	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	0	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	768	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	neuron	1023	


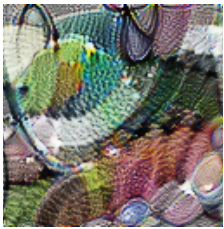
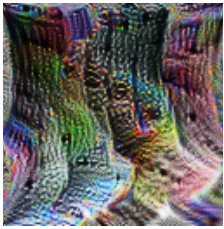
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	0	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	768	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	channel	1023	
ResNet 152 V2*	resnet_v2_152/block3/unit_30 /bottleneck_v2/add	layer	-	

ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	neuron	0	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	neuron	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	neuron	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	neuron	768	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	neuron	1023	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	channel	0	

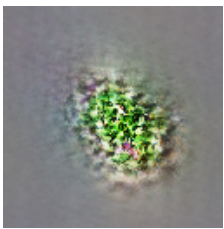
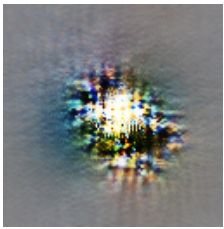
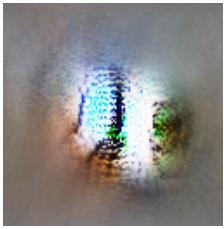
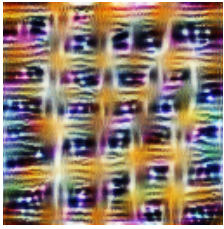

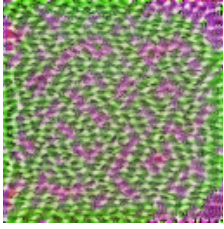


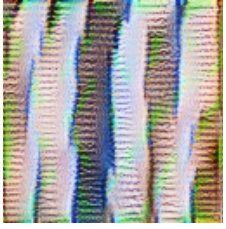
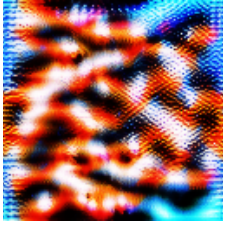
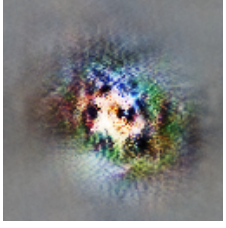
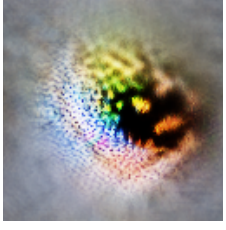
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	channel	256	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	channel	512	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	channel	768	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	channel	1023	
ResNet 152 V2*	resnet_v2_152/block3/unit_3/ bottleneck_v2/add	layer	-	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	neuron	0	

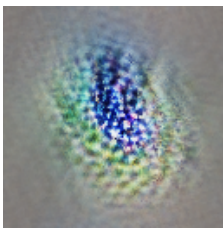
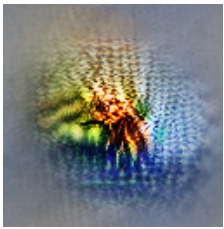
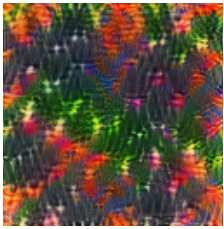
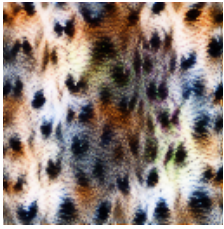

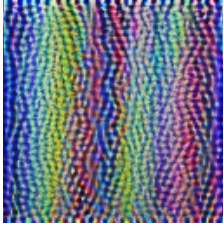
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	neuron	256	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	neuron	512	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	neuron	768	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	neuron	1023	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	channel	0	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	channel	256	

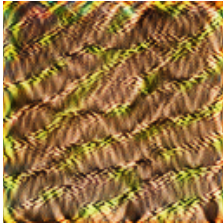
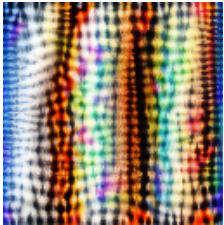
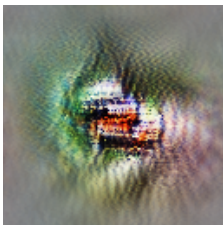
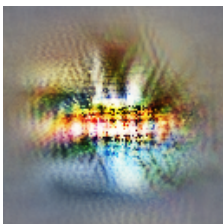
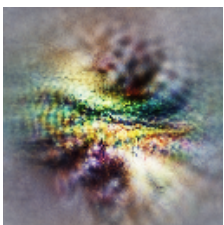
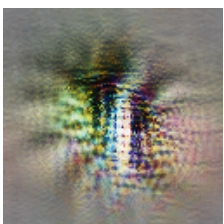
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	channel	512	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	channel	768	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	channel	1023	
ResNet 152 V2*	resnet_v2_152/block4/unit_1/ bottleneck_v2/add	layer	-	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	neuron	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	neuron	32	


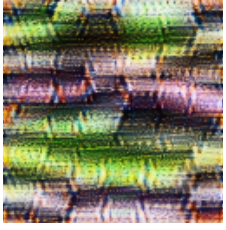
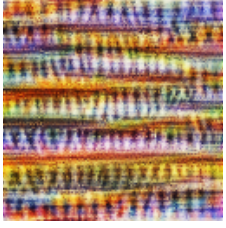
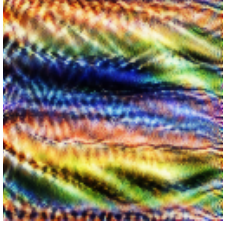
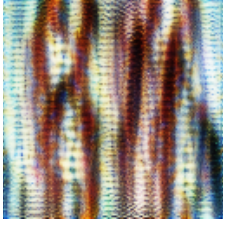
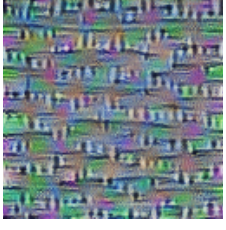


MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	neuron	64	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	neuron	96	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	neuron	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	channel	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	channel	32	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	channel	64	

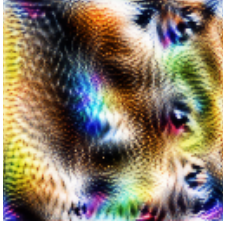
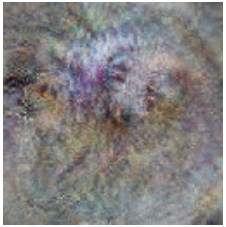
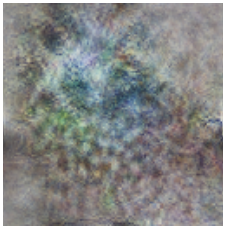
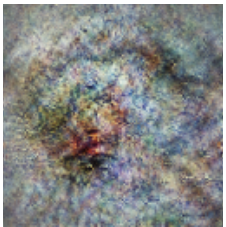
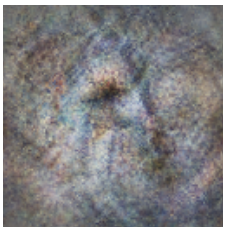
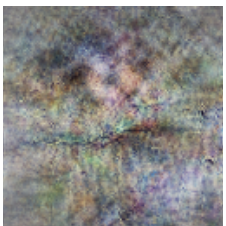
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	channel	96	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	channel	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_6_pointwise/Relu6	layer	-	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	neuron	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	neuron	32	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	neuron	64	

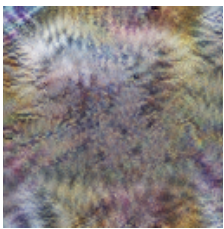
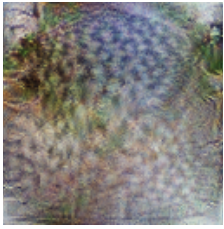
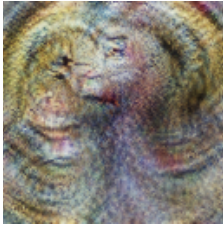
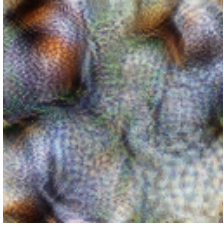
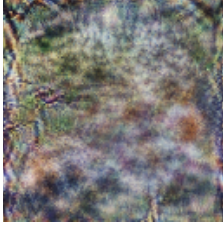
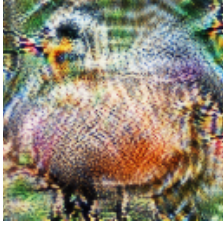
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	neuron	96	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	neuron	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	channel	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	channel	32	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	channel	64	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	channel	96	

MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	channel	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_7_pointwise/Relu6	layer	-	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	neuron	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	neuron	256	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	neuron	512	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	neuron	768	

MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	neuron	1023	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	channel	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	channel	256	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	channel	512	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	channel	768	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	channel	1023	



MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_8_pointwise/Relu6	layer	-	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	neuron	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	neuron	64	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	neuron	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	neuron	192	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	neuron	256	

MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	channel	0	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	channel	64	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	channel	128	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	channel	192	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	channel	256	
MobileNet V1 0.25*	MobilenetV1/MobilenetV1/C onv2d_13_pointwise/Relu6	layer	-	

### III. Licence

#### **Non-exclusive licence to reproduce the thesis and make the thesis public**

I, Carolin Lüübek,

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

“Investigating Psychedelic Imagery through Convolutional Neural Networks’ Feature Visualization”,

supervised by Jaan Aru.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons’ intellectual property rights or rights arising from the personal data protection legislation.

*Carolin Lüübek*  
**14/08/2023**