Title

Get more from less: Differential neural decoding for effective reconstruction of perceived naturalistic stimuli from noisy and scarce neural data

Permalink

https://escholarship.org/uc/item/7vt4w09r

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Dado, Thirza

Güçlü, Umut

Publication Date

2024

Peer reviewed

Get more from less: Differential neural decoding for effective reconstruction of perceived naturalistic stimuli from noisy and scarce neural data

Thirza Dado[1] **(thirza.dado@donders.ru.nl)**　　　**Umut Güçlü**[1] **(u.guclu@donders.ru.nl)**

[1] Neural Coding Lab,
Donders Institute for Brain, Cognition and Behaviour,
Radboud University, Nijmegen, Netherlands

Abstract

Decoding naturalistic stimuli from neural recordings is a significant challenge in systems neuroscience, primarily due to the high-dimensional and nonlinear nature of stimulus-response interactions, and is further exacerbated by the limited availability and noisiness of neural data. While contemporary approaches that incorporate generative models, such as Generative Adversarial Networks (GANs), attempt to address these issues by mapping neural responses to latent representations, they do not fully overcome these obstacles. In this work, we present a novel paradigm of differential neural decoding (**dicoding**) that focuses on the relative changes in response patterns, which not only expands the neural training data quadratically but also inherently denoises it. To determine the corresponding stimulus changes, this method leverages the Euclidean and feature-disentangled properties of the underlying latents through vector arithmetic. As such, we not only effectively exploit the latent space but also achieve semantically meaningful latent offsets in the context of the stimuli, resulting in improved sample efficiency. We trained a decoder to predict changes in latent vectors based on the corresponding changes in neural responses. The absolute latent vector itself was derived by vector addition of the predicted latent change (indicative of stimulus shift) to a reference latent, which was fed to the generator for the reconstruction of the perceived stimulus. Our results show that this geometrically principled approach facilitates more effective reconstruction of naturalistic stimuli from noisy and limited neural data.

Keywords: generative adversarial networks; neural decoding; latent space geometry; reconstruction

Introduction

Understanding the representation of sensory stimuli in neural activity patterns is a crucial yet challenging aspect of neural decoding — a framework of computational methods that seek to extract meaningful information from recorded neural responses (Naselaris, Kay, Nishimoto, & Gallant, 2011; M. A. van Gerven, Seeliger, Güçlü, & Güçlütürk, 2019). These methods include classification (Kamitani & Tong, 2005; Horikawa & Kamitani, 2017), identification (Mitchell et al., 2008; Kay, Naselaris, Prenger, & Gallant, 2008) and reconstruction (Thirion et al., 2006; Miyawaki et al., 2008; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; M. van Gerven, de Lange, & Heskes, 2010; Nishimoto et al., 2011; Schoenmakers, Barth, Heskes, & Van Gerven, 2013; Cowen, Chun, & Kuhl, 2014; Du, Du, & He, 2017; Güçlütürk et al., 2017; Shen, Horikawa, Majima, & Kamitani, 2019; VanRullen & Reddy, 2019; Dado et al., 2022) of perceived stimuli. While these methods establish *direct* mappings from neural recordings to stimulus features, their practical application encounters significant obstacles. First, these approaches rely on extensive paired datasets of stimuli and their corresponding neural responses. Such datasets are scarce and expensive to obtain at scale. The intrinsic limitation of neural recordings stems from the complex challenge of interfacing with the brain in a manner that is both safe and stable. As such, collecting sufficient labeled samples to train high-dimensional decoders is infeasible in most experimental settings. Second, biological sensory processing consists of complex and non-linear transformations which makes it difficult to learn robust decoding functions that can reliably interpret neural signals. Third, substantial trial-to-trial variability and noise in neural responses to identical stimuli confound learning consistent mappings.

A prevalent method to mitigate these issues is to leverage the latent space of high-dimensional generative models which have been pretrained on huge amounts of data (Seeliger, Güçlü, Ambrogioni, Güçlütürk, & van Gerven, 2018; Han et al., 2019; VanRullen & Reddy, 2019; Mozafari, Reddy, & VanRullen, 2020; Dado et al., 2022, 2024). Specifically, a decoder embeds neural responses as points in the Euclidean and feature-disentangled latent space such that the responses are aligned with the latents of the stimuli. Subsequently, photorealistic images are synthesized by passing these decoded latent codes through the pretrained generator (Figure 1). The inductive biases captured in the latent space provide useful regularization of the decoding process, but this strategy does not fully overcome the aforementioned obstacles because latents and responses do not perfectly align. As such, learning a consistent one-to-one mapping from noisy (potentially low-dimensional) neural responses to complex, high-dimensional stimuli remains fundamentally challenging.

To overcome these issues, we propose a new framework for DIfferential neural deCODING (**dicoding**) that focuses on reconstructing stimuli from *relative* differences between pairs of neural responses instead of directly predicting stimuli or latents from the *absolute* neural responses themselves[1]. This shift to decoding the response comparisons rather than

---

[1] Note that by 'absolute', we do not imply 'only positive values'; rather, we use the term to indicate that we utilize the data points themselves, independent of their comparative context or relationship to other data points.
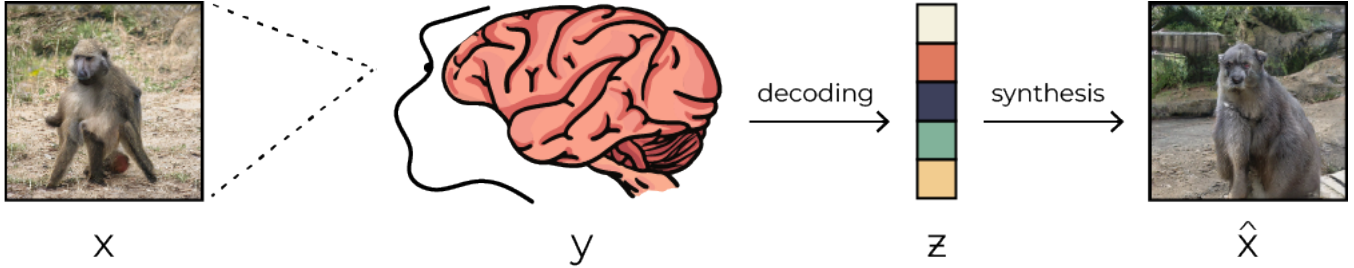
Figure 1: **Neural reconstruction.** A visual stimulus, X, is presented while recording the stimulus-evoked neural response, y, thereto. Neural decoding then maps the recorded response back to an image in two steps: *decoding* is the linear transformation from y to feature representation, z, and *synthesis* the nonlinear transformation from z to the reconstructed image, $\hat{X}$.

absolute values results in a quadratic increase in usable data points since for each neural response, we compute its difference with every other response in the dataset, transitioning from n individual data points to $n(n-1)$ relative data pairs.

Concretely, we start with a dataset of visual stimuli, $\{X\}$, paired with their corresponding neural responses, $\{Y\}$. For each combination of two neural responses (y_i, y_j) in the dataset, we calculate the difference in two directions: $\{\delta y_{ij} = y_i - y_j\}$ and $\{\delta y_{ji} = y_j - y_i\}$. Our decoder model is then trained to associate these differential neural responses with their respective changes in the latent space, denoted as δz_{ij} and δz_{ji} that capture the semantic and perceptual shifts between stimuli. To reconstruct a specific stimulus x_k, we choose a reference point (x_i, y_i) with known latent representation z_i and compute the response differential $y_{ik} = y_i - y_k$. The decoder then predicts the corresponding latent shift δz_{ik}. With this predicted latent offset, we synthesize the reconstructed stimulus x_k by applying the transformation $x_k = G(z_i - \delta z_{ik})$, where G is the pretrained generator.

Importantly, we posit that by translating differences in neural responses into latent offsets, we achieve three key advantages: enhanced noise robustness, more effective utilization of latent spaces, and improved sample efficiency. This is supported by empirical validation with neural datasets, confirming its effectiveness, particularly for lower amounts of training data. As such, our approach presents a meaningful contribution to the processing and interpretation of limited biological neural data.

Theoretical motivation

Here, we employ simplifying assumptions about the statistical properties of neural noise and geometric structure of the latent space to gain clearer insights into the underlying mechanics and potential of our method.

1. Noise robustness. We model neural noise as independent, additive Gaussian noise with a covariance matrix $\Sigma = \sigma^2 I$. Consider a fixed decoding mapping $f()$, and two stimuli x_i and x_j with corresponding latent vectors z_i and z_j. The neural responses can be expressed as:

$$y_i = f(z_i) + \varepsilon_i$$
$$y_j = f(z_j) + \varepsilon_j \tag{1}$$

where $\varepsilon_i, \varepsilon_j \sim \mathcal{N}(0, \sigma^2 I)$ represent independent noise components.

Importantly, the differential response $\delta y_{ij} = y_i - y_j$ retains this additive structure:

$$\delta y_{ij} = \delta z_{ij} + \delta \varepsilon_{ij} \tag{2}$$

where $\delta \varepsilon_{ij} = \varepsilon_i - \varepsilon_j$ has a variance of $2\sigma^2 I$, doubling the individual noise variance. However, the quadratic increase in data pairs — from n to $n(n-1)$ — leads to a superlinear improvement of the signal-to-noise ratio. This substantial increase significantly bolsters the robustness of decoding under noisy and limited conditions. Essentially, an implicit denoising mechanism is incorporated into the decoding process through this asymmetrical scaling: the volume of data increases quadratically while the associated noise only grows linearly. We amplify this denoising effect even further by averaging the latent offset predictions across all possible pairs in the dataset. Consequently, **dicoding** inherently exhibits greater robustness to inter-trial variability and uncorrelated noise, unlike traditional approaches that rely on absolute responses.

2. Latent space. We assume that the generator's latent space is an isotropic Euclidean metric space, characterized by orthogonal, feature-disentangled axes, each corresponding to distinct perceptual features. Feature disentanglement allows for the independent manipulation of specific features (e.g., shape or color) without inadvertently affecting other attributes. Utilizing the Euclidean geometry, we can achieve relative changes in stimuli — corresponding to the relative changes in neural responses — through vector arithmetic applied to their underlying latent representations. In this idealized setting, the latent difference vector δz_{ij} precisely captures the stimulus difference δx_{ij} without inter-feature interactions. As such, **dicoding** maximizes the utilization of the latent space's inherent properties to represent
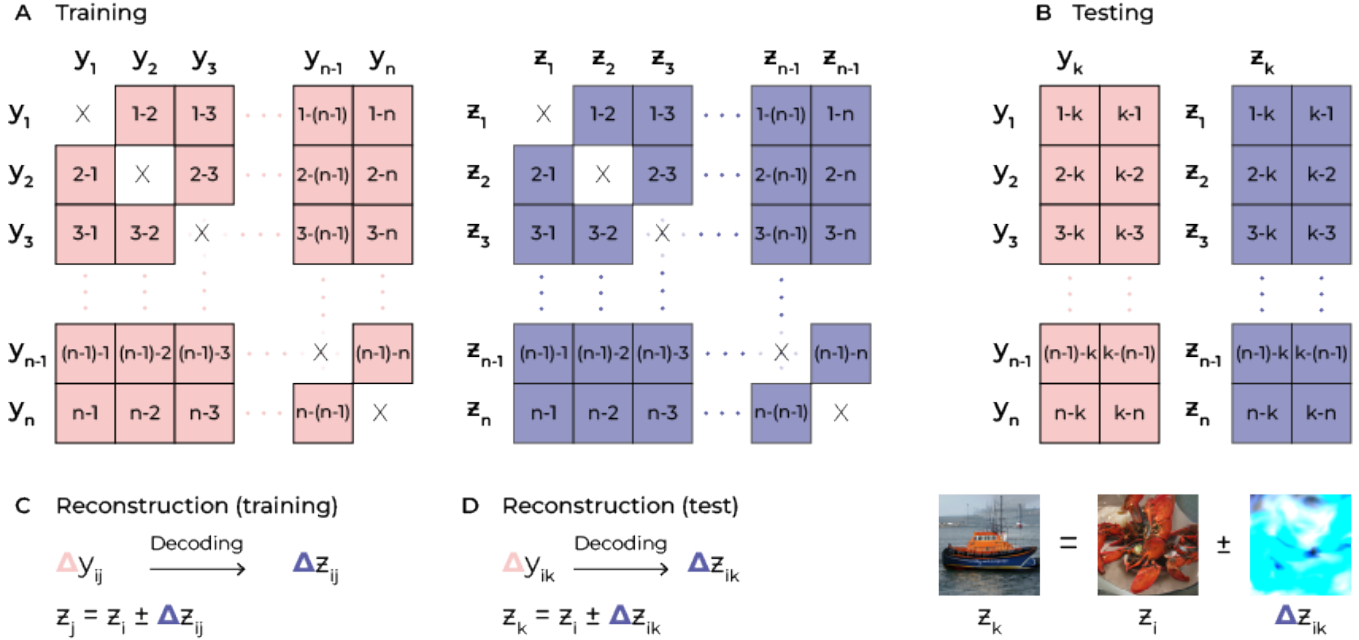
Figure 2: **DIfferential neural deCODING (dicoding).** **A.** The training set is constructed by taking the relative differences between pairs of training example. In the left matrix, each entry represents the difference between a neural response to one stimulus y_i and a neural response to every other example y_j, capturing all possible combinations. Meanwhile, the right matrix similarly captures the differences between latent representations of visual stimuli. Note that the diagonal differences are excluded since the difference between an arbitrary example with itself equals zero. Therefore, our total dataset will have $n(n-1)$ training points. **B.** The test set is constructed by taking the relative difference between a test set example k and an arbitrary other example. Here, we take the difference between k and n training set examples. **C.** To reconstruct a training stimulus, X_i, its corresponding latent representation, z_i, can be re-obtained by adding/subtracting the relative difference, δz_{ij}, from the other latent representation, z_j. This latent, z_i, is then fed to the generator for reconstruction of the stimulus. **D.** To reconstruct a test stimulus, X_k, the latent representation, z_k, corresponding to this unobserved stimulus can be obtained by adding or subtracting the predicted relative difference, δz_{ik}, from the observed latent representation, z_i, which is then fed to the generator for image reconstruction.

variations in stimuli. This enables efficient exploration of the stimulus landscape, allowing us to navigate the complex web of potential stimuli with straightforward mathematical operations, which naturally aligns with the generator's functional design.

3. Sample efficiency. We assume that the latent space exhibits proportionality, meaning the magnitude of latent difference vectors δz_{ij} scales linearly with stimulus differences δx_{ij}. This proportional relationship is key; when the training set size doubles from n to $2n$, we do not only increase the number of examples but also augment the semantic diversity captured by the latent differences. This broader coverage of potential stimulus variations leads to a more robust and effective training process.

Under these simplified assumptions, **dicoding** emerges as a remarkably effective method that demonstrates enhanced robustness against noise, strategic utilization of latent space geometry and improved sample efficiency. While these theoretical insights underscore the method's potential

benefits, they also set the stage for further empirical validation, acknowledging that real neural systems may differ from these idealized models.

Methods

Dataset

We reanalyzed two datasets of visual stimuli (faces and natural images) and corresponding neural responses from (Dado et al., 2024). These stimuli were generated using Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) which made their underlying w-latent representations (required for the generation process) directly accessible for neural decoding. To ensure diversity, each training example was presented once, while test examples were averaged over twenty repetitions to enhance the signal-to-noise ratio.

- *Face images.* StyleGAN3 (Karras et al., 2021) pretrained on the Flickr-Faces-HQ (FFHQ) dataset (Karras, Laine, & Aila, 2019) was used to generate 4000 and 100 training and test set images (1024^2 pixels), respectively.
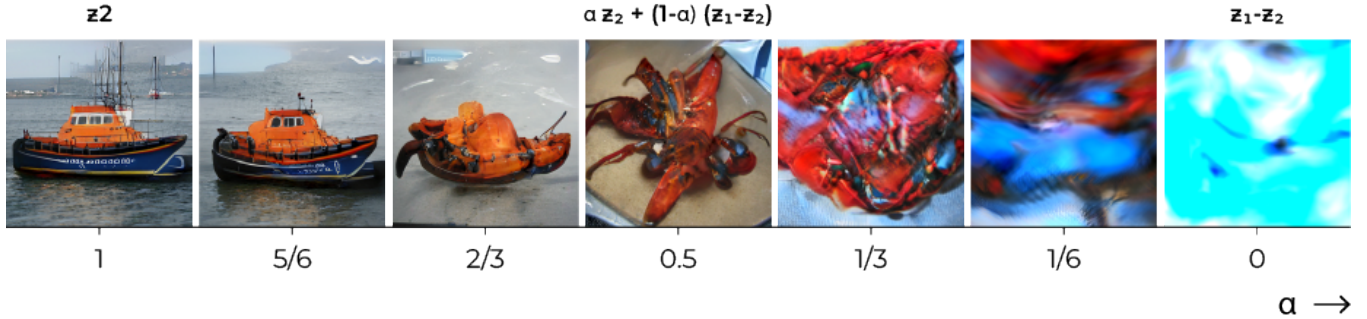
Figure 3: **Relative differences in latent space.** This sequence of images illustrates the linear interpolation from latent representations z_2 to the relative difference $z_1 - z_2$. As we gradually add the relative difference to z_2, the images transition from the starting point z_2, through the intermediate representation z_1 (since $z_2 + z_1 - z_2 = z_1$), and eventually culminate in the (inherently meaningless) latent direction corresponding to $z_1 - z_2$. As such, it provides intuition about the information encoded in the relative difference between two latent representations.

- *Natural images.* StyleGAN-XL (Sauer, Schwarz, & Geiger, 2022) pretrained on ImageNet (Deng et al., 2009) was used to generate 4000 and 200 training and test set images (512^2 pixels), respectively. For this, a subset of categories was used, namely the 200 classes from Tiny ImageNet (Le & Yang, 2015). Each category was represented by twenty training images and one test image.

In a passive fixation experiment, the images were resized to 500^2 pixels and presented to a macaque with 15 implanted multi-unit electrode arrays (64 units each) to record the multi-unit activity (MUA) (Super & Roelfsema, 2005), with seven arrays were positioned in primary visual cortex (V1), four arrays in area V4 and four arrays in inferior temporal cortex (IT).

Neural decoding

Decoding Neural decoding involves the process of learning a mapping function, f, that predicts the stimulus properties from its recorded neural response. The direct response-stimulus transformation can be divided into two distinct stages. In the first stage, we train a dense layer to decode a neural response, y_i, into a latent representation, z_i, by minimizing the Euclidean loss function:

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^{N} \left(z_i - \mathbf{w}^T \mathbf{y}_i\right)^2 \quad (3)$$

where i ranges over the samples.

In the second stage, the latent representations obtained from the linear model are nonlinearly transformed into images. This transformation is facilitated by the pretrained generator of the GAN which maps the 512-dimensional latent code z_i to a photorealistic image x_i.

Dicoding Given a dataset of neural responses $Y = \{y_i\}$ and latent representations $Z = \{z_i\}$, we first compute all pairwise differences in both directions for each pair of data examples:

$$\delta y_{ij} = y_i - y_j, \quad \delta y_{ji} = y_j - y_i, \quad \forall i, j \text{ with } i \neq j$$

and similarly for the latent representations:

$$\delta z_{ij} = z_i - z_j, \quad \delta z_{ji} = z_j - z_i, \quad \forall i, j \text{ with } i \neq j$$

where i and j range over all indices in the datasets.

Next, as with regular decoding (described above), we perform neural decoding in two stages. In the first stage, we decode a response difference, δy_{ij}, into latent difference, δz_{ij}, by optimization of a dense layer using the following loss function:

$$\mathcal{L} = \frac{1}{4} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \left(\left(z_{ij} - \mathbf{w}^T \mathbf{y}_{ij}\right)^2 + \left(z_{ji} - \mathbf{w}^T \mathbf{y}_{ji}\right)^2\right) \quad (4)$$

where i ranges over the samples.

The second stage remains the same as for regular decoding and thus requires the absolute latent representation. For this, it is important that the unobserved latent, z_k, is paired with an observed reference latent, z_i, so that after predicting the latent difference $\delta z_{ik} = f(\delta y_{ik})$, we can obtain the reconstructed stimulus $\hat{x}_k = G(z_i - \delta z_{ik})$ where G represents the pretrained generator of the GAN. Finally, averaging predictions across multiple reference pairs can provide greater robustness.

Implementation details

Analyses were performed on a cloud VM with Python 3.8, Intel Xeon CPU, NVIDIA Tesla T4 GPU, and Linux OS.

Results

To validate the effectiveness of our **dicoding** approach compared to conventional decoding using absolute values, we trained a series of **dicoders** and regular decoders on increasingly larger subsets of data. Specifically, we utilized
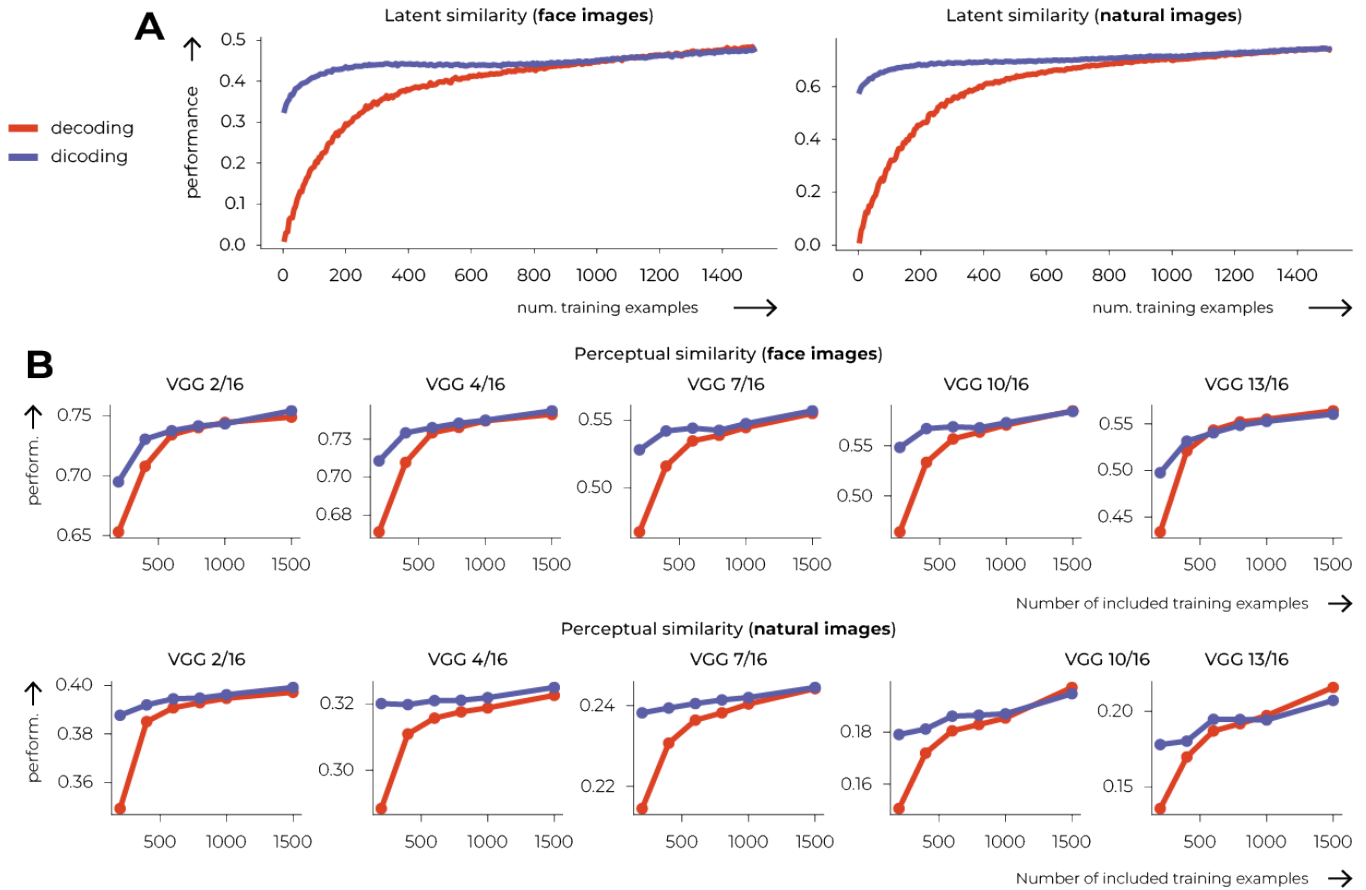
Figure 4: Decoding performance (Y) as a function of number of training examples (X).

training sets ranging from as few as 5 examples up to a maximum of 1500 examples, with increments of 5, resulting in a total of 300 **dicoders** and 300 decoders. For the natural images dataset, we ensured exposure to a wide variety of image categories. That is, when the training set size, N, equaled 200, the model was trained using the first image from each of the 200 categories. As the size of the training set increased to $N = 400$, the model was also trained on each second image from each category. So, as the training set expanded, the model was progressively exposed to a more diverse array of stimuli.

We evaluated decoding performance by measuring the cosine similarity between the predicted and target latent vectors on a held-out test set, which comprised 100 images for the faces dataset and 200 images for the natural images dataset. We then averaged the evaluation metrics across these five permutations to enhance the robustness of our performance assessment. Our results indicated that **dicoding** outperformed conventional decoding for lower amounts of training data but their performances converged as the training dataset size increased (Figure 4A). Subsequently, we selected the four **dicoder** and decoder models that were trained on 200, 400, 600, and 800 training faces and images and fed their predicted latents to the generator for the reconstruction

of the corresponding images (Figure 5). Visual inspection revealed that **dicoding** already produced reconstructions that closely resembled the original stimuli for lower amounts of training data. We quantitatively evaluated the predictions of the models trained on 200, 400, 600, 800, 1000 and 1500 training examples based on perceptual similarity using five activations of VGG16 for face recognition (Parkhi, Vedaldi, & Zisserman, 2015) for faces and VGG16 for object recognition (Simonyan & Zisserman, 2014) for natural images. This analysis also confirmed the superior performance of **dicoding** for lower sample sizes (Figure 4B).

Discussion

In this work, we introduced **dicoding** - a novel approach in differential neural decoding that integrates relative sensory processing with generative modeling. By shifting its focus from absolute to relative neural responses, this approach significantly enhanced the reconstruction of perceptual stimuli through quadratic data expansion from limited neural data. We posited that this method would not only demonstrate enhanced robustness against noise but also maximize the exploitation of the latent space's geometric properties and leverage the inherent semantically meaningful directional information thereof. Subsequent empirical analyses validated

Face images

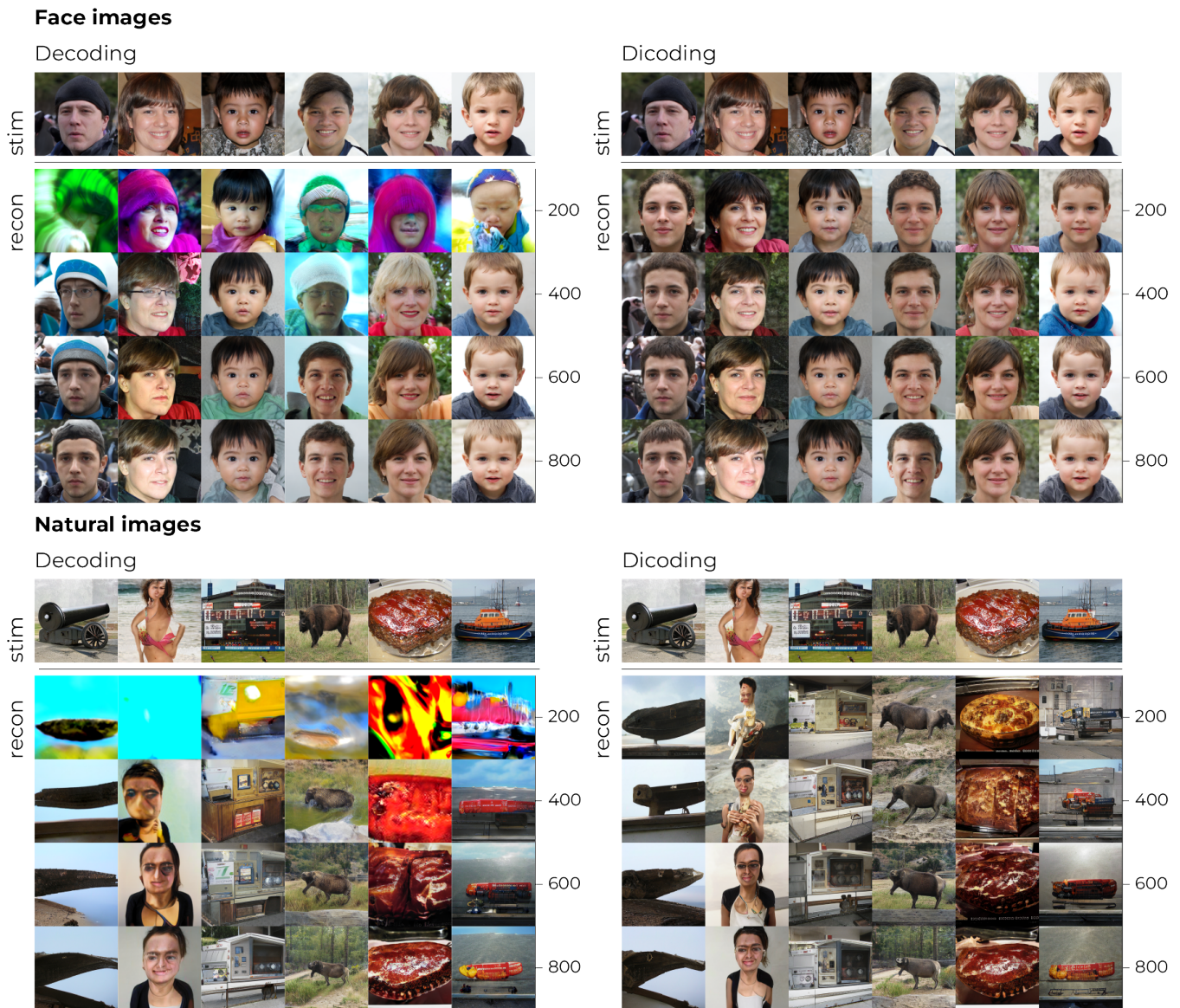Decoding Dicoding

Natural images

Decoding Dicoding

Figure 5: Stimuli and reconstructions from brain activity using decoders and **dicoders** trained on different subset amounts.

advantages in practice for lower amounts of training data, confirming its effectiveness in scenarios where neural data is noisy and scarce — a common challenge due to the difficulty and expense of data collection.

It is important to note that as the size of the training dataset increased, the performances of differential and regular decoding converged. This suggests that when substantial neural data is available, relying on absolute values may be more advantageous as they contribute uniquely to the learning process without the computational overhead of generating differential pairs and managing the expanded dataset. Nevertheless, the benefits of our approach are clear. Future research should concentrate on developing robust inversion techniques that enable the conversion of photographs to their corresponding latent representations,

extending its applicability to a broader range of real-world data. By doing so, differential decoding could help enable the next generation of neural interfaces and prosthetic devices through more effective utilization of limited biological data.

To contextualize **dicoding** within the broader spectrum of sensory processing, we note that it employs principles that resonate with some aspects of how the brain processes sensory information relative to a baseline or context, such as sensory adaptation and contrast effects. However, as a computational strategy, differential decoding is fundamentally different. It serves specific objectives and operates within a distinct context, separate from the ongoing, adaptive processes observed in biological systems that directly respond to sensory inputs as they occur, without artificially expanding data.

References

Cowen, A. S., Chun, M. M., & Kuhl, B. A. (2014). Neural portraits of perception: reconstructing face images from evoked brain activity. *NeuroImage*, *94*, 12–22.

Dado, T., Güçlütürk, Y., Ambrogioni, L., Ras, G., Bosch, S., van Gerven, M., & Güçlü, U. (2022). Hyperrealistic neural decoding for reconstructing faces from fmri activations via the gan latent space. *Scientific reports*, *12*(1), 1–9.

Dado, T., Papale, P., Lozano, A., Le, L., Wang, F., van Gerven, M., ... Güçlü, U. (2024). Brain2gan: Feature-disentangled neural encoding and decoding of visual perception in the primate brain. *PLOS Computational Biology*, *20*(5), e1012058.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

Du, C., Du, C., & He, H. (2017). Sharing deep generative representation for perceived image reconstruction from human brain activity. In *2017 international joint conference on neural networks (ijcnn)* (pp. 1049–1056).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Güçlütürk, Y., Güçlü, U., Seeliger, K., Bosch, S., van Lier, R., & van Gerven, M. (2017). Reconstructing perceived faces from brain activations with deep adversarial neural decoding. *Advances in neural information processing systems*, *30*.

Han, K., Wen, H., Shi, J., Lu, K.-H., Zhang, Y., Fu, D., & Liu, Z. (2019). Variational autoencoder: An unsupervised model for encoding and decoding fmri activity in visual cortex. *NeuroImage*, *198*, 125–136.

Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, *8*(1), 1–15.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679–685.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021). Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, *34*.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 4401–4410).

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355.

Le, Y., & Yang, X. (2015). Tiny imagenet visual recognition challenge. *CS 231N*, *7*(7), 3.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., ... Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, *60*(5), 915–929.

Mozafari, M., Reddy, L., & VanRullen, R. (2020). Reconstructing natural scenes from fmri patterns using bigbigan. In *2020 international joint conference on neural networks (ijcnn)* (pp. 1–8).

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fmri. *Neuroimage*, *56*(2), 400–410.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology*, *21*(19), 1641–1646.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015, September). Deep face recognition. In X. Xie, M. W. Jones, & G. K. L. Tam (Eds.), *Proceedings of the british machine vision conference (bmvc)* (p. 41.1-41.12). BMVA Press. Retrieved from https://dx.doi.org/10.5244/C.29.41 doi: 10.5244/C.29.41

Sauer, A., Schwarz, K., & Geiger, A. (2022). Stylegan-xl: Scaling stylegan to large diverse datasets. In *Acm siggraph 2022 conference proceedings* (pp. 1–10).

Schoenmakers, S., Barth, M., Heskes, T., & Van Gerven, M. (2013). Linear reconstruction of perceived images from human brain activity. *NeuroImage*, *83*, 951–961.

Seeliger, K., Güçlü, U., Ambrogioni, L., Güçlütürk, Y., & van Gerven, M. (2018). Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage*, *181*, 775–785.

Shen, G., Horikawa, T., Majima, K., & Kamitani, Y. (2019). Deep image reconstruction from human brain activity. *PLoS computational biology*, *15*(1), e1006633.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Super, H., & Roelfsema, P. R. (2005). Chronic multiunit recordings in behaving animals: advantages and limitations. *Progress in brain research*, *147*, 263–282.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage*, *33*(4), 1104–1116.

van Gerven, M., de Lange, F. P., & Heskes, T. (2010). Neural decoding with hierarchical generative models. *Neural computation*, *22*(12), 3127–3142.

van Gerven, M. A., Seeliger, K., Güçlü, U., & Güçlütürk, Y. (2019). Current advances in neural decoding. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 379–394.

VanRullen, R., & Reddy, L. (2019). Reconstructing faces from fmri patterns using deep generative neural networks. *Communications biology*, 2(1), 1–10.