Brain2Image: Converting Brain Signals into Images

I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano
PeRCeiVe Lab - University of Catania
Via Santa Sofia, 102
Catania, Italy 95127
www.perceivelab.com

M. Shah
Center for Research in Computer Vision - University of Central Florida
4328 Scorpius St.
Orlando, USA 32816-2365
http://crcv.ucf.edu

ABSTRACT

Reading the human mind has been a hot topic in the last decades, and recent research in neuroscience has found evidence on the possibility of decoding, from neuroimaging data, how the human brain works. At the same time, the recent rediscovery of deep learning combined to the large interest of scientific community on generative methods has enabled the generation of realistic images by learning a data distribution from noise. The quality of generated images increases when the input data conveys information on visual content of images. Leveraging on these recent trends, in this paper we present an approach for generating images using visually-evoked brain signals recorded through an electroencephalograph (EEG). More specifically, we recorded EEG data from several subjects while observing images on a screen and tried to regenerate the seen images. To achieve this goal, we developed a deep-learning framework consisting of an LSTM stacked with a generative method, which learns a more compact and noise-free representation of EEG data and employs it to generate the visual stimuli evoking specific brain responses.

Our *Brain2Image* approach was trained and tested using EEG data from six subjects while they were looking at images from 40 ImageNet classes. As generative models, we compared variational autoencoders (VAE) and generative adversarial networks (GAN). The results show that, indeed, our approach is able to generate an image drawn from the same distribution of the shown images. Furthermore, GAN, despite generating less realistic images, show better performance than VAE, especially as concern sharpness. The obtained performance provides useful hints on the fact that EEG contains patterns related to visual content and that such patterns can be used to effectively generate images that are semantically coherent to the evoking visual stimuli.

MM '17, October 23–27, 2017, Mountain View, CA, USA

ACM ISBN 978-1-4503-4906-2/17/10…$15.00
https://doi.org/10.1145/3123266.3127907

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Learning latent representations**;

KEYWORDS

Variational autoencoder, EEG, image generation

ACM Reference format:
I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano and M. Shah. 2017. *Brain2Image*: Converting Brain Signals into Images. In *Proceedings of MM '17, Mountain View, CA, USA, October 23–27, 2017,* 9 pages.
https://doi.org/10.1145/3123266.3127907

1 INTRODUCTION

The idea of using the mind to control machines is not new and has been long investigated, with fair success especially in the brain computer interface research field [7, 8, 16, 24], which mainly aims to decode simple patterns for direct-actuated control of machines. Beside BCI research, brain signals have been also employed to drive the learning of intelligent systems for emotion classification [14], medical tasks [1], etc. The common factor between these approaches is that they all attempt to learn a latent space to perform "simple" classification between few mind states. While it is relatively easy and well-documented to identify brain patterns related to audio stimuli or associated to specific diseases, it is much more complex to understand what happens in the human brain while performing visual tasks. To this end, neurocognitive studies [12, 20, 21] have found out that brain activity contains detectable patterns related to visual stimuli categories [2, 3, 5, 25, 28] and, recently, in [26] we have demonstrated that, indeed, brain signals recorded through an electroencephalograph (EEG) can be used to understand what humans are seeing. However, although this work reveals important aspects on brain functioning, it is not really a reading-the-mind process as latent features are only employed for automated classification and no means to "see" what humans are seeing is considered.

In this paper we aim at closing the loop by adding a "generative layer", which, starting from a latent space learned from brain signals, generates a meaningful — and human-readable — data representation of it, i.e., *images semantically coherent with the visual stimuli evoking those brain responses*. More specifically, in this paper we propose a *Brain2Image* framework consisting of a discriminative model to learn latent features from brain signals and a generative model, which, starting from the learned manifold, is able to generate visual samples from it. As discriminative model we employed an LSTM and as generative model

we used and compared generative adversarial networks [6] (GAN) and variational encoders [11] (VAE), which have already demonstrated great performance in generating realistic images from an input data distribution. However, training such generative methods, especially GANs, is rather unstable and requires large data – often unavailable in experiments involving human subjects – and here we propose a training strategy to overcome this limitation. The achieved results demonstrate the effectiveness of our approach in generating visual samples semantically coherent with visual stimuli.

The remainder of the paper is as follows: Sect. 2 reports on the literature about methods attempting to reconstruct visual images from neuroimaging data as well as the recent methods to generate realistic images starting from a specific data distribution. Sect. 3 describes the proposed framework, from data acquisition to the discriminative methods for EEG latent variable learning to the generative models for image generation. Sect. 4 reports the achieved performance, while Sect. 5 draws conclusions on the further steps towards the ambitious goal of understanding the cognitive processes behind visual perception.

2 RELATED WORK

Research in neuroscience and neuroimaging [9] has shown that human cognitive processes related to human perception (and visual perception in particular) can be decoded through non-invasive imaging techniques such as fMRI, EEG, MEG. These studies have specifically found evidence about the possibility of decoding what humans are thinking from brain activity. Simple statistical pattern recognition techniques have been put in place to address the challenging goal of reading the mind; while those methods mainly aim at identifying differential brain activity patterns between cognitive states, our work is different as it not only attempts to do that, but it also tries to reconstruct the stimuli that have evoked specific brain responses. One work that is close to ours in terms of objective is [19], where authors propose an approach to estimate what humans are seeing using fMRI images. In particular, the method aims at maximizing the posterior probability of a certain visual stimulus, inducing a specific brain response to belong to a data distribution learned from a large pool of images [18]. In practice, fMRI showed great potential in this mind-generative process, but its main limitation lies in the experimental costs. This drawback is overcome by lower-cost techniques such as electroencephalography, which provide a higher temporal resolution compared to fMRI, but on the other hand, suffer from lower spatial resolution and more noisy data, which make the stimuli reconstruction process more difficult.

Nevertheless, the key aspects for a successful *Brain2Image* process are: 1) the assumption the input brain signals retain information about visual content; 2) the possibility to decode and extract such visual information; 3) a generative model able to use the decoded information and to learn a data distribution using limited and noisy signals. The first two issues are addressed in the recent literature that has demonstrated that visual stimuli elicit detectable changes in EEG brain responses [4, 17] and through recurrent neural networks able to extract such information [26] for being used by machines.

Image generation from a latent feature space is what we want to investigate in this paper. Image generation is an active research topic and with the concomitant advent of deep generative models, two successful and fundamentally diverse methods attracted the attention of the researchers showing promising results: Variational Autoencoders (VAE) [11] and Generative Adversarial Networks (GAN) [6]. The former category of methods follow a straightforward strategy: they use a classical autoencoder (i.e. an encoder/ decoder scheme), but add noise to the intermediate representation in order to impose a Gaussian distribution. This way the decoder learns to generate images not by compressing the input feature space with a lossy non-linear filter (as in "vanilla" autoencoders), but by following specific data feature distributions extracted from the input. As a result, if the input contains one such distribution (identified by the encoder), then the generator outputs an image that contains a visual feature corresponding to that distribution.

The latter category, GANs, adopt a completely different concept: a generator network creates an image starting from noise and a discriminator tries to identify whether an input image is fake or real. The two modules practically engage in a competition where the generator struggles to create realistic images and the discriminator gets better in identifying forged ones.

Both approaches present a number of advantages and disadvantages with respect to each other: VAEs are more intuitive, can be trained easily, make it easier to produce the desired result but are proner to overfitting with small datasets; besides, the output images lack sharpness because of the noise introduction during the intermediate representation. GANs are more novel, they create very sharp-looking images but are unstable to train, especially under imposed conditions, although further enhancements can help to alleviate, but not eliminate, such problems [15, 23].

3 METHOD

The objective of this work is to create an approach able to "translate" visually-evoked EEG signals into meaningful images, as shown in Fig. 1. As mentioned in the previous section, the key aspect to achieve this goal is to be sure that EEG signals encode visual class discriminative information that can be extracted by processing EEG data and this has been demonstrated in [26].

Our *Brain2Image* approach consists of an *encoder*, which aims at identifying a latent feature space for brain signal classification, and a *decoder*, which turns the learned feature into images using a deconvolution approach. In our case, the encoder consists of an LSTM layer, while as decoder we employ both VAE (see Fig. 2) and GAN (Fig. 3). The two generative approaches share the same architecture, i.e.:

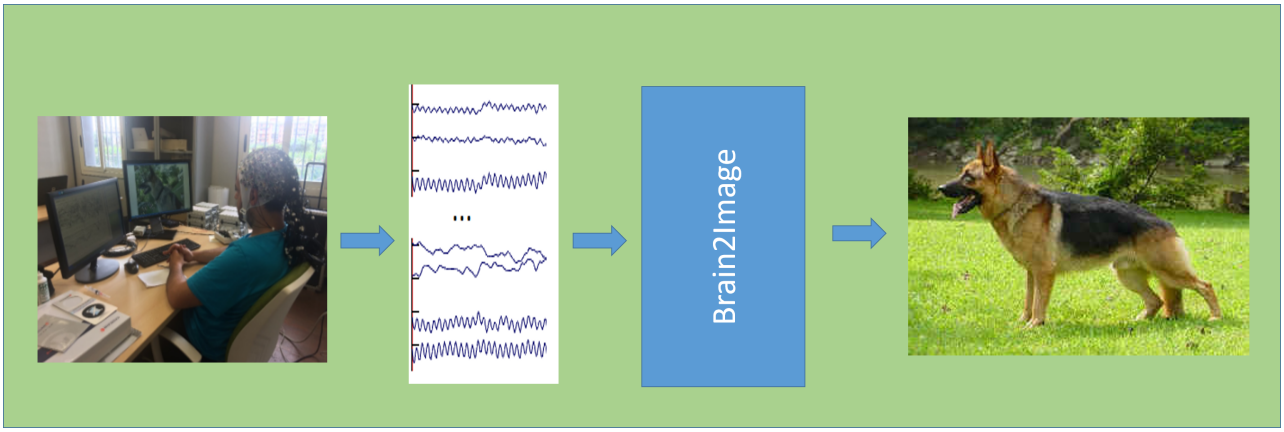


Figure 1: The objective of our work: a subject is looking at an image and at the same time its brain activity is captured by an EEG. The acquired EEG signals are fed to our *Brain2Image* system and translated to an image semantically-similar (ideally the same) to the one that the subject is looking at.

- **EEG data acquisition**: a human subject participating in the experiment looks at images of various object classes at a computer screen while their brain activity is recorded.
- **EEG feature extraction**: the acquired signals are processed by an encoder, which is trained to output a feature vector (*EEG features*), containing class-discriminative information.
- **EEG-conditioned image generation**: generative models (decoder in VAE or a generator–discriminator pair in GAN) are trained to produce images from the EEG feature vectors.

The data acquisition and EEG feature extraction parts have been largely described in [26], so here are briefly overviewed, while we mainly concentrate on the image generation part, i.e., how we can generate images starting from EEG signals.

3.1 EEG data acquisition

Our experiment involved six subjects who were shown images of objects while EEG data was recorded. As visual stimuli, we employed 50 images from 40 different ImageNet classes (see Tab. 3 for a list of the employed classes) for a total of 2,000 images. Each image class was presented in batches of 25 seconds, followed by a 10 second black screen to "clear" the visual pathway. The total duration of each experiment was 1,400 seconds (23 minutes and 20 seconds). After the EEG data acquisition, we obtained 11,466 128-channel EEG sequences (536 recordings were discarded because they were too short or too altered to be included in the experiment).

A summary of the adopted experimental paradigm is shown in Table 1.

Number of classes	40
Number of images per class	50
Total number of images	2,000
Visualization order	Sequential
Time for each image	0.5 s
Pause time between classes	10 s
Number of sessions	4
Session running time	350 s
Total running time	1,400 s

Table 1: The parameters of the experimental protocol.

3.2 Learning EEG latent space

The first processing module of our approach consists of an encoder, which receives as input an EEG time series and provides as output a more compact and class-discriminative feature vector. In [26] we tested several encoder models and the most performing one is shown in Fig. 4. It consists of a standard LSTM layer followed by a nonlinear layer. An input EEG sequence is fed into the LSTM layer, whose output at the final time step goes into a fully-connected layer with a ReLU activation function. This simple architecture when stacked with a 40-way softmax layer yielded good performance — over 80% classification accuracy.

3.3 Image Generation using EEG latent space

Image generation from a brain signal feature vector encoding information about visual classes is the main contribution of this paper. Thus, we have developed and compared two different approaches: one based on a variational autoencoder and the other one based on a generative adversarial network.

The VAE version of the *Brain2Image* generator is a traditional variational autoencoder with the difference that the encoder part is not a convolutional neural network, but the LSTM-based

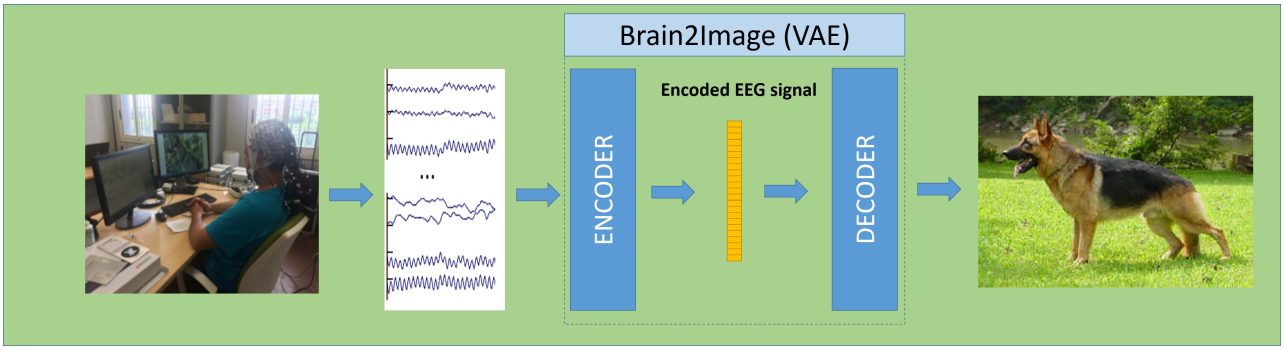


Figure 2: Overview of the VAE-based architecture design of the proposed *Brain2Image* module driving the EEG-based image generation approach, showing the constituting parts (encoder and decoder).

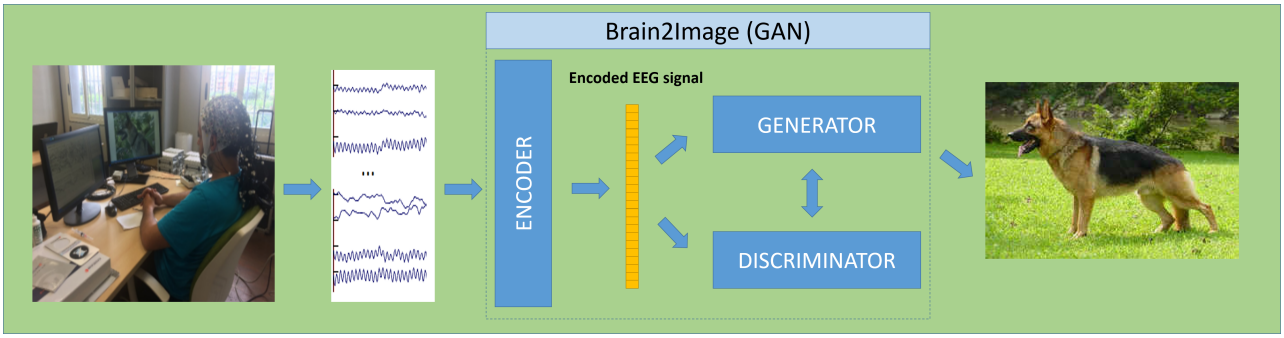


Figure 3: Overview of the GAN-based architecture design of the proposed *Brain2Image* module driving the EEG-based image generation approach, showing the constituting parts (encoder, generator and discriminator).

neural network described in the previous section. To that network, an additional fully-connected layer is added in order to impose the learned feature vector to have a Gaussian distribution as required by VAEs. The LSTM combined with this last fully-connected layer represents our encoder $E(z|x)$, which receives as input a raw EEG sequence x_I, with I indicating the related image, and outputs a projection to a lower-dimension latent multivariate Gaussian representation z (Fig. 4).

The decoder $D(\hat{I}|z)$ is a network consisting of a two-layer fully connected neural network with non-linear activation functions (ReLU) followed by a cascade of deconvolutional layers (more details are given in Sect. 4) that converts the latent distribution z produced by the encoder into an output image $\hat{I}$ (Fig. 5).

The objective of training is to minimize a loss function given by the sum between the Kullback–Leibler divergence — dissimilarity between the generated latent distribution and a Gaussian one — and the image generation loss which represents the accuracy of the image reconstruction (mean squared error between the generated image $\hat{I}$ and the target one I):

$$\mathcal{L}_{\text{VAE}} = D_{KL}[E(z|x_I)||D(\hat{I}|z)] + MSE(\hat{I}, I)$$

While as encoder we use a pre-trained LSTM for a brain signal classification task, the decoder network is trained from scratch: for each EEG sequence presented as the encoder's input, we use its output to train the network for generating the image that the subject is looking at that precise moment.

The second method we developed for image generation exploits generative adversarial networks. Unlike traditional GANs, our approach generates an image sample using random noise and a condition vector coming out from the encoder network. More specifically, our GAN consists of a generator and a discriminator. The generator $G(z|y)$ maps random input from a $p(z)$ noise distribution and an EEG-based conditioning vector y to a target image distribution $p_{\text{data}}(x)$. The discriminator $D(x|y)$ predicts, instead, the probability that a data point belongs to the target distribution of conditioning vectors. Both networks are trained simultaneously in a minimax settings: the discriminator attempts to distinguish correctly — maximizing the the probability of assigning correct labels — "real" data (from $p_{\text{data}}(x)$) from "fake" data (from $p_G(z|y)$), while the generator tries to make the discriminator fail by generating realistic images as if derived from data distribution. The minimax function $V(D, G)$ is the following:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \in p_{\text{data}}}(x)\left[\log D(x|y)\right] + \mathbb{E}_{z \in p_z(z)}\left[\log\left(1 - D\left(G\left(z|y\right)|y\right)\right)\right]$$

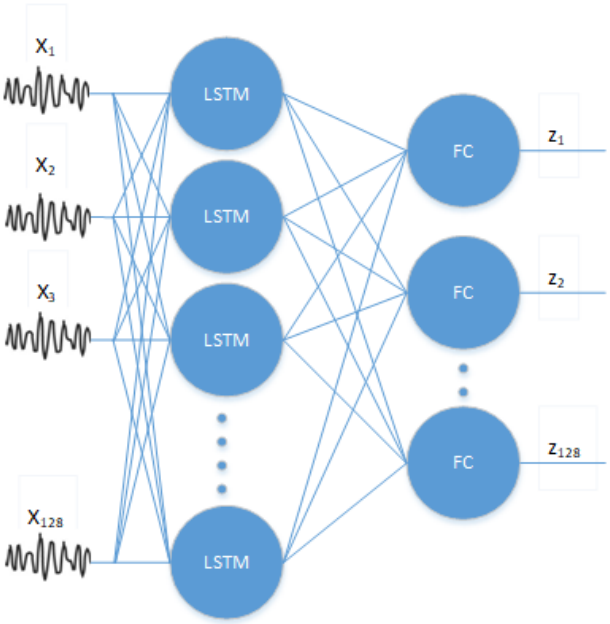


Figure 4: EEG feature encoder. The raw EEG signals are fed to an LSTM network, where the temporal dynamics of the input signals are modeled. Afterwards, the data is sent to a fully connected layer with non-linear activation functions (ReLU), whose output represents the EEG feature vector passed to either VAE or GAN.

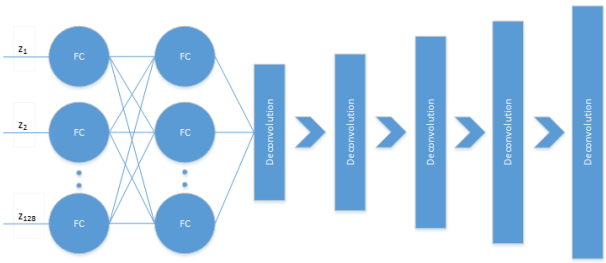


Figure 5: The image decoder of our VAE architecture.

In terms of loss function, a true sample $s_t = (x_t, y_t)$ consisting of real data with correct condition and a fake sample $s_f = (x_f, y_f)$ consisting of fake data with arbitrary condition, the negative log-likelihood discriminator loss is:

$$\mathcal{L}_D = -\log D(x_t|y_t) - \log\left(1 - D\left(x_f|y_f\right)\right) \quad (1)$$

while the generator loss, for an analogous s_f sample, is:

$$\mathcal{L}_G = -\log D\left(x_f|y_f\right) \quad (2)$$

Fig. 6 shows our GAN architecture with both D and G being convolutional networks.

As conditioning vector y we use the latent features learned by our encoder. It is used — concatenated with the random noise — both as input for the generator and appended to the feature maps of the second-to-last convolutional layer of the discriminator. Furthermore, given that we employ conditioning vectors for image generation, the generator needs to learn the correct association between the data distribution and the conditioning vectors. To simplify the learning dynamics, we change the discriminator loss function, following the approach in [22], by providing a real image with a wrong condition. Thus, we train our discriminator using true samples $s_t = (x_t, y_t)$ and wrong samples $s_{w_1} = (x_c, y_w)$ and $s_{w_2} = (x_w, y_w)$, and compute the loss as follows:

$$\begin{aligned}\mathcal{L}_D = &-\log D(x_t|y_t) \\ &-\log(1 - D(x_c|y_w)) \\ &-\log(1 - D(x_w|y_w))\end{aligned} \quad (3)$$

4 PERFORMANCE ANALYSIS

Performance analysis aimed at evaluating the accuracy of our *Brain2Image* approach in generating accurate and realistic images resembling those evoking the recorded brain signals.

While it is relatively easy to assess image quality in a qualitative manner, quantitative assessment of image fidelity and resemblance to real images is not trivial and not clearly defined. One recent metric to test generative methods is employing the Inception score [23], i.e., using the output of the Inception network [27] to assess the resemblance of a generated image to an object class. Even better, it also includes a metric for the quality of a generated image by evaluating how easy it is for the Inception network to classify it correctly.

For experimental evaluation, we split our EEG signal dataset into training, validation and test sets, with respective fractions 80% (1,600 images), 10% (200), 10% (200). Splitting by images, rather than by EEG signals (which, for each image, are as many as the number of participant subjects), makes sure that the signals generated by all subjects for a single image are not spread over different splits.

4.1 VAE and GAN's architectures

Adam gradient descent method [10] is used for training both approaches (learning rate initialized to 0.001), with mini-batches of size 16. The encoder's LSTM layer size is set to 128, as is the number of the fully connected non linear output layer.

The VAE's decoder and the GAN's generator share the same architecture, except from the input layer. In the VAE case, the fully connected input layers of the decoder has 128 nodes while in the GAN case, the generator's input layer has 228 (100-dimensional random noise and 128- dimensional EEG features) nodes. The models and training hyperparameters are tuned on the validation set.

In both models' generative parts (i.e., VAE's decoder and GAN's generator), the input passes through 5 deconvolutional layers: the first spatially upsamples the vector by four times, while each of the others doubles the size at every step, so that the output image size is 64×64. The number of features maps

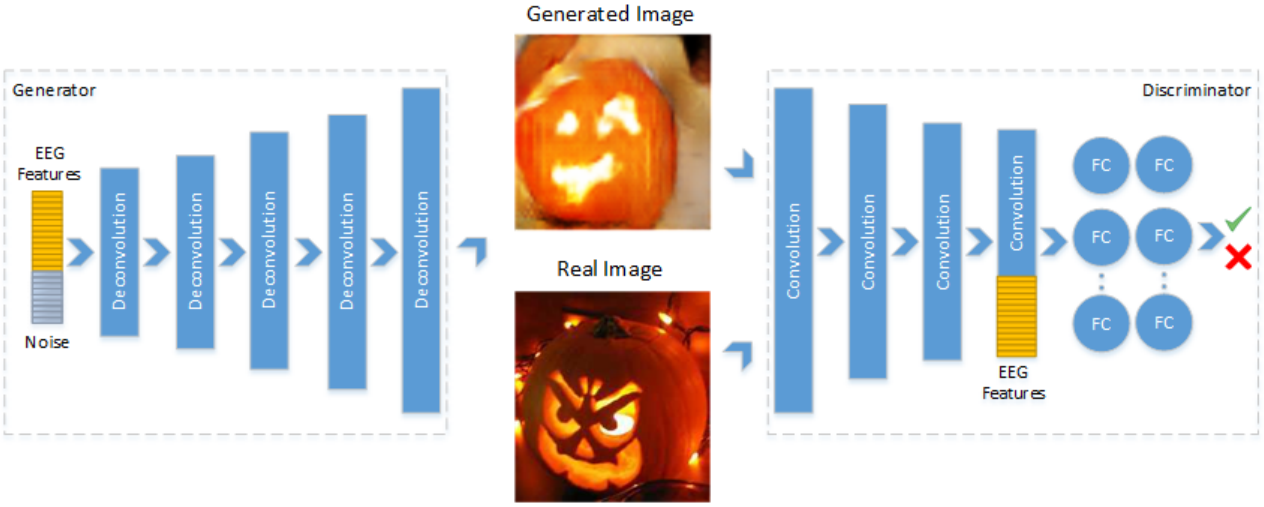


Figure 6: GAN Architecture. The generator receives a feature vector, which is created by concatenating the encoded EEG features (in yellow) — with random noise. The discriminator is trained from generated images and real images from the training dataset to identify whether an input image is real or fake. EEG features are used also by the discriminator in order to derive a decision about the correctness of the generated image with respect to the conditioning vector.

starts at 512 at the first layer, and is reduced by half for each successive deconvolutional layer before the last one, which generates a 3-channel (color) image.

In the GAN case, the discriminator is made up of four convolutional layers and two fully-connected layers. It takes as input 64×64 images, and analogously halves the feature map size at every convolutional step. After the final convolutional layer, where the feature map size is 4×4 (to which the condition vector is spatially appended), two fully-connected layers reduce the number of features to 1024 and 1, the latter being the sigmoidal probability estimate on the input image/condition pair. The number of feature maps in the convolutional layers starts at 64 at the first layer, and is doubled at every layer before the fully-connected ones, to a maximum of 512 feature maps at the last convolutional layer. While VAE implicitly use the latent features learned by the encoder, in the GAN case we explicitly append conditioning vectors to the input noise. All (de) convolutional layers in both models include batch normalization modules and ReLU activation functions. All model details are given in Tab. 2.

Model	**VAE**		**GAN**	
Component	**E**	**DEC**	**G**	**D**
Number of deconvolution layers	5	5	5	4
Initial number of feature maps	32	512	512	64
Kernel Size	4×4	4×4	4×4	4×4
Striding Size	2×2	2×2	2×2	2×2
Padding Size	1×1	1×1	1×1	1×1

Table 2: The employed models's details. In the second row, E stands for the VAE's encoder (of the first training phase, see Sec. 4.2), and DEC for the VAE's decoder, G for the GAN's generator and D for the GAN's discriminator.

4.2 VAE and GAN's training details

Deep-learning based generative models need a lot of data in order to avoid overfitting. However, acquiring EEG data to satisfy such data quantity needs is impractical, because of the nature of the experiment. In fact, we tried to minimize the time that subjects had to undergo the experiment because we noted that after a while, brain signals showed alterations due to subject tiredness and nervousness.

According to the data acquisition protocol discussed in Section 3.1, only 50 images per class were presented during the experiments, even though the ImageNet dataset contains about 1300 images per class. However, the unused images contained visual features that could be exploited in the image generation process, even if EEG data was not available. So, in order to use all possible information and to avoid overfitting, we trained the models in two stages:

(1) The models were initially pre-trained with images, for which EEG data was not available. For the VAE, a mirrored version of the decoder was used as an encoder, but with convolution layers instead of the deconvolution ones. For the GAN, all condition vectors y were set to the zero vector, and the loss term related to real images and wrong conditions was ignored.

(2) The models were then trained (fine-tuned) with those images for which EEG data was available. For the VAE, the EEG feature encoder was used for the encoder part, and only the additional fully-connected layer was trained during this stage (i.e. the parameters of the EEG feature

encoder in Fig. 4 were frozen). Given the complexity and instability of GAN training, we could not use one condition vector for each input EEG signals as this would have turned out in the impossibility to get realistic images, thus, as condition vector y associated to each image we used the average of the EEG feature encoder's output of all images in the same class and of all subjects.

Furthermore, data augmentation was performed during both stages by resizing images at 96×96 pixels, and extracting random 64×64 (horizontally flipped with 50% chance).

4.3 Image generation: qualitative and quantitative performance

Fig. 7 and 8 show samples generated, respectively, by VAE and GAN for some of the 40 ImageNet visual classes. Differences can be observed in the quality of the produced results. Indeed, as expected, the GAN network managed to generate sharp - although very artificial - images with respect to the VAE ones, which, instead, was able to capture better basic distinguishing patterns and to generate more realistic images. Thus, VAE was better in representing the object structure but with much less definition and sharpness. The results confirm that both models are able to translate EEG features into meaningful images of the appropriate image classes.

To quantitatively assess the performance of the two approaches in generating images from brain signals, we computed the Inception score both globally (across all classes) and on a per-class base. In the first case, we generated a sample of 50,000 images (1,250 per class); in the second case, we generated a sample of 50,000 images for each class, and computed per-class Inception scores. The results achieved by VAE and GAN are shown in Table 3. The results indicate that GAN outperformed VAE in the *Brain2Image* task. However, both methods yielded satisfactory results in converting raw EEG signals into meaningful images. The performance achieved by both methods are also inline with state-of-the-art methods, which, to our knowledge, have been tested only on simpler datasets than ImageNet such as CIFAR-10, on which the current best published result is 8.07 [23]. While the performance is not as high as those obtained on CIFAR-10, it should be noted that: 1) the resolution of the generated images is higher than CIFAR-10's one (64×64 vs 32×32); 2) we attempt to generate images from much more classes (40 vs 10 in CIFAR-10); 3) the images available for each class is lower than the ones in CIFAR-10 (1,200-1,300 in our case, whose only 50 images coming with EEG data vs 6,000 in CIFAR-10); 4) image generation process is driven by brain signals, which is true that they contain information about visual classes but such information is mixed with a lot of noise, while traditional generative methods use as conditioning vectors directly the labels of the image class.

Moreover, the Inception score does measure only the level of realism of the generated images, whereas, our goal is to translate brain signals evoked by visual stimuli into the corresponding images. To assess this capability, we employ the generated image samples (by the two approaches) and compute the class probability distribution of Inception, with the softmax layer changed for a 40-class classification task. The correct classification rate when using VAE generate images was 0.35, while for GAN-generated images we obtain 0.43. Table 3 shows per-class correct classification rate. As expected, similarly to the qualitative results, GAN outperformed VAE. This may be explained with the fact that the initial layers of GAN match better fine details and simple visual patterns, which VAE is unable to reproduce adequately.

5 CONCLUDING REMARKS

In this paper, we propose an approach able to reconstruct images seen by human subjects using only brain signals. In particular, we developed a general deep-learning framework that given input time series encoding information about visual classes is able to reverse them into images. We tested two approaches: one using variational autoencoders and the other one employing generative adversarial networks. The results show that both approaches are able to generate images semantically coherent with the visual stimuli evoking specific brain signals. GAN, in general, outperform VAE, which, in turn, generate more realistic images that however lack fine details. Thus, as future work, we aim at combining them as in [13]. We are currently working also on acquiring fMRI data to complement EEG data and provide the generative methods less noise and more significant conditioning vectors. However, the results achieved in this paper are highly satisfactory and represent an important step forward the reading the mind goal.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of NVIDIA Corporation for the donation of two Titan X Pascal GPUs used for this research. We also acknowledge Dr. Martina Platania for carrying out the EEG data acquisition.

REFERENCES

[1] U Rajendra Acharya, S Vinitha Sree, G Swapna, Roshan Joy Martis, and Jasjit S Suri. 2013. Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems* 45 (2013), 147–165.

[2] T. Carlson, D. A. Tovar, A. Alink, and N. Kriegeskorte. 2013. Representational dynamics of object vision: the first 1000 ms. *Journal of Vision* 13, 10 (2013).

[3] T. A. Carlson, H. Hogendoorn, R. Kanai, J. Mesik, and J. Turret. 2011. High temporal resolution decoding of object position and category. *Journal of Vision* 11, 10 (2011).

[4] Hubert Cecotti and Axel Graser. 2011. Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence* 33, 3 (2011), 433–445.

[5] K. Das, B. Giesbrecht, and M. P. Eckstein. 2010. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *Neuroimage* 51, 4 (Jul 2010), 1425–1437.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

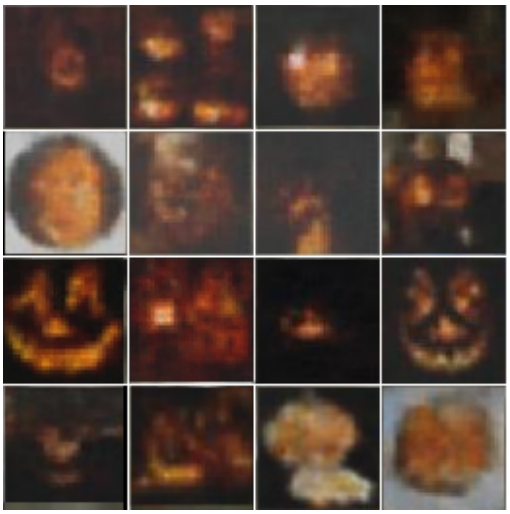
[7] Andrea M Green and John F Kalaska. 2011. Learning to move machines with the mind. *Trends in neurosciences* 34, 2 (2011), 61–75.

[8] Christoph Guger, Werner Harkam, Carin Hertnaes, and Gert Pfurtscheller. 1999. Prosthetic control by an EEG-based brain-computer interface (BCI). In *Proc. aaate 5th european conference for the advancement of assistive technology*. 3–6.

[9] John-Dylan Haynes and Geraint Rees. 2006. Decoding mental states from brain activity in humans. *Nature reviews. Neuroscience* 7, 7 (2006), 523.

[10] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

(a) Airliner



(b) Jack-o'-Lantern

(c) Panda

Figure 7: VAE results in terms of generated images

(a) Airliner

(b) Jack-o'-Lantern

(c) Panda

Figure 8: GAN results in terms of generated images

Generative method	VAE		GAN	
Class	**IS**	**IC**	**IS**	**IC**
German shepherd (n02106662)	3.04	0.11	4.91	0.23
Egyptian cat (n02124075)	2.98	0.21	4.45	0.29
Lycaenid butterfly (n02281787)	4.44	0.34	5.03	0.37
Sorrel (n02389026)	2.91	0.37	5.86	0.62
Capuchin (n02492035)	4.89	0.39	4.99	0.41
Elephant (n02504458)	5.01	0.49	5.35	0.57
Panda (n02510455)	4.97	0.58	6.35	0.72
Anemone fish (n02607072)	5.11	0.46	6.11	0.81
Airliner (n02690373)	6.14	0.34	6.20	0.86
Broom (n02906734)	4.07	0.41	4.76	0.35
Canoe (n02951358)	3.85	0.22	4.59	0.24
Cellphone (n02992529)	4.99	0.11	5.17	0.31
Mug (n03063599)	2.17	0.14	4.62	0.23
Convertible (n03100240)	4.09	0.27	4.54	0.34
Desktop PC (n03180011)	5.77	0.59	5.81	0.61
Digital watch (n03197337)	4.26	0.44	4.54	0.51
Electric guitar (n03272010)	4.34	0.17	4.91	0.32
Electric locomotive (n03272562)	3.76	0.18	4.88	0.24
Espresso maker (n03297495)	5.31	0.31	5.33	0.32
Folding chair (n03376595)	4.72	0.19	4.88	0.27
Golf ball (n03445777)	3.13	0.18	5.06	0.28
Piano (n03452741)	4.66	0.21	4.47	0.22
Iron (n03584829)	3.74	0.23	4.32	0.23
Jack-o'-lantern (n03590841)	6.89	0.96	6.64	0.91
Mailbag (n03709823)	4.17	0.37	5.51	0.49
Missile (n03773504)	5.70	0.45	5.87	0.54
Mitten (n03775071)	3.59	0.31	5.10	0.36
Mountain bike (n03792782)	4.51	0.14	4.86	0.33
Mountain tent (n03792972)	4.18	0.29	4.70	0.30
Pyjama (n03877472)	3.71	0.21	4.21	0.20
Parachute (n03888257)	4.39	0.23	4.59	0.38
Pool table (n03982430)	4.28	0.29	4.68	0.35
Radio telescope (n04044716)	4.99	0.33	5.08	0.37
Reflex camera (n04069434)	3.09	0.22	4.64	0.29
Revolver (n04086273)	4.42	0.26	4.55	0.26
Running shoe (n04120489)	4.01	0.23	4.31	0.22
Banana (n07753592)	7.31	0.91	6.28	0.83
Pizza (n07873807)	6.01	0.87	5.87	0.79
Daisy (n11939491)	5.67	0.66	5.81	0.74
Bolete (n13054560)	4.37	0.42	5.37	0.60
All	*4.49*	*0.35*	*5.07*	*0.43*

Table 3: Inception scores (IS) and Inception classification accuracies (IC) for each class of the dataset (specified by their ImageNet synset identifier and by a short description), and overall.

[11] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[12] Z. Kourtzi and N. Kanwisher. 2000. Cortical regions involved in perceiving object shape. *J. Neurosci.* 20, 9 (May 2000), 3310–3318.

[13] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. 2015. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300* (2015).

[14] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. 2010. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering* 57, 7 (2010), 1798–1806.

[15] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[16] Gernot R Muller-Putz and Gert Pfurtscheller. 2008. Control of an electrical prosthesis with an SSVEP-based BCI. *IEEE Transactions on Biomedical Engineering* 55, 1 (2008), 361–364.

[17] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, Yasue Mitsukura, and Tzyy-Ping Jung. 2014. A high-speed brain speller using steady-state visual evoked potentials. *International journal of neural systems* 24, 06 (2014), 1450019.

[18] Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 6 (2009), 902–915.

[19] Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology* 21, 19 (2011), 1641–1646.

[20] H. P. Op de Beeck, K. Torfs, and J. Wagemans. 2008. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J. Neurosci.* 28, 40 (Oct 2008), 10111–10123.

[21] M. V. Peelen and P. E. Downing. 2007. The neural basis of visual body perception. *Nat. Rev. Neurosci.* 8, 8 (Aug 2007), 636–648.

[22] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, Vol. 3.

[23] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*. 2226–2234.

[24] Andrew B Schwartz, X Tracy Cui, Douglas J Weber, and Daniel W Moran. 2006. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron* 52, 1 (2006), 205–220.

[25] Pradeep Shenoy and Desney Tan. 2008. Human-Aided Computing: Utilizing Implicit Human Processing to Classify Images. In *CHI 2008 Conference on Human Factors in Computing Systems*. http://research.microsoft.com/apps/pubs/default.aspx?id=64271

[26] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. 2017. Deep Learning Human Mind for Automated Visual Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017).

[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.

[28] C. Wang, S. Xiong, X. Hu, L. Yao, and J. Zhang. 2012. Combining features from ERP components in single-trial EEG for discriminating four-category visual objects. *J Neural Eng* 9, 5 (Oct 2012), 056013.