INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

IIIT Hyderabad

CSAI
Project Presentation

Image Reconstruction from fMRI Data

Akshit Sharma (2021101029)

Rudransh Agarwal (2021101033)

Objective

Our objective is to explore different techniques for image reconstruction from fMRI data (which is obtained from a subject who is shown an image and his fMRI data is collected) based on recent developments in the field of Deep Learning

Research Question

The question is whether natural images can be reconstructed merely from fMRI data and which ROIs are responsible for encoding meaningful information in the brain while viewing such images

Datasets Used

- **Interim Submission:** For the interim submission we have used the **Binary Contrast Pattern dataset from Miyawaki et.al.** which consists of visual stimuli that are just binary images.

- **Final Submission:** We have used the **Deep Image Reconstruction datatset** which contains **6000 samples** of fMRI scans, captured with **3 subjects** on natural visual stimuli. We will be using this dataset in order to achieve reconstruction on natural images.

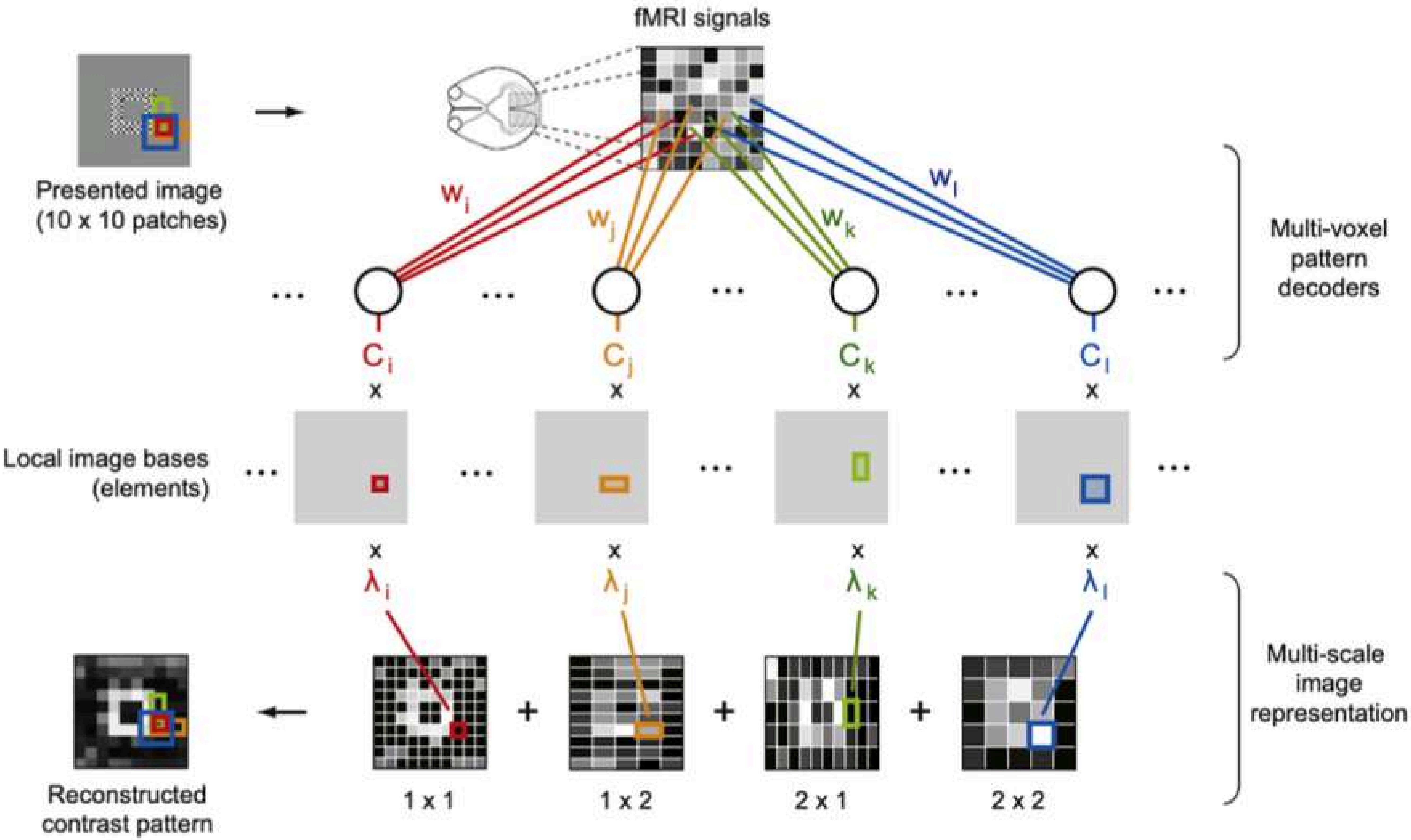
Interim Submission

Based on **Neuron journal’s** publication titled: **Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders (2008)**

- Reconstructed visual images by combining local image bases of multiple scales, whose contrasts were independently decoded from fMRI activity by automatically selecting relevant voxels and ex- ploiting their correlated patterns. Binary-contrast, 10x10-patch images (2^100 possible states) were accurately reconstructed.

- An approach to visual image reconstruction using multivoxel patterns of fMRI signals and multiscale visual representation. Assumed that an image is represented by a linear combination of local image elements of multiple scales. The stimulus state at each local element (Ci,Cj,...) is predicted by a decoder using multivoxel patterns (weight set for each decoder, wi,wj,...), and then the outputs of all the local decoders are combined in a statistically optimal way (combination coefficient, λi, λj , ...) to reconstruct the presented image. Only the voxels in the early visual cortex, that is the V1 and V2 regions are taken into account.

Interim Submission



Extending this Method

- Nilearn implementation uses Orthogonal Matching Pursuit(OMP) for calculating the weights, we have tried different models with the likes of linear regression and bayesian ridge regression. OMP performs the best becasuse of the feature selection associated with it (takes only the features with weights which are not zero). While the Nilearn code assigns fixed weights to all the decoders(0.25), we have trained the weights for each decoder using SGD. This is done because while performing MVPA, the contribution of a single voxel pattern might be different from a shared voxel pattern.

- Furthermore, the paper only suggests a multi-scale pattern of (1*1, 1*2, 2*1 and 2*2) whereas we have tried various other patterns as well like (1*1, 1*3, 3*1 and 3*3) and (1*1,1*5, 5*1, 5*5). We have done this because consideration of voxels which are farther apart might also effect the reconstruction. As a baseline, we have also experimented with a univariate analysis.

Interim Submission Results

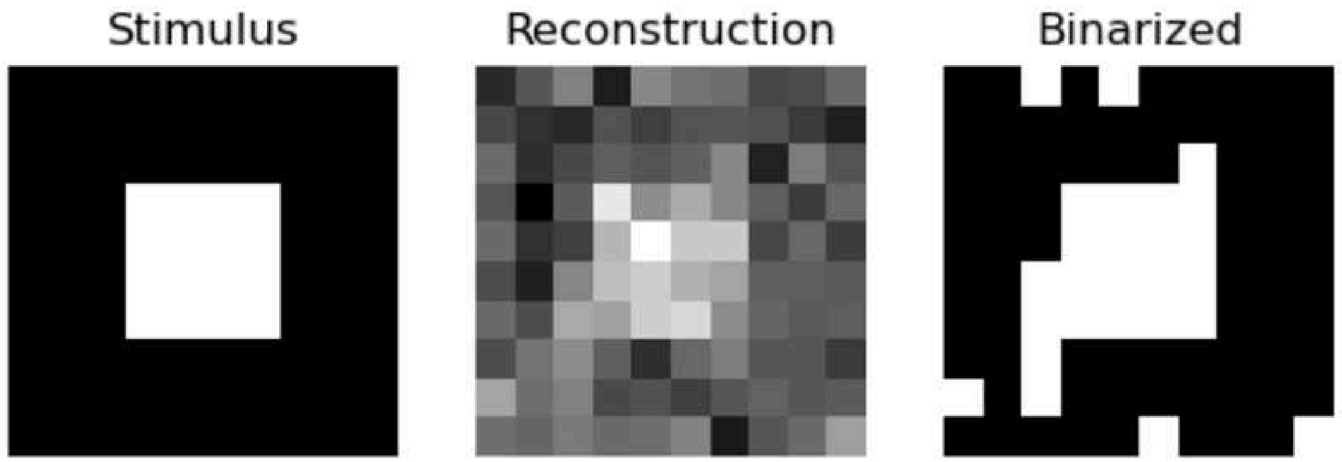


Figure 1: (1*1), Accuracy: 75.1%

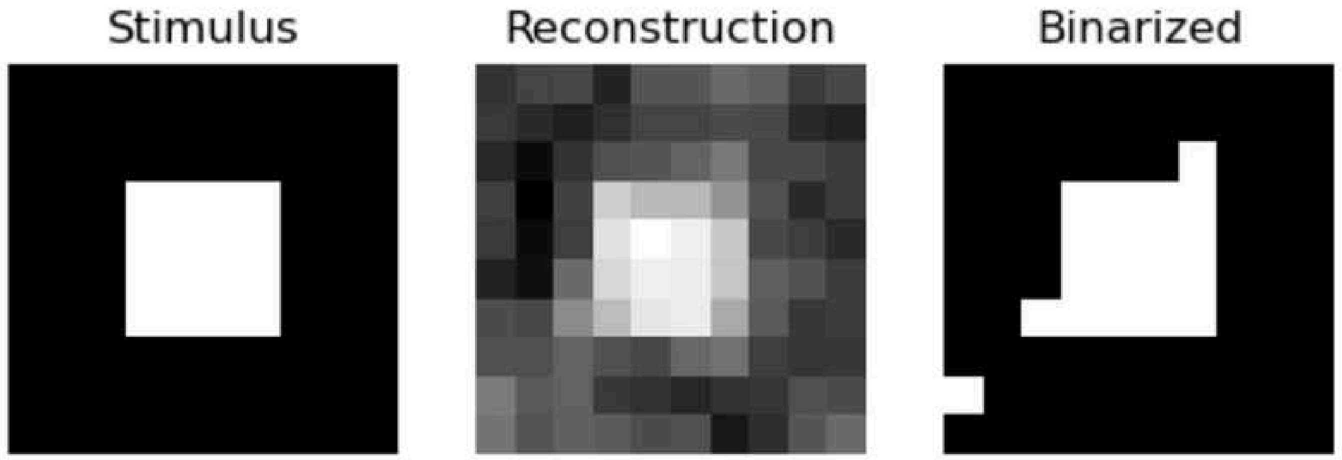


Figure 2: (1*1, 1*2, 2*1 and 2*2), Accuracy: 80.2%

Scale(x)	1*1	1*x	x*1	x*x
2	0.18	0.34	0.27	0.19
3	0.21	0.33	0.26	0.18
5	0.25	0.32	0.25	0.16

Table 1: trained λ values (weights) for each scale in the 3 cases for multi-scale reconstruction

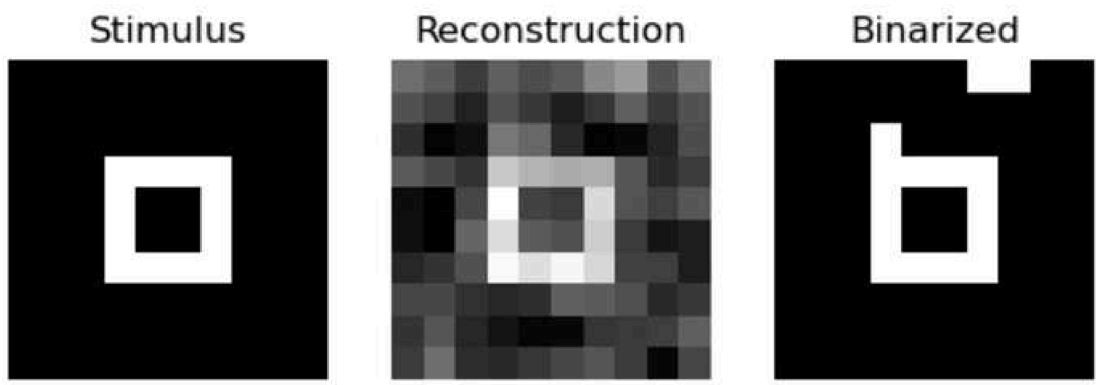


Figure 3: OMP, Accuracy: 80.2%

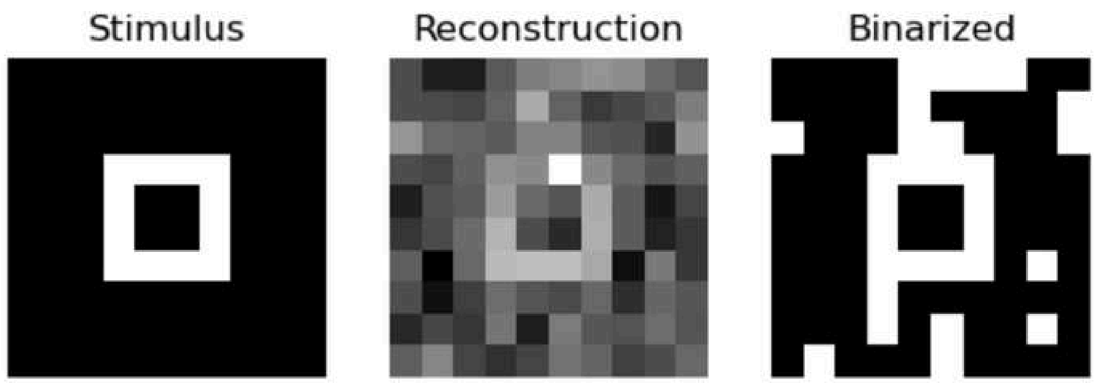


Figure 4: LinearRegressor, Accuracy: 72.9%

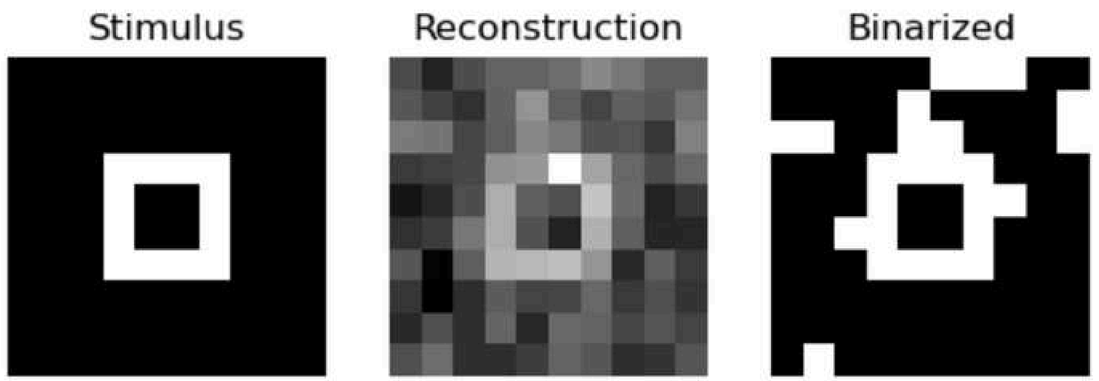


Figure 5: BayesianRidge, Accuracy: 76.1%

Interim Submission Results

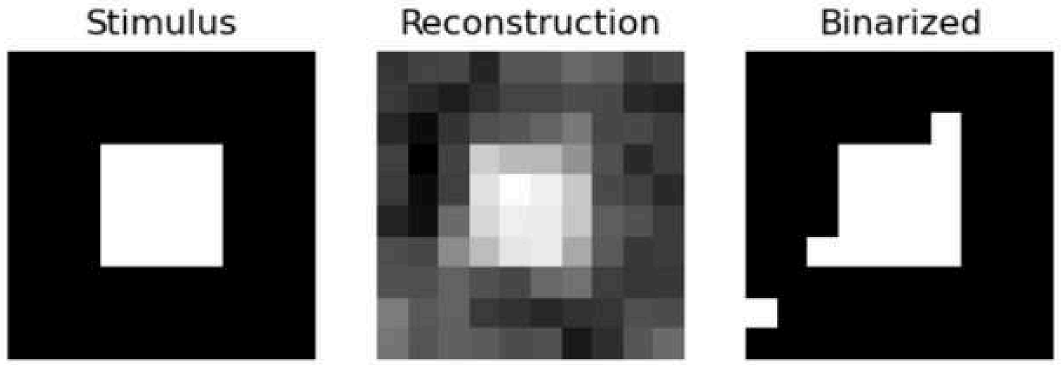


Figure 6: (1*1, 1*2, 2*1 and 2*2), Accuracy: 80.2%

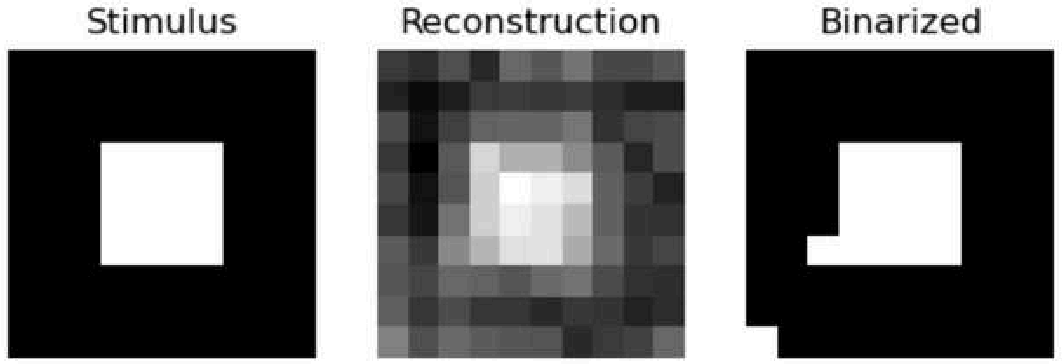


Figure 7: (1*1, 1*3, 3*1 and 3*3), Accuracy: 83.2%

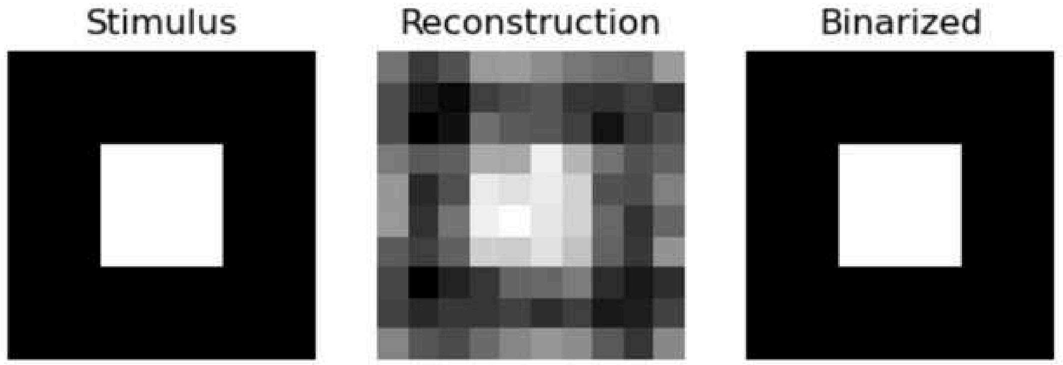


Figure 8: (1*1, 1*5, 5*1 and 5*5), Accuracy: 82.1%

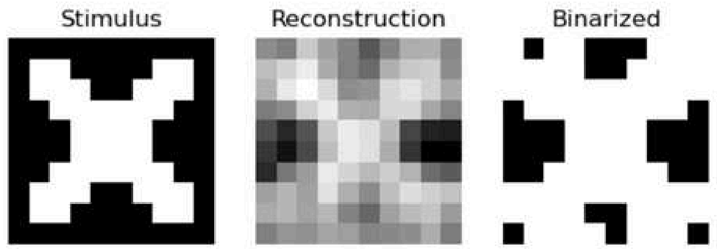


Figure 9: (1*1, 1*2, 2*1 and 2*2), Accuracy: 80.2%

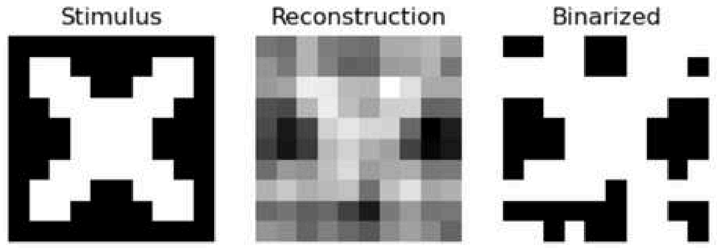


Figure 10: (1*1, 1*3, 3*1 and 3*3), Accuracy: 83.2%

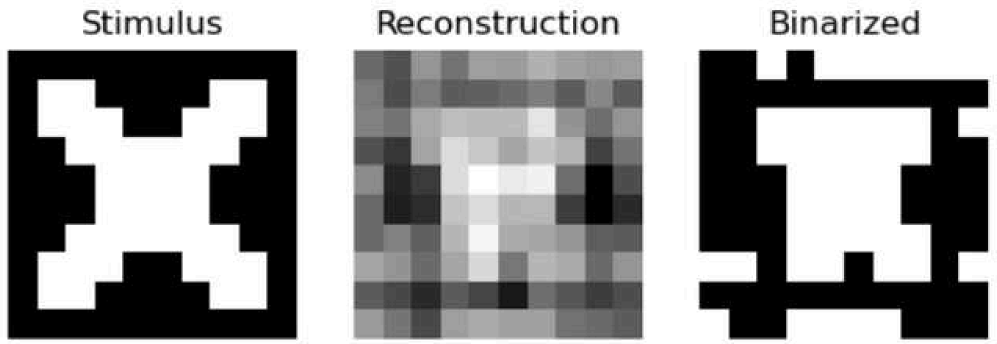


Figure 11: (1*1, 1*5, 5*1 and 5*5), Accuracy: 82.1%

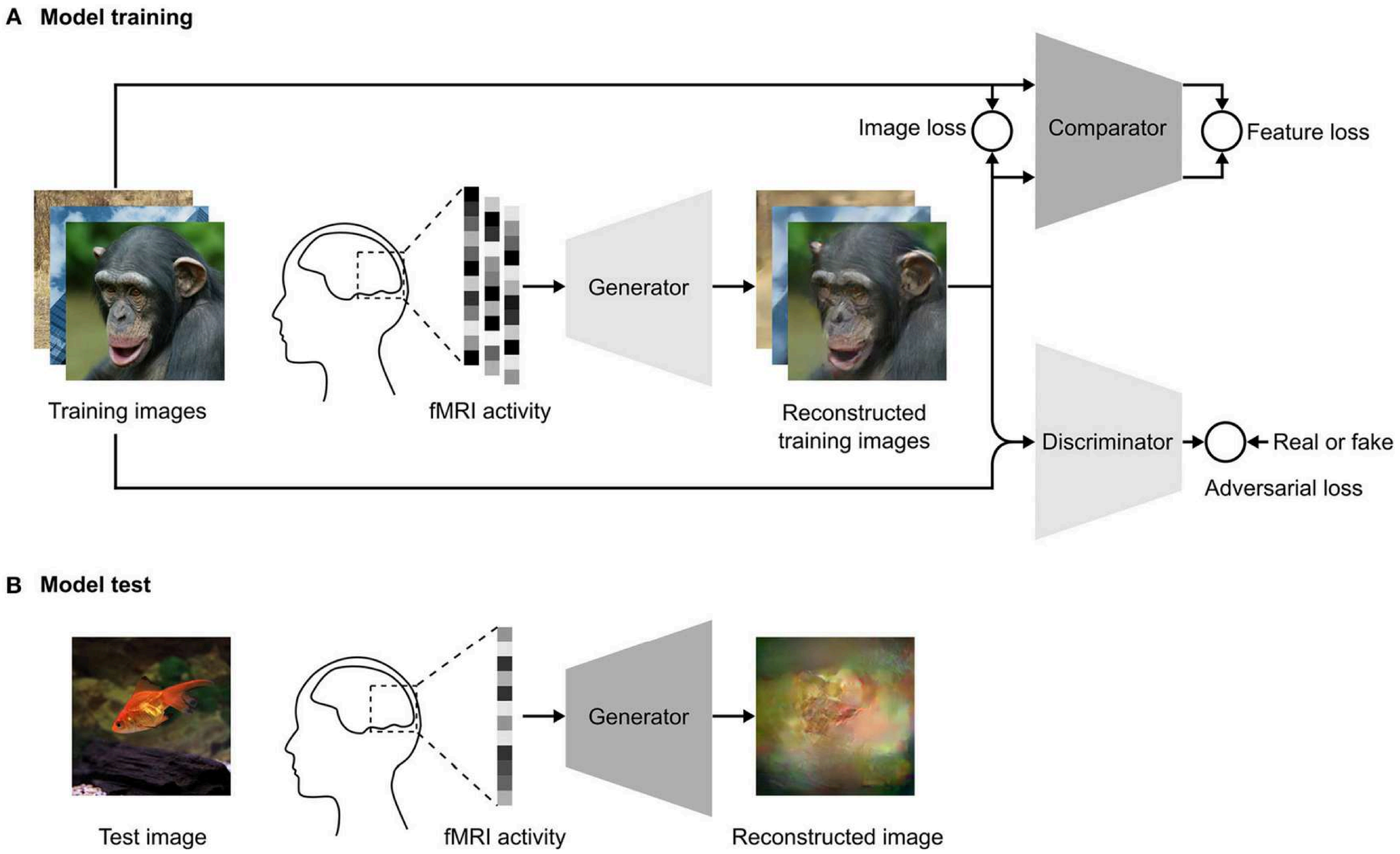
Final Submission

Based on **Frontiers in Computer Neuroscience journal**’s publication titled: **End-to-End Deep Image Reconstruction from Human Brain Activity (2019)**

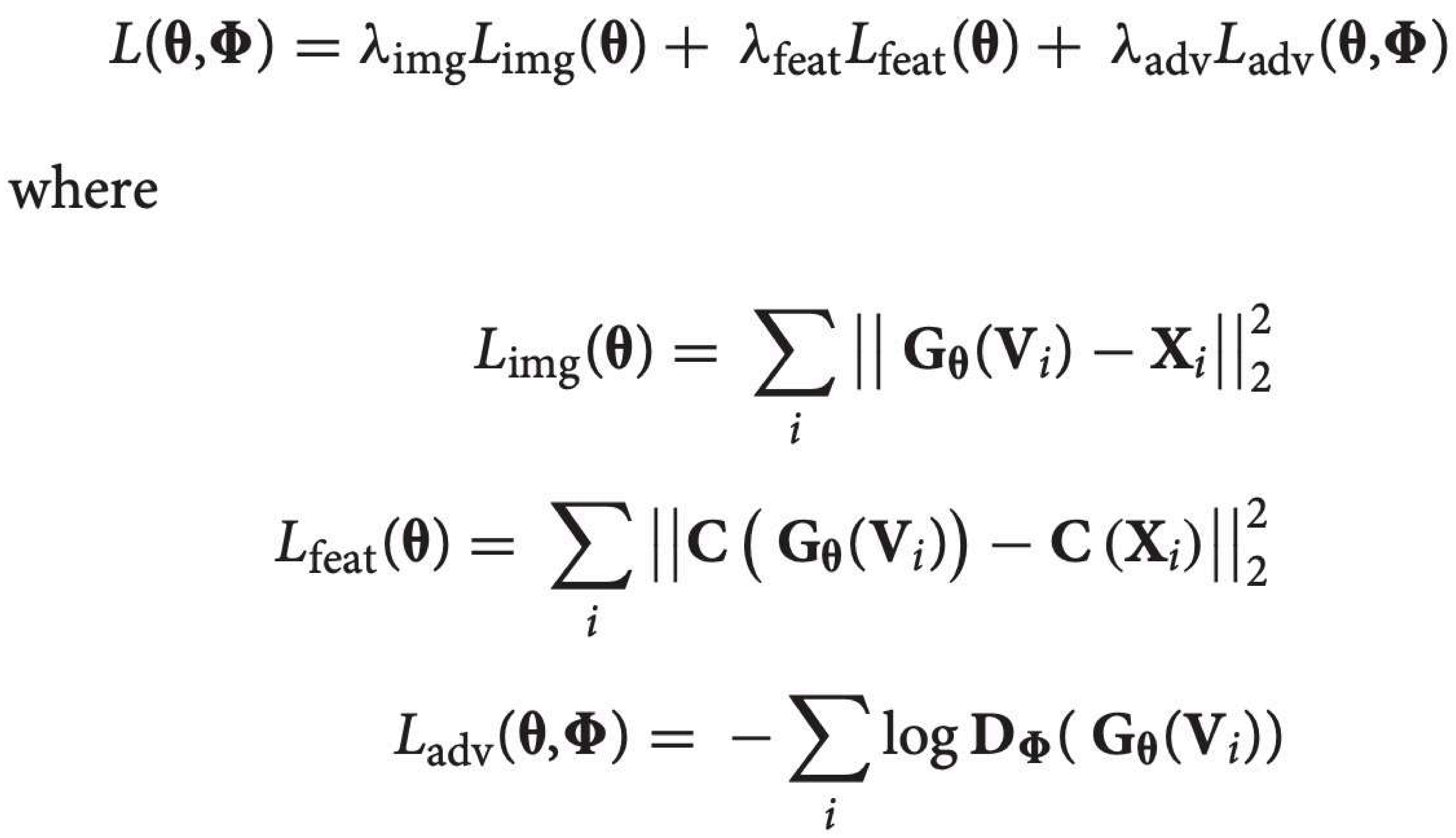
- We trained a DNN model with fMRI data and the corresponding stimulus images to build an end-to-end reconstruction model. We accomplished this by training a generative adversarial network (GAN) with an additional loss term that was defined in high-level feature space.

- The idea here is to train an end-to-end reconstruction network which takes fmri as input and outputs the reconstructed image. For this, we train a GAN architecture. Basically, each fmri is fed into the GAN, and it generates an image, the disriminator then tries to predict if this image is real or fake.

Training and Testing the Model



The 3 types of Loss Terms and the Generator Loss



$$L(\boldsymbol{\theta},\boldsymbol{\Phi}) = \lambda_{\text{img}} L_{\text{img}}(\boldsymbol{\theta}) + \ \lambda_{\text{feat}} L_{\text{feat}}(\boldsymbol{\theta}) + \ \lambda_{\text{adv}} L_{\text{adv}}(\boldsymbol{\theta},\boldsymbol{\Phi})$$

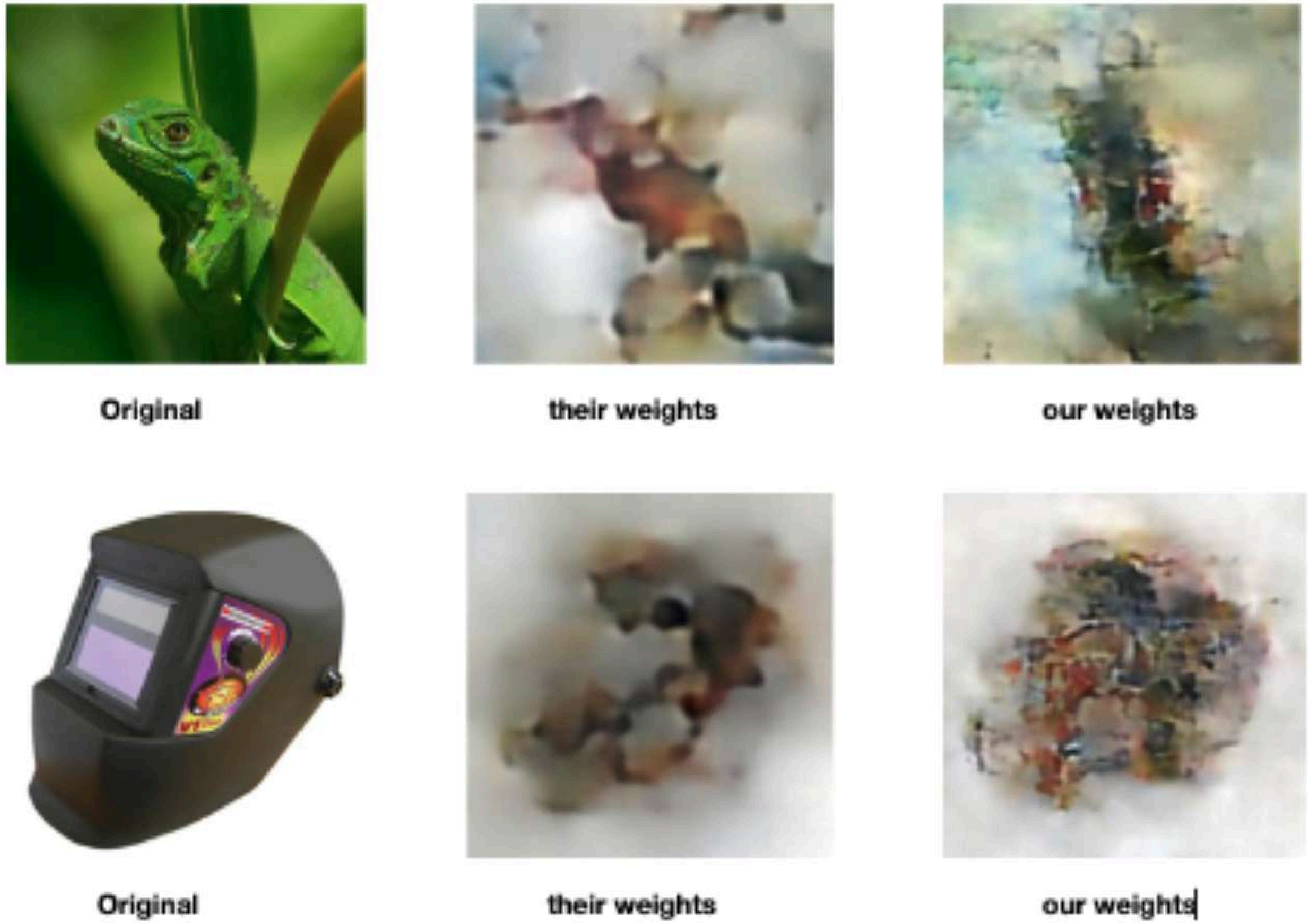
where

$$L_{\text{img}}(\boldsymbol{\theta}) = \ \sum_i \left|\left| \, \mathbf{G}_{\boldsymbol{\theta}}(\mathbf{V}_i) - \mathbf{X}_i \right|\right|_2^2$$

$$L_{\text{feat}}(\boldsymbol{\theta}) = \ \sum_i \left|\left| \mathbf{C} \left(\, \mathbf{G}_{\boldsymbol{\theta}}(\mathbf{V}_i) \right) - \mathbf{C} \left(\mathbf{X}_i \right) \right|\right|_2^2$$

$$L_{\text{adv}}(\boldsymbol{\theta},\boldsymbol{\Phi}) = \ -\sum_i \log \mathbf{D}_{\boldsymbol{\Phi}}(\, \mathbf{G}_{\boldsymbol{\theta}}(\mathbf{V}_i))$$

Reconstruction Results (On test data)



Original

their weights

our weights

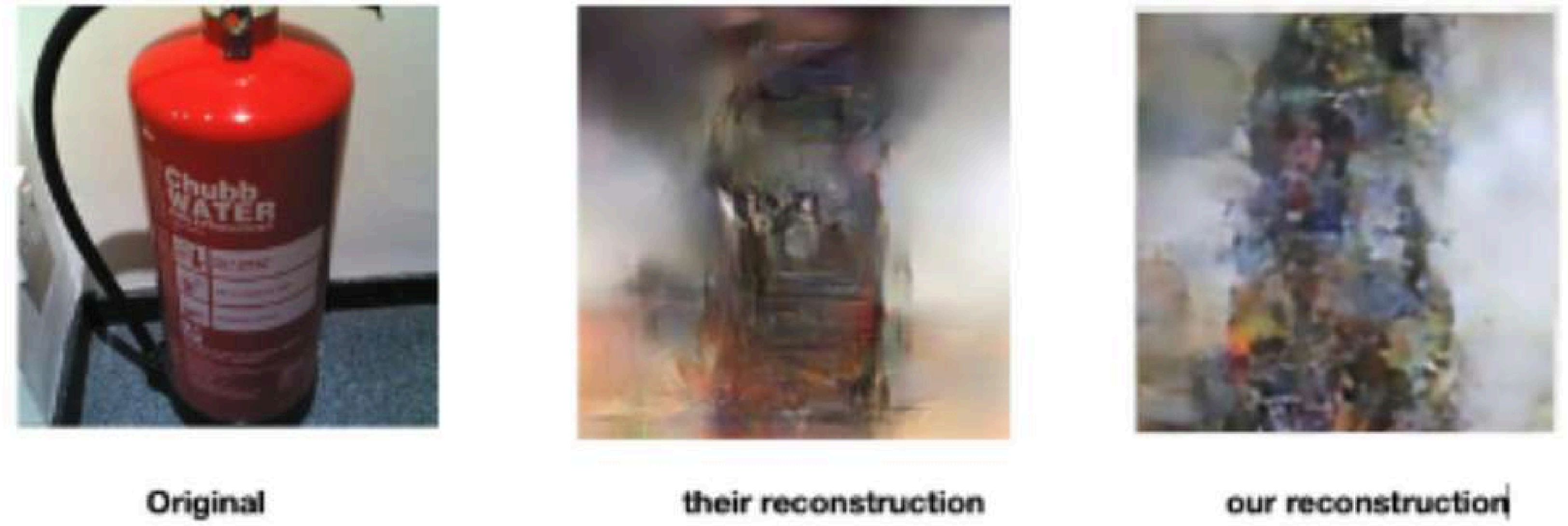
Original

their weights

our weights

Here, our weights refers to the best model weights, that we have used, giving better results compared to the weights meintioned in paper

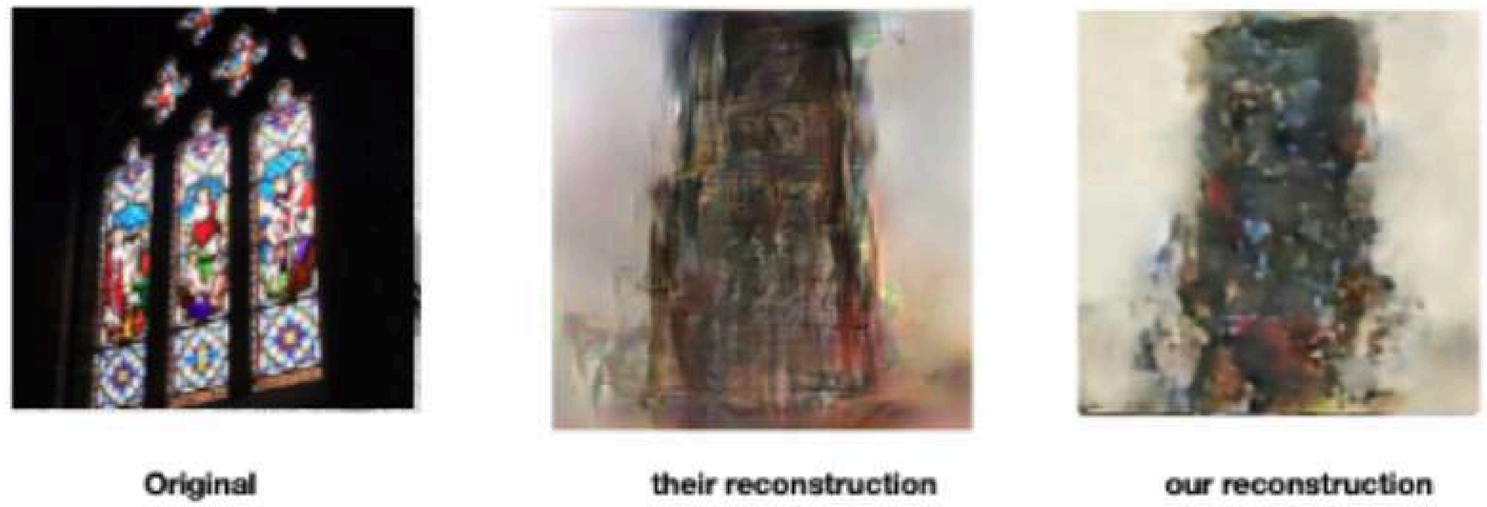
Reconstruction Results (On test Data)


Original

their reconstruction

our reconstruction

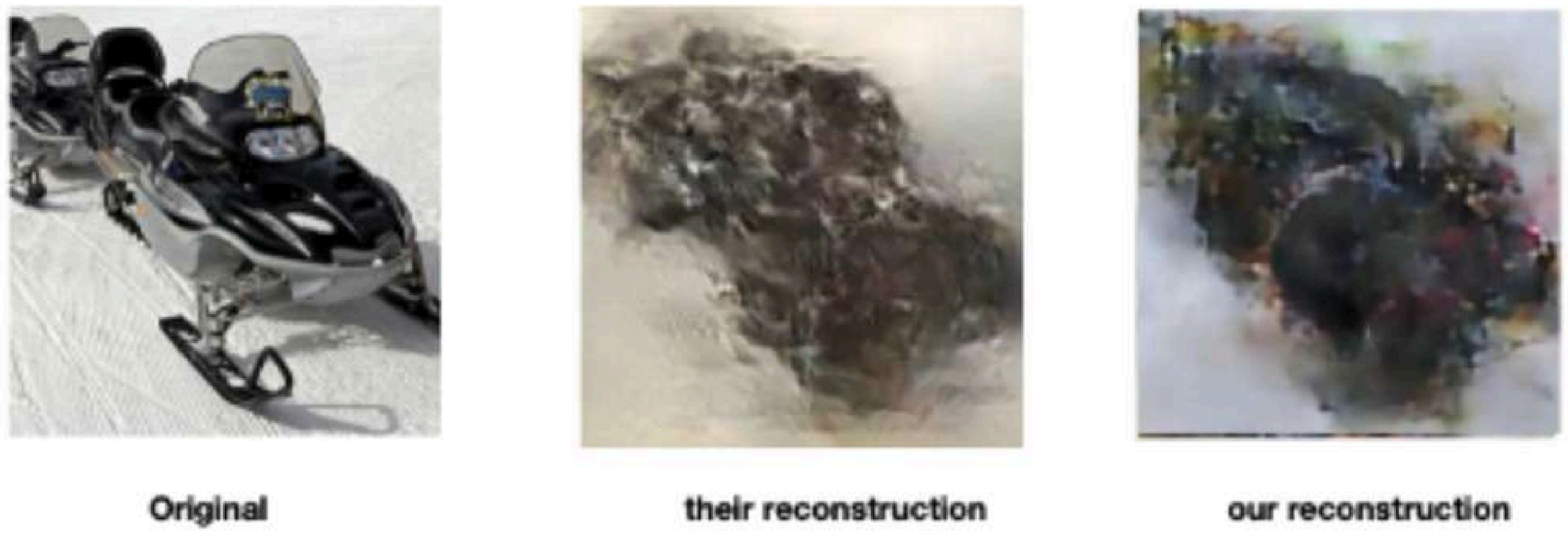
Fig. 17.


Original

their reconstruction

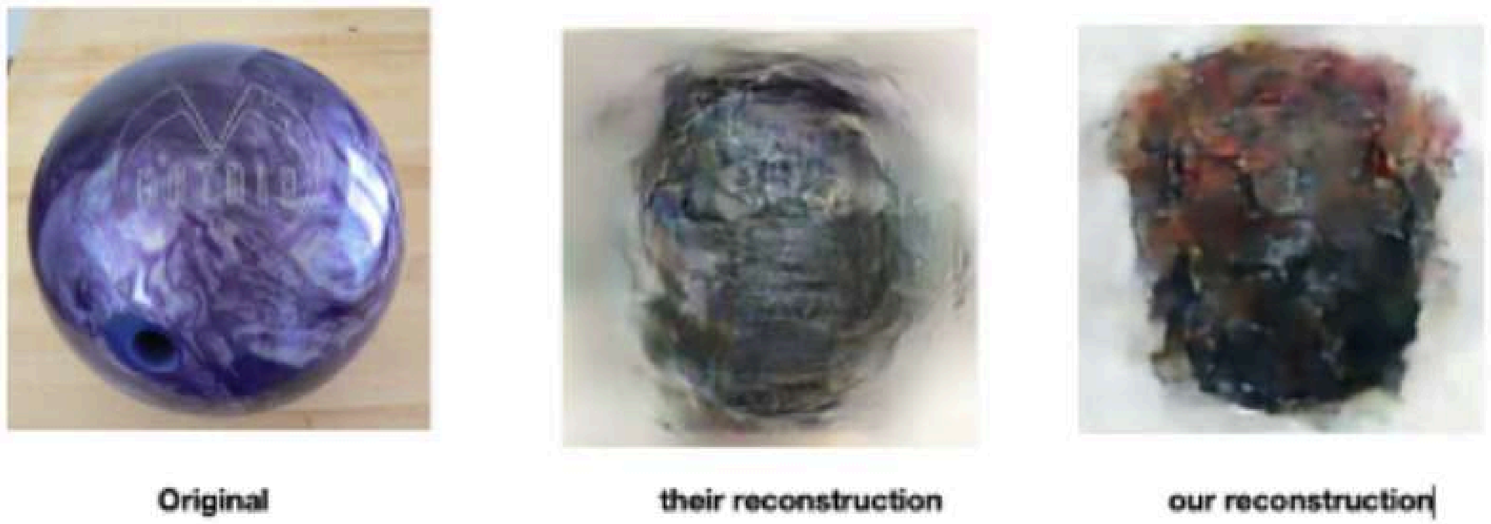
our reconstruction

Fig. 19.


Original

their reconstruction

our reconstruction

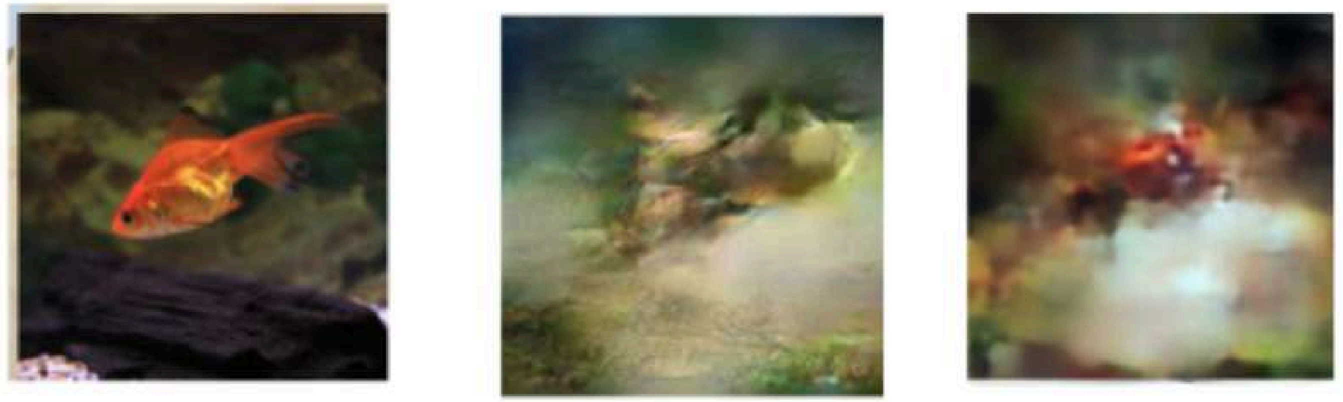

Original

their reconstruction

our reconstruction

Here, our weights refers to the best model weights, that we have used, giving better results compared to the weights meintioned in paper

Reconstruction Results for Different ROIs (On test Data)

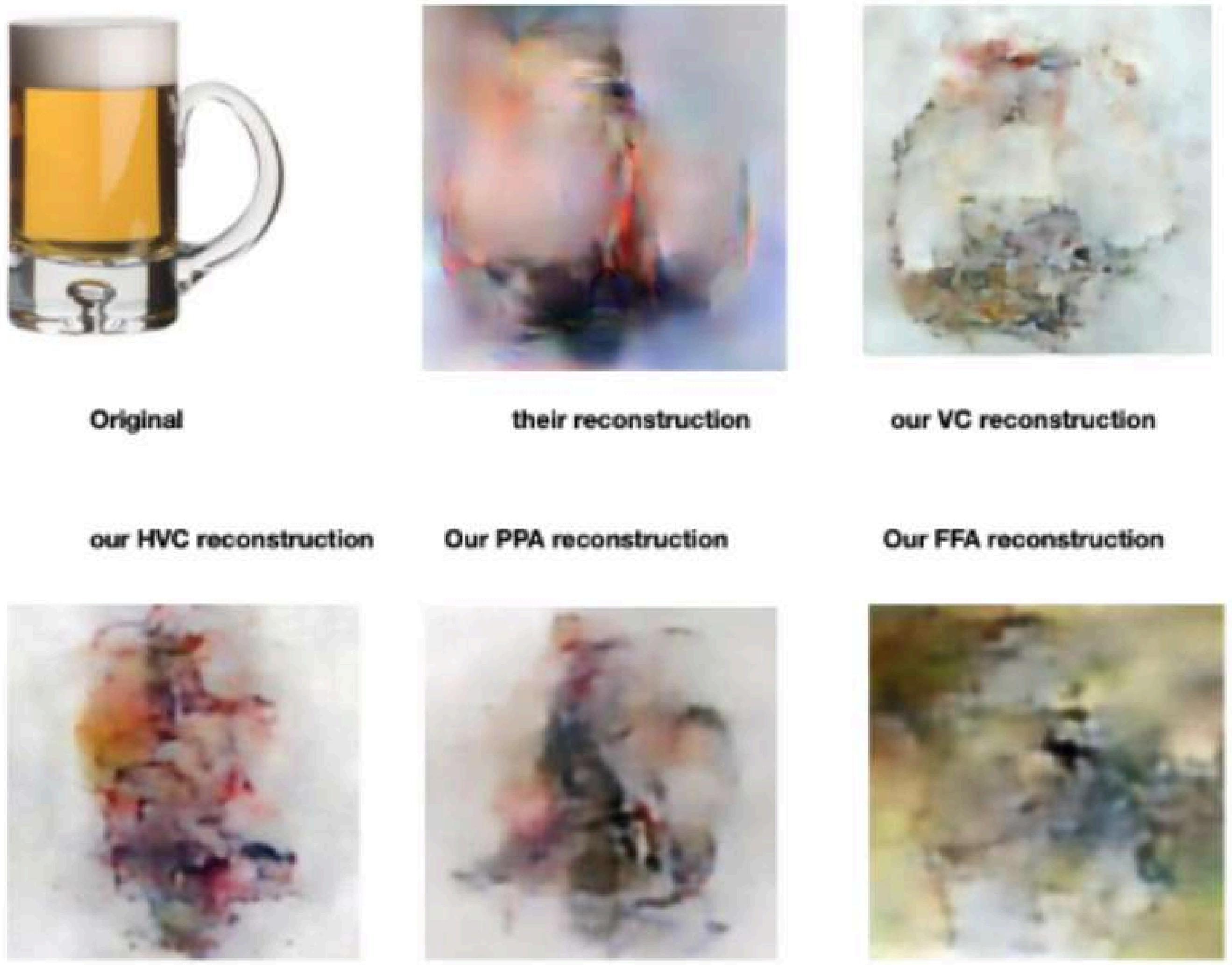


Original

their reconstruction

our V1-V2 reconstruction

Fig. 21.



Original

their reconstruction

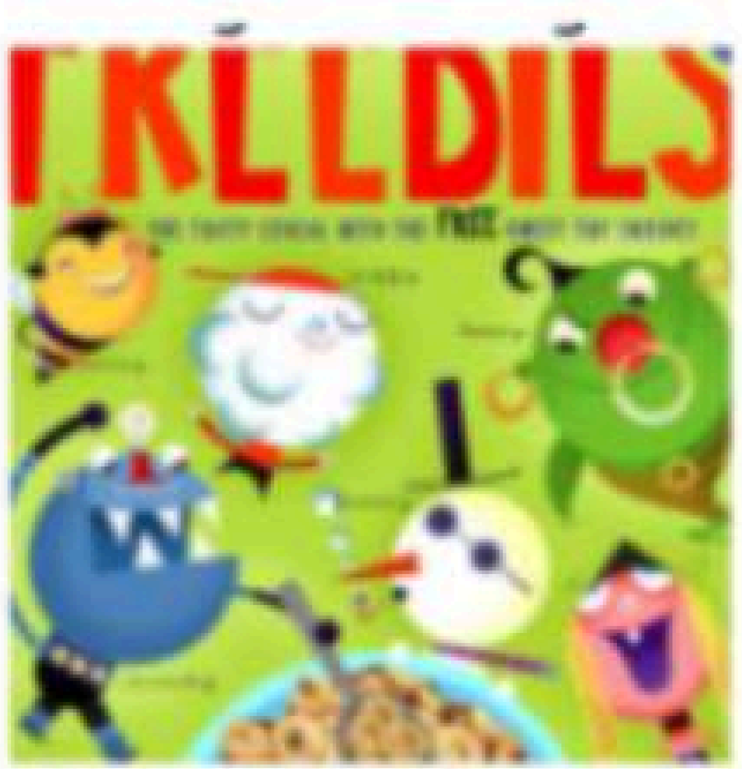
our VC reconstruction

our HVC reconstruction

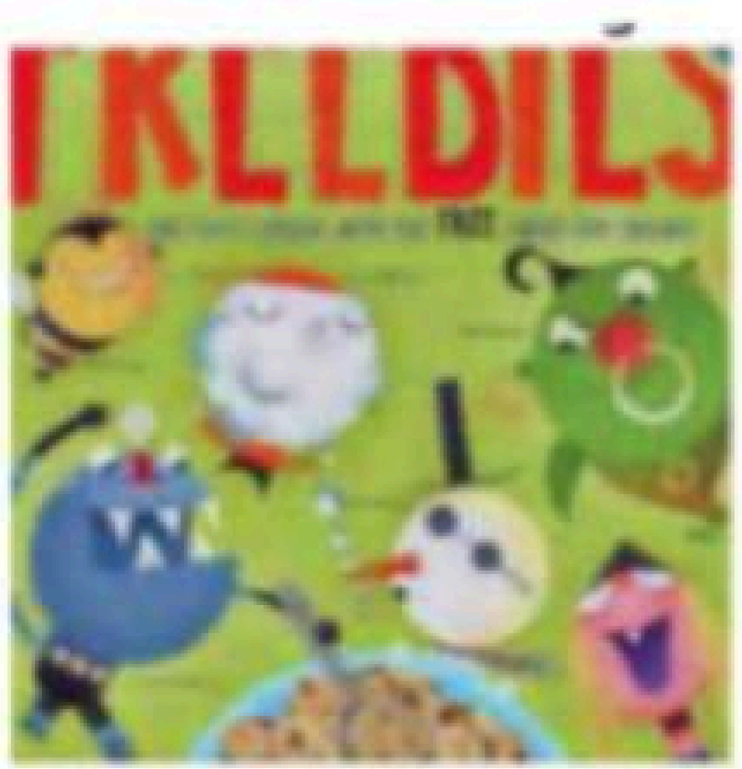
Our PPA reconstruction

Our FFA reconstruction

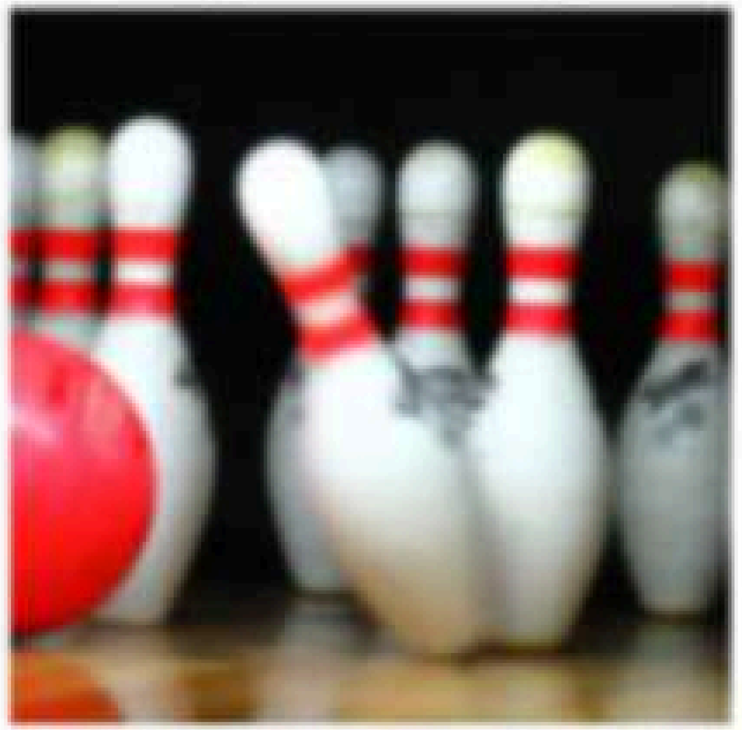
Reconstruction Results (On train Data)



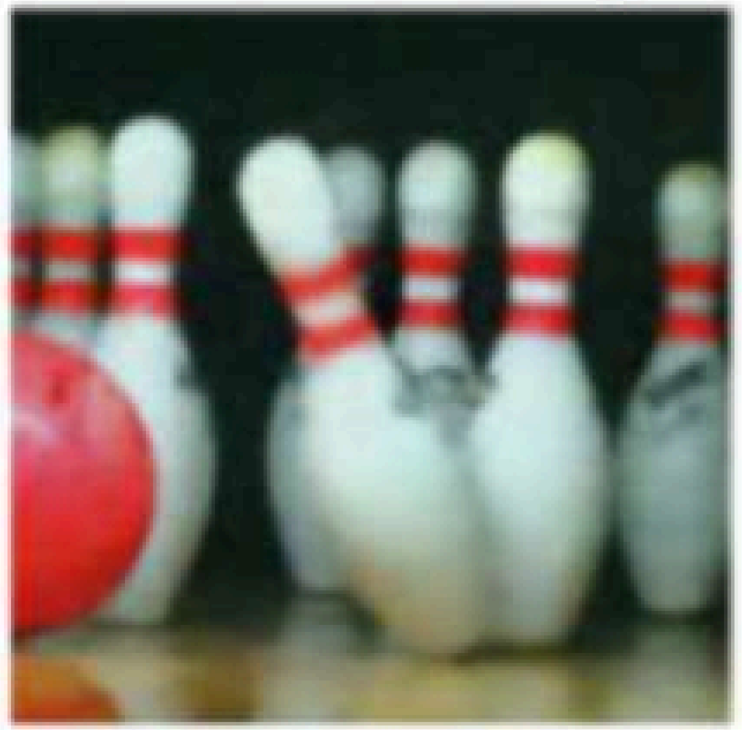
Original Image



Generated Image



Original Image



Generated Image

Observations

- Reconstructions done with the lower visual cortex regions, V1 and V2 are almost comparable to the whole visual cortex.

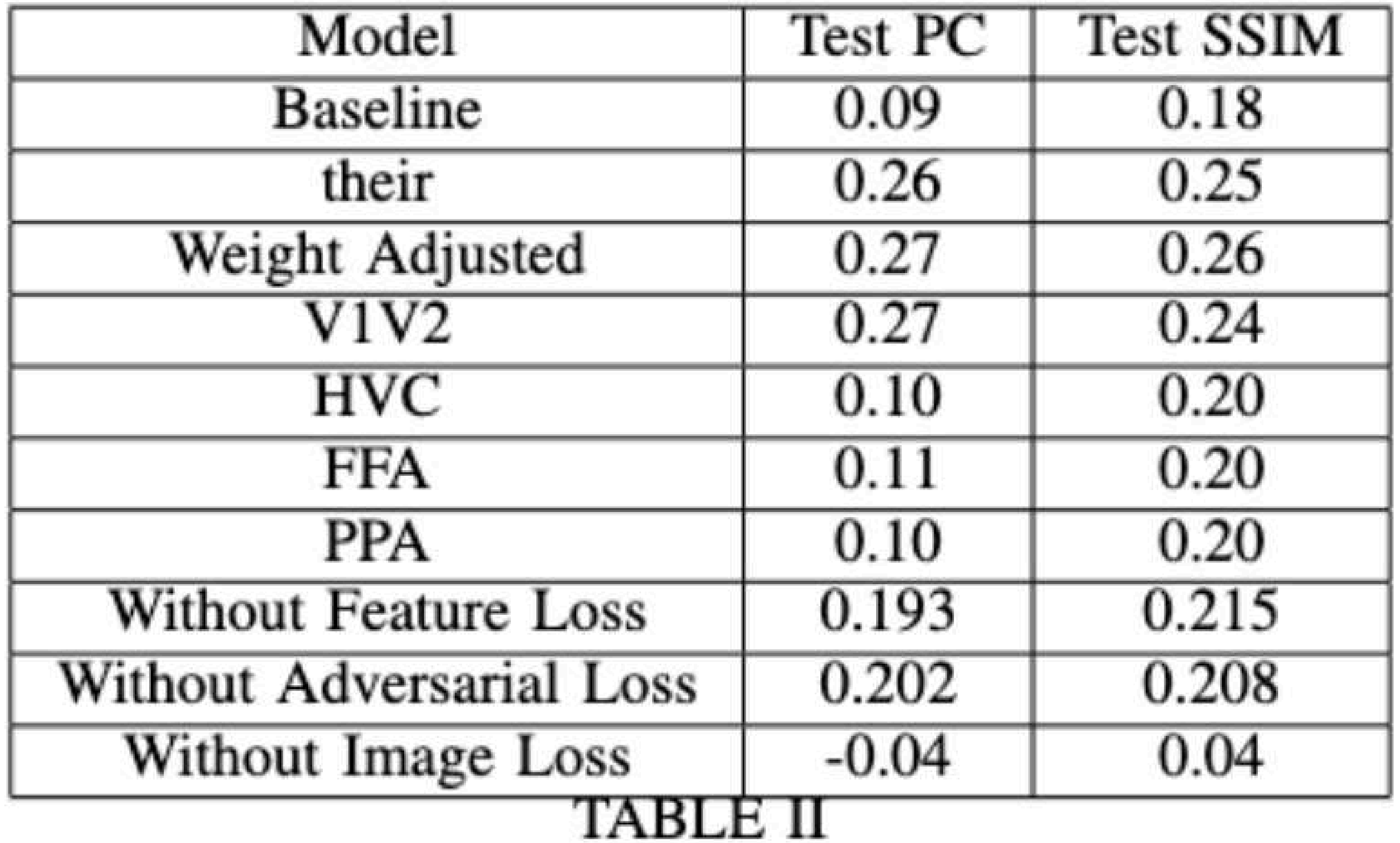
- Lower visual cortex also performs very well on some natural scenes' color reconstruction as well, as it is responsible for color processing.

- Reconstructions done with the higher visual cortex, and its sub regions, particularly, the PPA and FFA, are nowhere close to the previous reconstructions, even in terms of shapes etc. The pearson and SSIM are almost equal to half the ones from before.

- The best model we have trained with the changed weights, gives a Pearson correlation of 96% and SSIM score of 75%.

Ablation Studies, Baseline and Metrics

We have trained the GAN by removing different loss terms, one at a time, as part of our ablation analysis. The metrics that we obtain on the test set for all the models we have trained are shown in the table below. We see that the image loss (MSE) is the most important loss term in the generator's loss, as the metrics obtained after removing image loss are the worst. Also, we have shown the results for the baseline, where the input given to the generator was randomly initialised (with normal distribution).



Model	Test PC	Test SSIM
Baseline	0.09	0.18
their	0.26	0.25
Weight Adjusted	0.27	0.26
V1V2	0.27	0.24
HVC	0.10	0.20
FFA	0.11	0.20
PPA	0.10	0.20
Without Feature Loss	0.193	0.215
Without Adversarial Loss	0.202	0.208
Without Image Loss	-0.04	0.04

TABLE II

TEST SET METRICS FOR DIFFERENT MODELS

Ablation Study Results

Baseline model

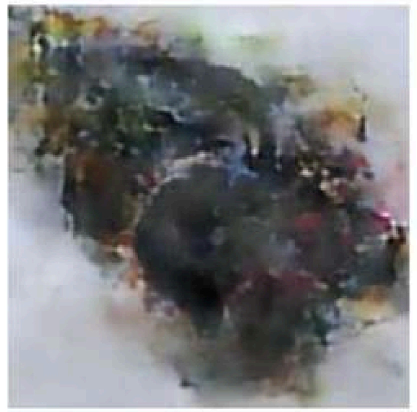
without adversarial loss

without feature loss

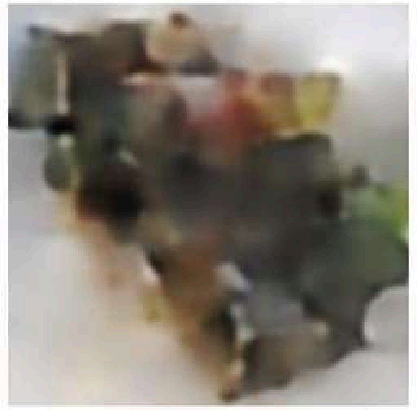
without image loss

Novelty

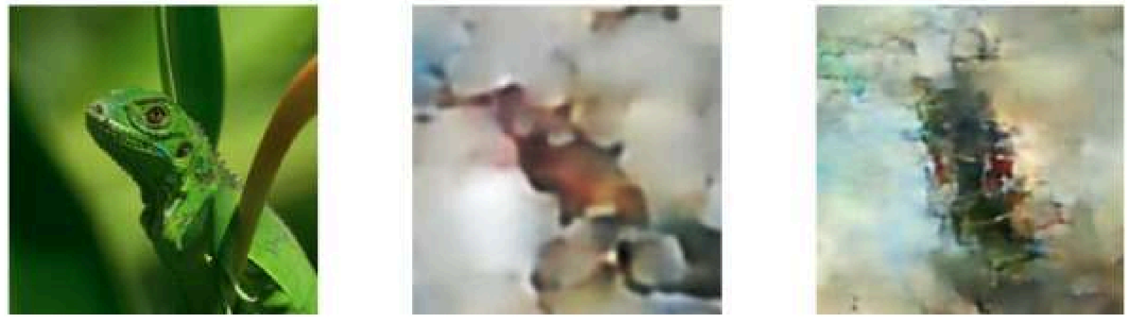
- While the the original paper uses Caffenet, we tried out various models for the comparator. We found that Alexnet worked best and gave a improvement from 0.26 for their model to 0.30 for our model.
- Next we tried to do a ROI wise analysis. Specifically when the reconstruction was done for V1V2, we found that there was considerable improvement in the colors. This is natural because the V1V2 regions are responsible for basic image processing like colors.
- For the HVC(Higher visual cortex) we notice that the metrics are considerably low, because lower level features are all absent here. But some higher level features are clear. Like we can see the artificial objects are constructed much better as we would expect from the HVC. Even with PPA, we notice the same thing. Although, we had fMRI features data for FFA, there are no images of faces to be reconstructed.



Reconstruction with HVC features



Reconstruction with PPA features

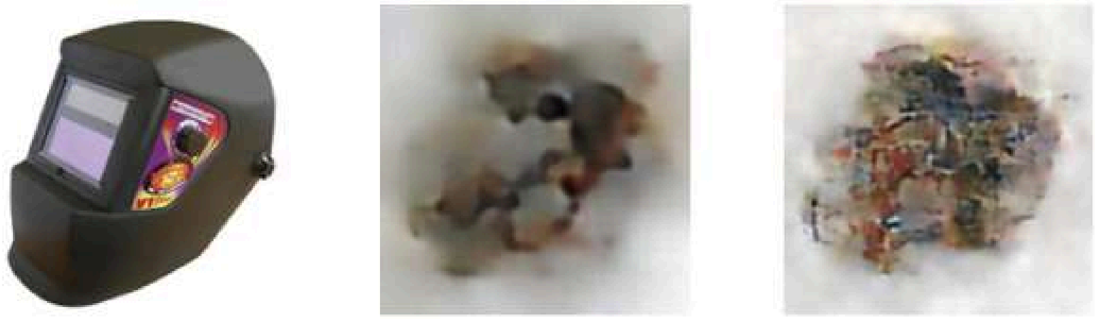


Original

their weights

our weights

Fig. 23. Our weight adjusted (best) model on natural image



Original

their weights

our weights

Fig. 24. Our weight adjusted (best) model on artificial object image

Individual Contribution

- **Rudransh Agarwal:** For interim submission, training of λs for wighting different scaled local element decoders and their analysis for different scale local elements used (2, 3 and 5) (novelty), report writing for interim submission, writing code for training GAN model for half of the ROIs and their analysis, experimenting with weights for combining losses for generator, inferencing and results for models, ablations, final report and slides.

- **Akshit Sharma:** For interim submission, experimenting with different linear models for local element decoders like OMP, Bayesian Ridge and Linear Regressor and generating their results and analysis along with trying differnt scales (3 and 5) for local elements (novelty), report writing for interim submission, training GAN model for the remaining ROIs and their analysis, inferencing and results, ablations, final report and slides.

So, the work done was equally divided between the two team members and the reports and slides were made together.

Challenges Faced

- **Unfamiliarity with GANs-** We had never worked with GANs before. So, we had to understand their architectecture and working before we started working on this part of project.

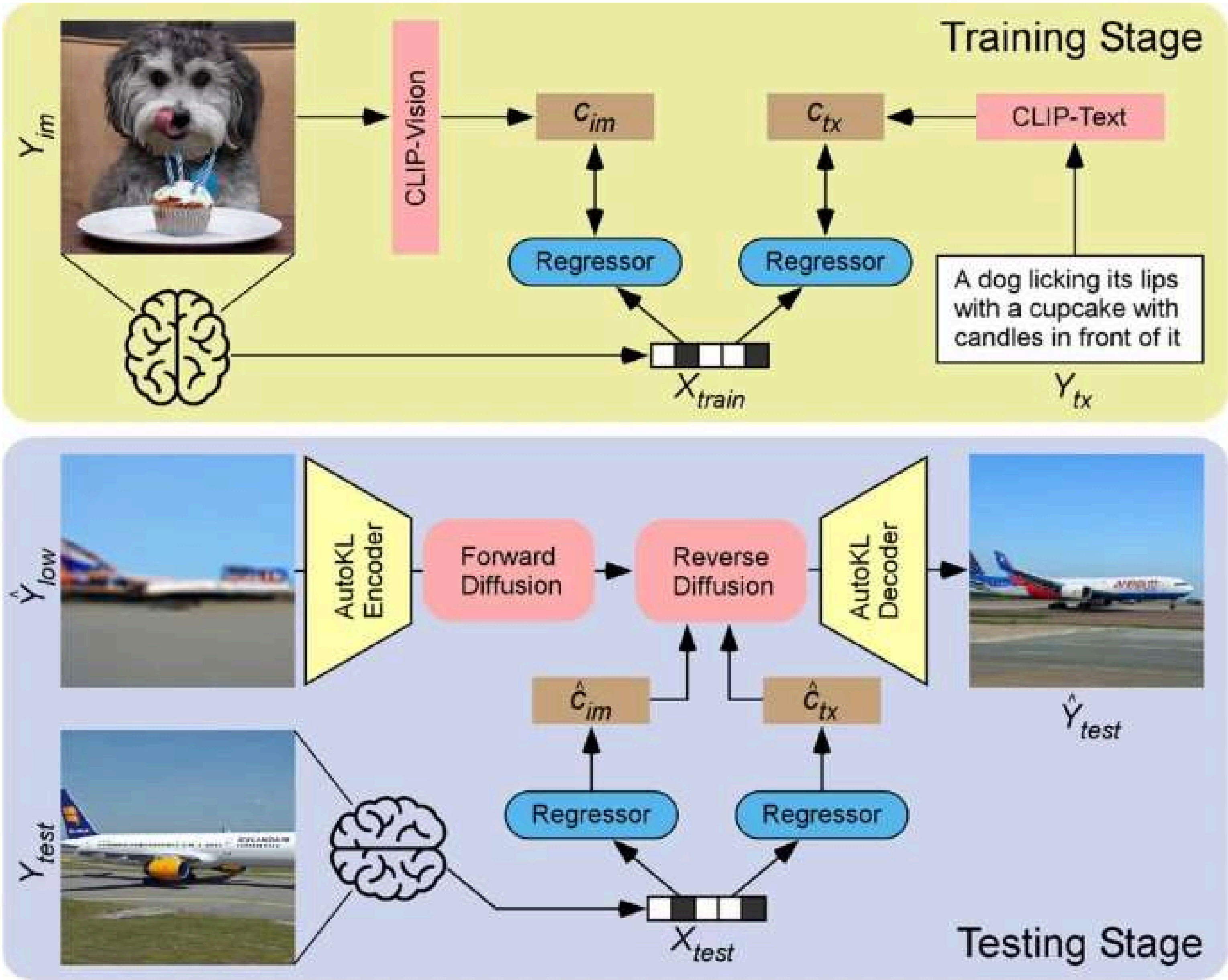
- **Lack of compute and large size of datasets-** Training of GANs requires a lot of GPU compute. With the resources that we have, it took a lot of time and effort to train multiple models, varying ROIs, weights, etc. and then inferencing from them. Also, the datasets for the project were quite large which meant more download, loading time.

Future Perspectives

Natural scenes reconstruction from Generative Latent Diffusion

Link: Click Here

This Nature paper Presents a two-stage scene reconstruction framework called “Brain-Diffuser”. In the first stage, starting from fMRI signals, they reconstruct images that capture low-level properties and overall layout using a VDVAE (Very Deep Variational Autoencoder) model. In the second stage, they use the image-to-image framework of a latent diffusion model (Versatile Diffusion) conditioned on predicted multimodal (text and visual) features, to generate final reconstructed images.

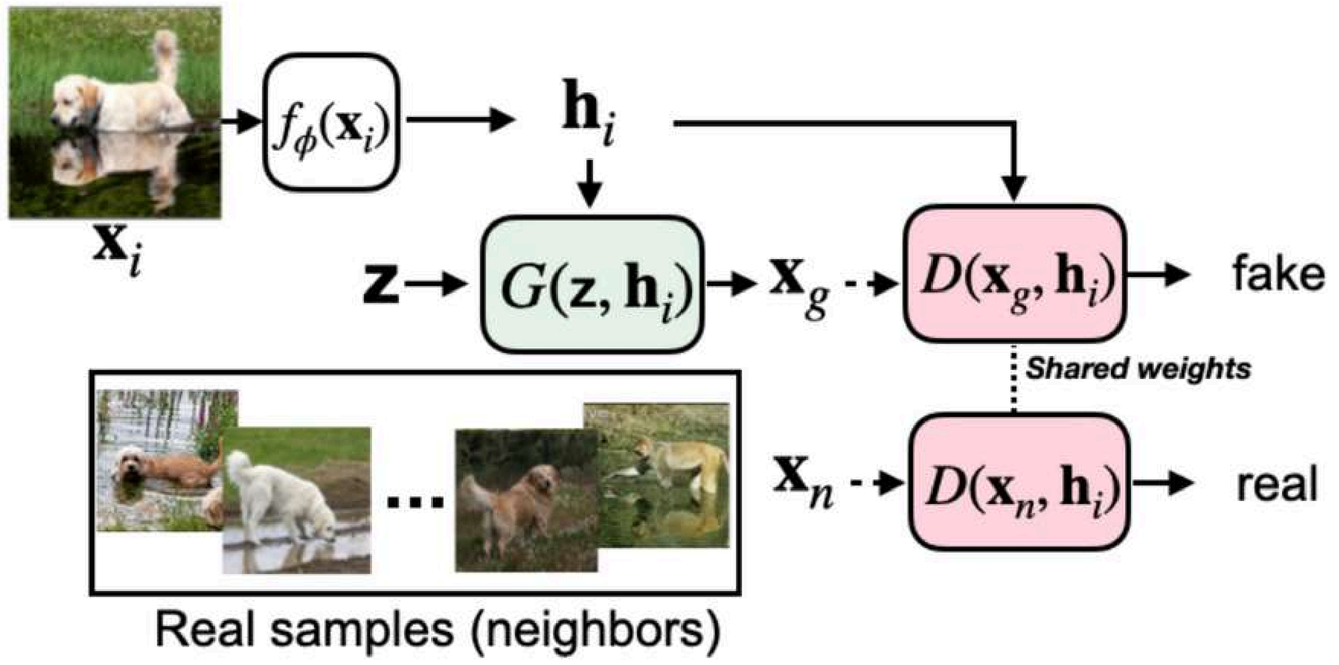


Future Perspectives

Instance-Conditioned GAN

Link: Click Here

They take inspiration from kernel density estimation techniques and introduce a non-parametric approach to modeling distributions of complex datasets. They partition the data manifold into a mixture of overlapping neighborhoods described by a datapoint and its nearest neighbors, and introduce a model, called instance-conditioned GAN (IC-GAN), which learns the distribution around each datapoint. Experimental results on ImageNet and COCO-Stuff show that IC-GAN significantly improves over unconditional models and unsupervised data partitioning baselines. Moreover, they show that IC-GAN can effortlessly transfer to datasets not seen during training by simply changing the conditioning instances, and still generate realistic images.



(b) Schematic illustration of the IC-GAN workflow

LINK FOR REPORT (OVERLEAF): Click Here

(PDF included in submission on moodle for report)

THANK YOU!