GANalyze: Toward Visual Definitions of Cognitive Image Properties

ICCV'19

Lore Goetschalckx*, Alex Andonian, Aude Oliva, Phillip Isola

Motivation

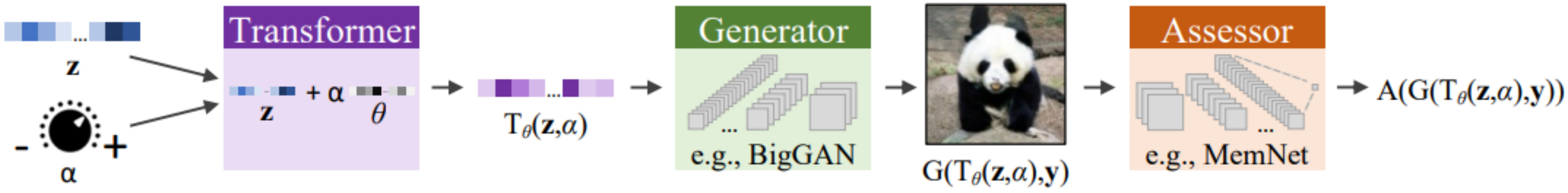
- Why do we remember the things we do?
- A slew of work have been provided to explain ‘something’ that positively affects our memory.
- But a picture is worth a thousand words.
- What does it look like to make an image more or less memorable?

Contribution

- Introducing GANalyze, that uses GANs to provide a visual definition of image properties, like memorability and aesthetics, that we can measure but are not easy, in words, to define.
- Showing that this framework surfaces previously overlooked attributes that correlate with memorability. (this may be concerned with cognitive science?)
- Demonstrating that the discovered transformations have a causal effect on memorability.

Method

- Overview



$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{z},\mathbf{y},\alpha}[(A(G(T_\theta(\mathbf{z},\alpha),\mathbf{y})) - (A(G(\mathbf{z},\mathbf{y})) + \alpha))^2]$$

$$T_\theta(\mathbf{z},\alpha) = \mathbf{z} + \alpha\theta$$

$$\theta^* = \mathrm{argmin}_\theta\, \mathcal{L}(\theta)$$

$$\mathbf{y} \in \{0;1\}^{1\times C}$$

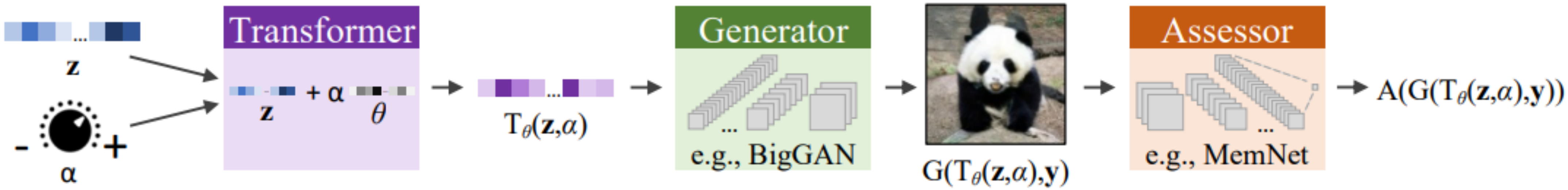
$$\mathbf{z} \in \mathbb{R}^{1\times M}$$

$$\theta \in \mathbb{R}^{1\times M}$$

A is a fixed assessor that quantitatively evaluates memory.
Y is a class.

Method

- Overview

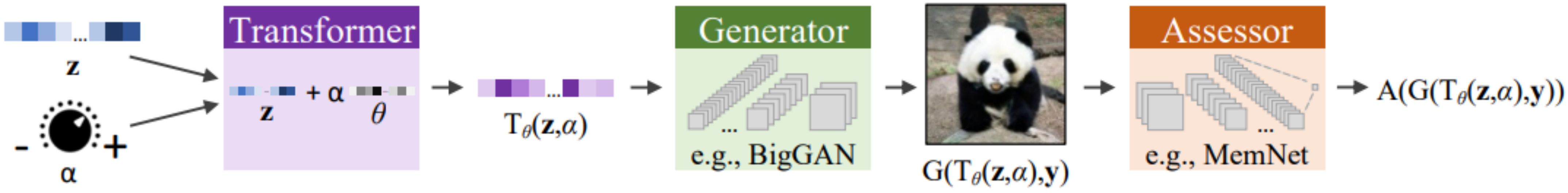


training

To train our model and find θ^*, we built a training set by randomly sampling 400K $\mathbf{z}$ vectors from a standard normal distribution truncated to the range $[-2, 2]$. Each $\mathbf{z}$ was accompanied by an α value, randomly drawn from a uniform distribution between -0.5 and 0.5, and a randomly chosen $\mathbf{y}$. We used a batch size of 4 and an Adam optimization procedure.

Method

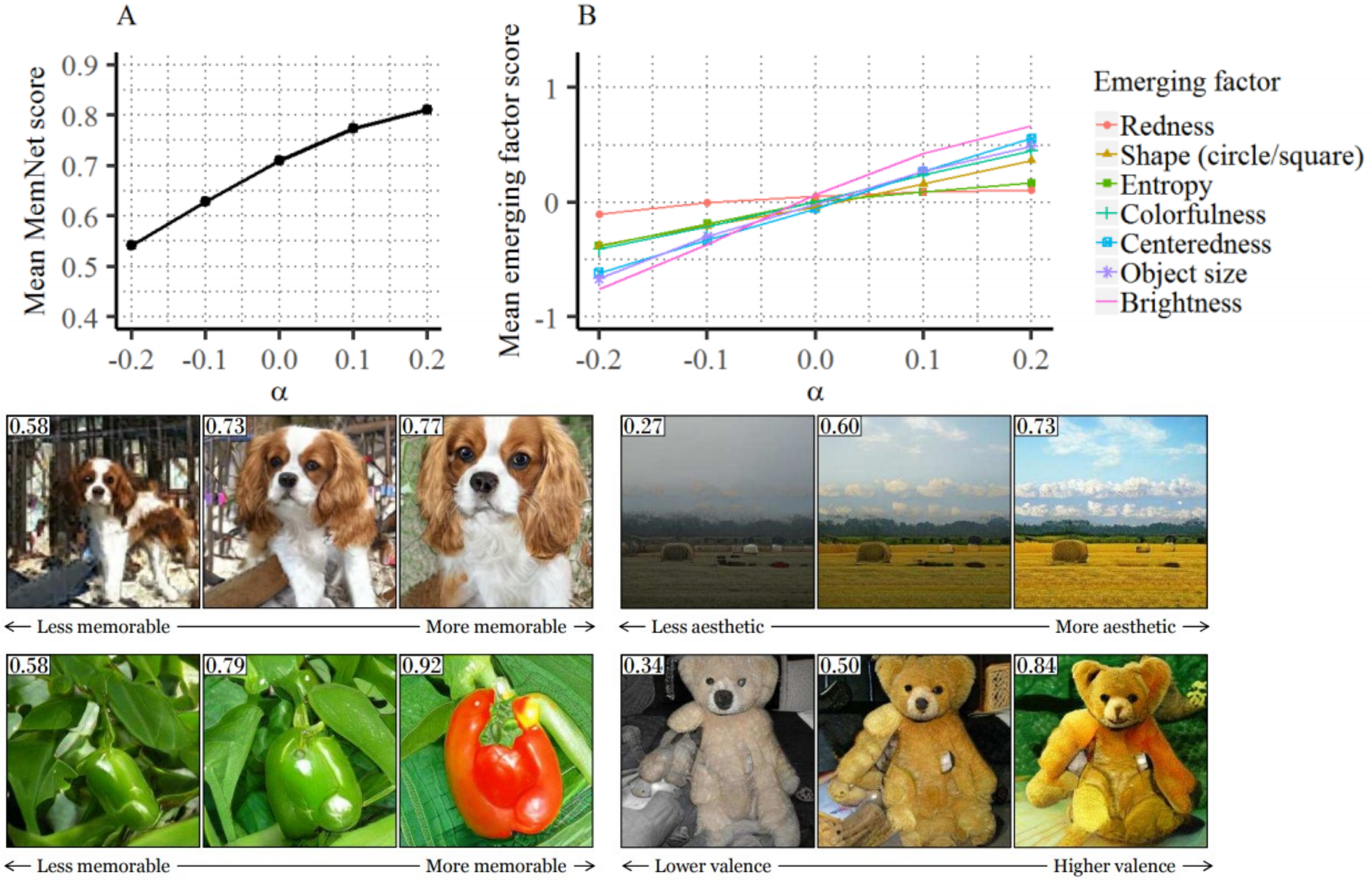
- Overview



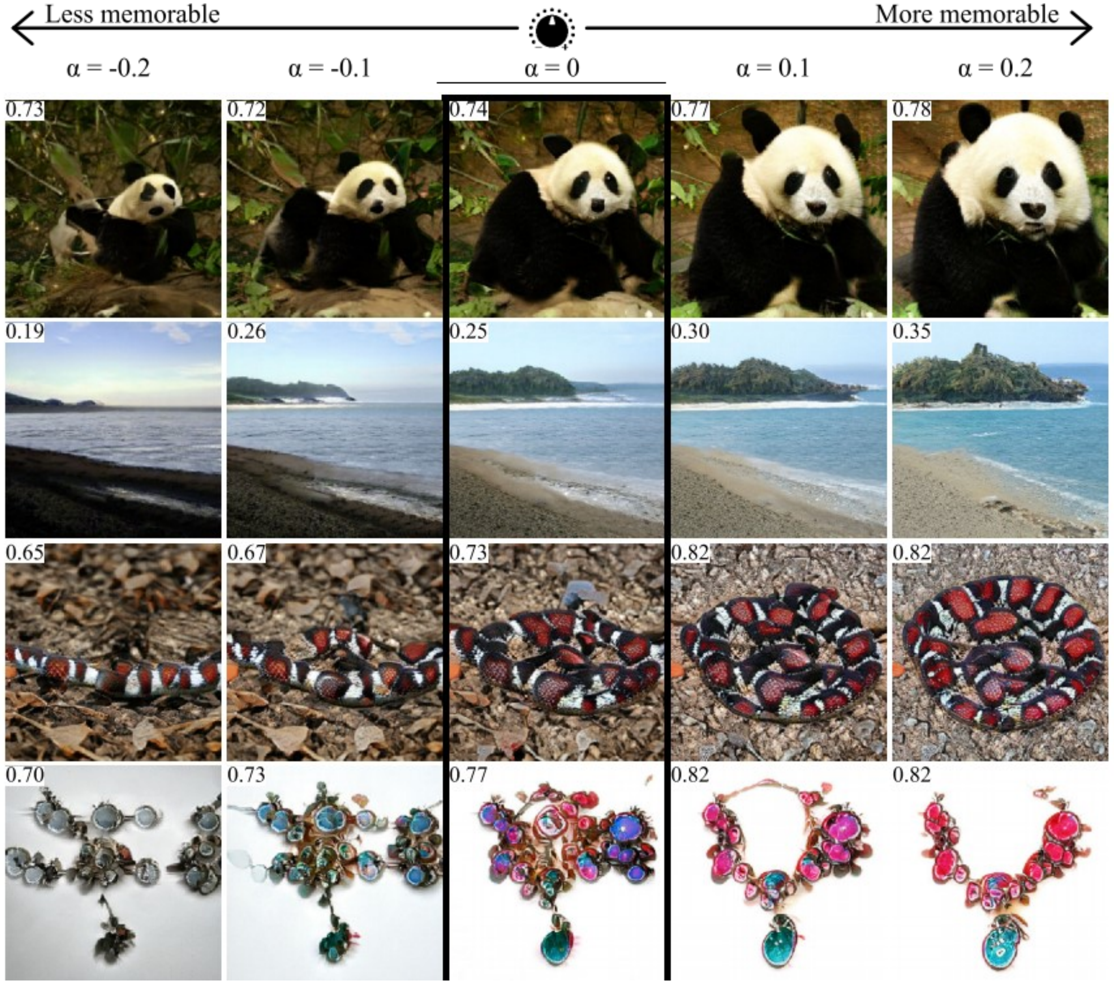
test

In view of the behavioral experiments (see Section 3), we restricted the test set to 750 randomly chosen ImageNet classes and two $\mathbf{z}$ vectors per class. Each $\mathbf{z}$ vector was then paired with five different α values: $[-0.2, -0.1, 0, 0.1, 0.2]$. Note that this includes an α of 0, representing the original image $G(\mathbf{z}, \mathbf{y})$. Finally, the test set consisted of 1.5K sets of five images, or 7.5K test images in total.

Experiments



Experiments

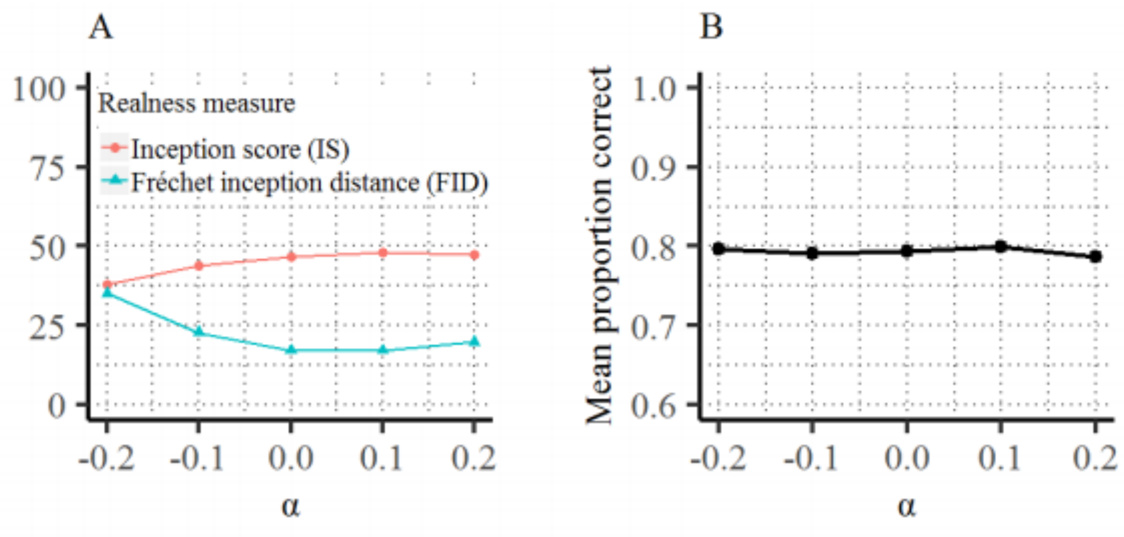


Experiments

Emerging factors

- **Brightness** is measured as the average pixel value after transforming the image to grayscale.
- For **colorfulness**, authors used the metric proposed by previous work.
- **Redness** is evaluated by computing the normalized number of red pixels.
- Entropy of the pixel intensity histogram was taken as proxy for **simplicity**.
- To capture **object size**, authors calculate the difference in the mask's area (normalized number of pixels) as the step size α varied.
- To measure **centeredness**, authors compute the deviation of the mask's centroid from the center of the frame.

Experiments

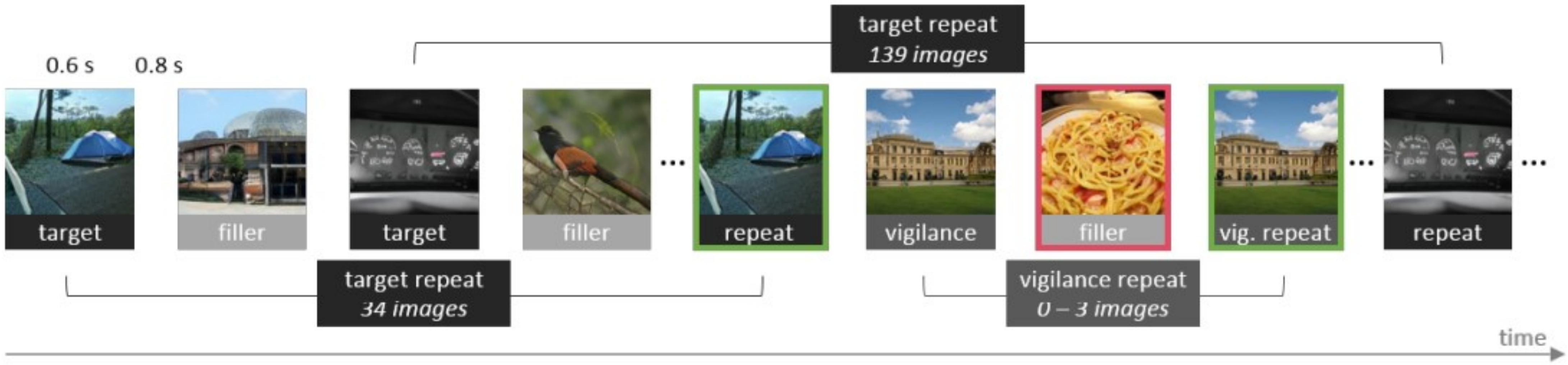


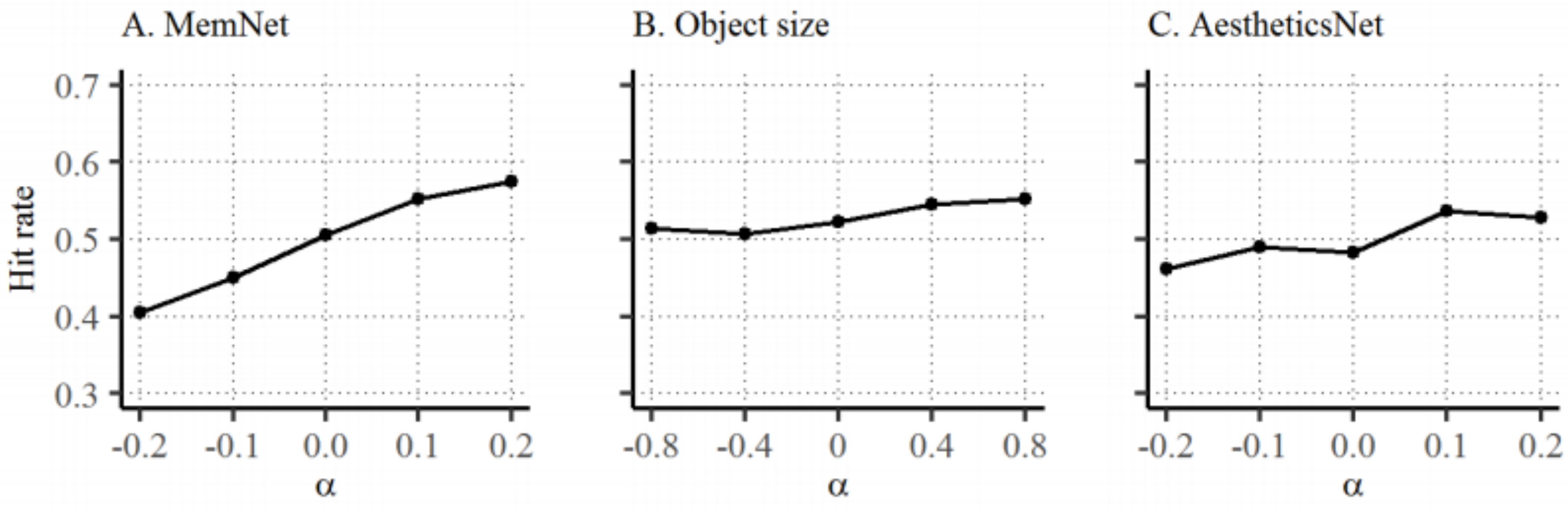
A two-alternative forced choice (2AFC) task, hosted on Amazon Mechanical Turk.

Each series consisted of 100 trials, of which 20 were vigilance trials.

For the vigilance trials, we generated GAN-images from z vectors that were sampled from the tails of a normal distribution (to make them look less real).

Experiments





Experiments