A. Implementation Details

We use the first two pixel-based stages of the DeepFloyd IF [24] diffusion model. Specifically, we use the first stage which produces images of size 64×64, and the second stage which upsamples images to 256×256. Our method is applied in both stages, by implementing view transformations for both resolutions. DeepFloyd IF additionally predicts the variance, along with a noise estimate. We reduce multiple variance estimates by also taking a mean. We use a classifier guidance strength between 7 and 10, and between 30 and 100 inference steps depending on the prompt. We use the M size models for both stages.

Because DeepFloyd IF also estimates variances, we need to apply inverse views to these variance estimates, in addition to the noise estimates. For pixel permutation based views, we simply apply the inverse permutation to the variance estimates. For inversion, the inverse transformation would be negating the predicted logged variance, which does not make sense. We find that simply not inverting the variance estimates works well in this case.

DeepFloyd IF additionally uses a third super resolution stage, which is the Stable Diffusion x4 upscaler. This model upscales from 256×256 to 1024×1024. Because this model is a latent model, we do not apply our method to it. However, we find that we can use it with no modification to upscale our illusions without any loss in quality in the different views. We do this by upsampling conditioned on the prompt associated with the identity view. All results in Fig. 1 have been upsampled in this way.

B. Dataset Collection

Our dataset consists of a list of styles, such as `"a street art of..."` or `"an oil painting of..."`, and a list of subjects such as `"an old man"` or `"a snowy mountain village"`. Subjects and styles were chosen by hand, using GPT-3.5 for inspiration. Prompt pairs are generated by randomly sampling a style prompt and prepending it to two randomly chosen subject prompts.

The CIFAR dataset was constructed by taking the 10 classes of CIFAR-10 as our subjects, and using the prompt `"a painting of"` as the style prompt. We take all 45 pairs of subjects, and prepend the style prompt to the subject prompts, resulting in 45 prompt pairs.

C. Additional Results

We provide additional qualitative results in this section. In Fig. 10, we compare our method to baselines, using prompts from our dataset and the CIFAR prompt dataset. This is an extension of Fig. 5. We also generate more illusions with 90° and 180° rotations, ambigrams, "polymorphic" jigsaw puzzles, color inversion, and vertical flips, which can all be found in Fig. 12. In Fig. 13, we generate several flip illusions with the same flipped prompt, and different unflipped prompts, and we show flipped versions of these illusions in Fig. 14.

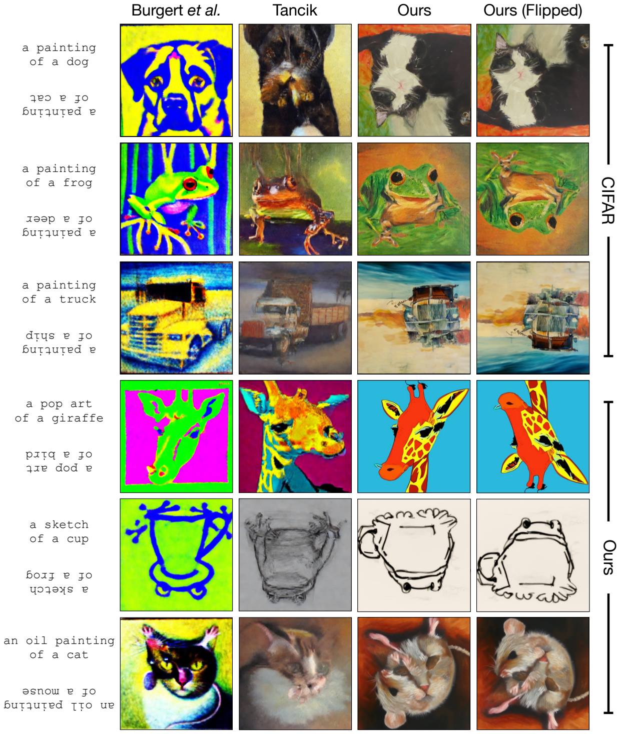


Figure 10. **Qualitative Comparisons.** We compare more illusions generated by baselines to our illusions. We show examples from both our prompt dataset and the CIFAR prompt dataset.

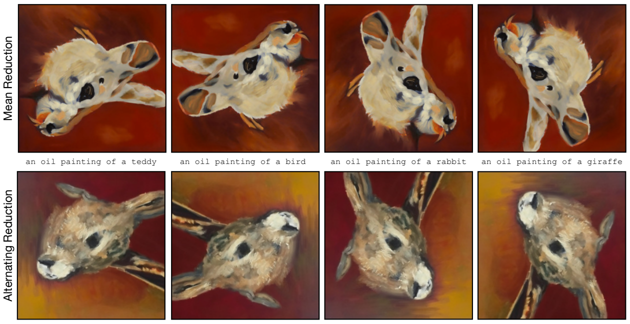

Mean Reduction
Alternating Reduction
an oil painting of a teddy
an oil painting of a bird
an oil painting of a rabbit
an oil painting of a giraffe

Figure 11. **Combining Noise Estimates.** We show that mean reduction does better than alternating with an example of a 4-view sample image.

D. Random Samples

We provide more random samples generated using our method. For rotations, color inversion, and vertical flips, please refer to Fig. 16. For three-view, inner rotation, "polymorphic" jigsaw puzzles, and patch and pixel permutation views, please refer to Fig. 17. We also provide random samples generated with prompts from the CIFAR dataset in Fig. 15.

The CIFAR prompt pair results, in Fig. 15 as well

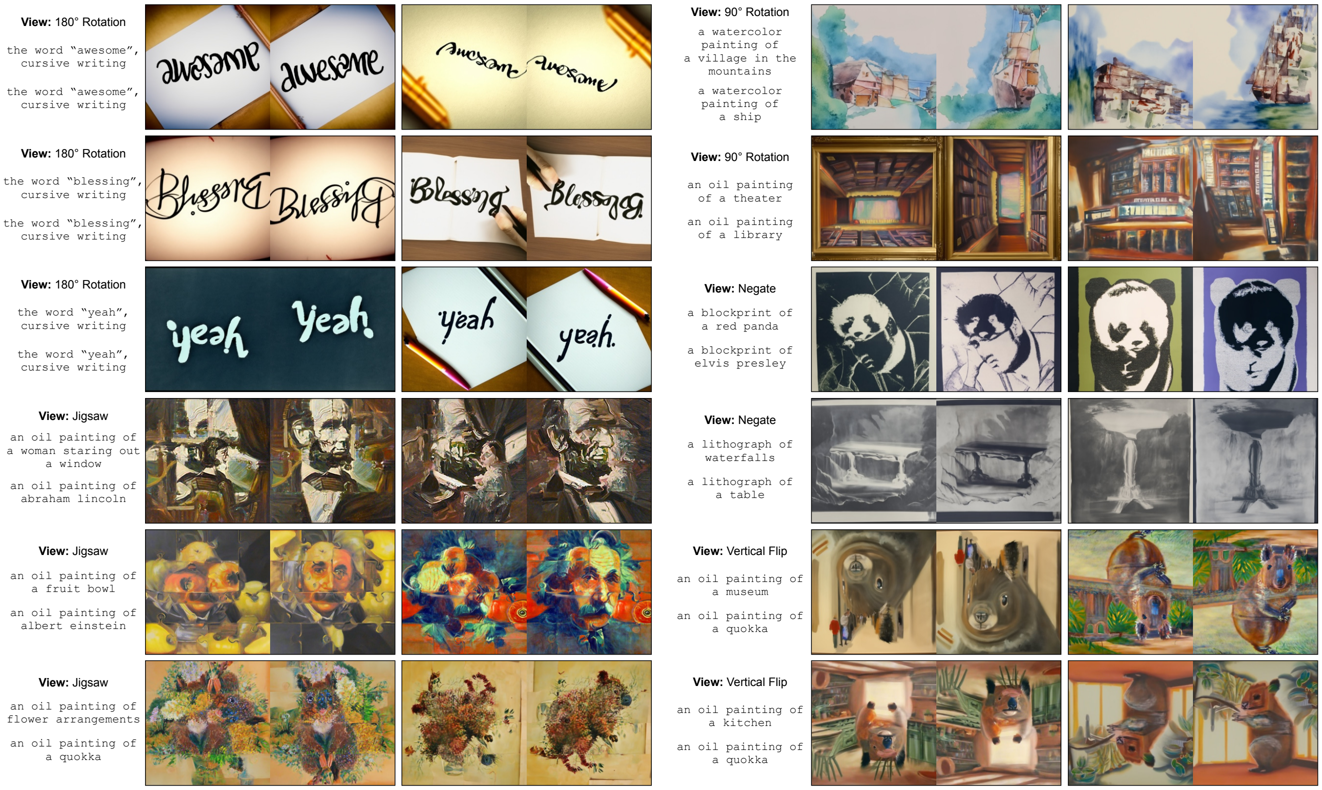

View: 180° Rotation
the word "awesome", cursive writing
the word "awesome", cursive writing
View: 180° Rotation
the word "blessing", cursive writing
the word "blessing", cursive writing
View: 180° Rotation
the word "yeah", cursive writing
the word "yeah", cursive writing
View: Jigsaw
an oil painting of a woman staring out a window
an oil painting of abraham lincoln
View: Jigsaw
an oil painting of a fruit bowl
an oil painting of albert einstein
View: Jigsaw
an oil painting of flower arrangements
an oil painting of a quokka
View: 90° Rotation
a watercolor painting of a village in the mountains
a watercolor painting of a ship
View: 90° Rotation
an oil painting of a theater
an oil painting of a library
View: Negate
a blockprint of a red panda
a blockprint of elvis presley
View: Negate
a lithograph of waterfalls
a lithograph of a table
View: Vertical Flip
an oil painting of a museum
an oil painting of a quokka
View: Vertical Flip
an oil painting of a kitchen
an oil painting of a quokka

Figure 12. **Qualitative Samples.** We show more illusions with views such as rotations, flips, color inversion, and jigsaw puzzles.

as Tab. 1, are included as a proxy for random prompts. We note that systematically sampling truly random prompts for evaluation is tricky. Firstly, there is no standard method for sampling a random prompt. And secondly, not all prompt pairs make for good illusions. A straightforward example of this would be prompt pairs that differ in style. Therefore, evaluating illusion generation on completely random prompts may result in meaningless or misleading results, and as such prompts in Fig. 8, Fig. 16, and Fig. 17 are to some extent curated.

E. The Art of Choosing Prompts

We find that choosing good prompts is important to achieving good illusions. We lay out a few rules of thumb here. Firstly, it is very hard to reason as to what will make good illusions. Prompts that one may believe to work easily can fail consistently, and prompts that one may believe to have no chance of working may work fantastically. We find that more abstract styles, such as "a painting" or "a drawing" work much better than realistic styles such as "a photo of". We believe this is because the constraints on realistic styles is too strong for illusions to work well. We also find that human faces make for good illusions, perhaps due to the sensitivity of the human visual system to face-like stimuli.

F. Jigsaw Puzzle Implementation

We produce jigsaw puzzles by implementing a rearrangement of puzzle pieces as a permutation of pixels. We first hand-draw three puzzle pieces—a corner, edge, and center piece—such that they can disjointly tile a 64×64, a 256×256, or a 1024×1024 image. All pieces in the puzzle are one of these three pieces, in different orientations. We then sample a random permutation of corner, edge, and center pieces respectively, and translate this permutation of pieces to a permutation of pixels.

G. Combining Noise Estimates

Rather than taking the mean of noise estimates, we also experimented with alternating or cycling through noise estimates by timestep, as is done in [42]. However, we find that this can lead to "thrashing," in which the sample is optimized in different directions at different timesteps, leading to poor quality. Moreover, in illusions with more than two views, each view gets fewer denoising steps, resulting in lower quality illusions. For example, given four prompts each matched to a rotation of the image (i.e., "a teddy", "a bird", "a rabbit", and "a giraffe"), the mean reduction outputs images with higher quality than the alternating method as shown in Fig. 11.

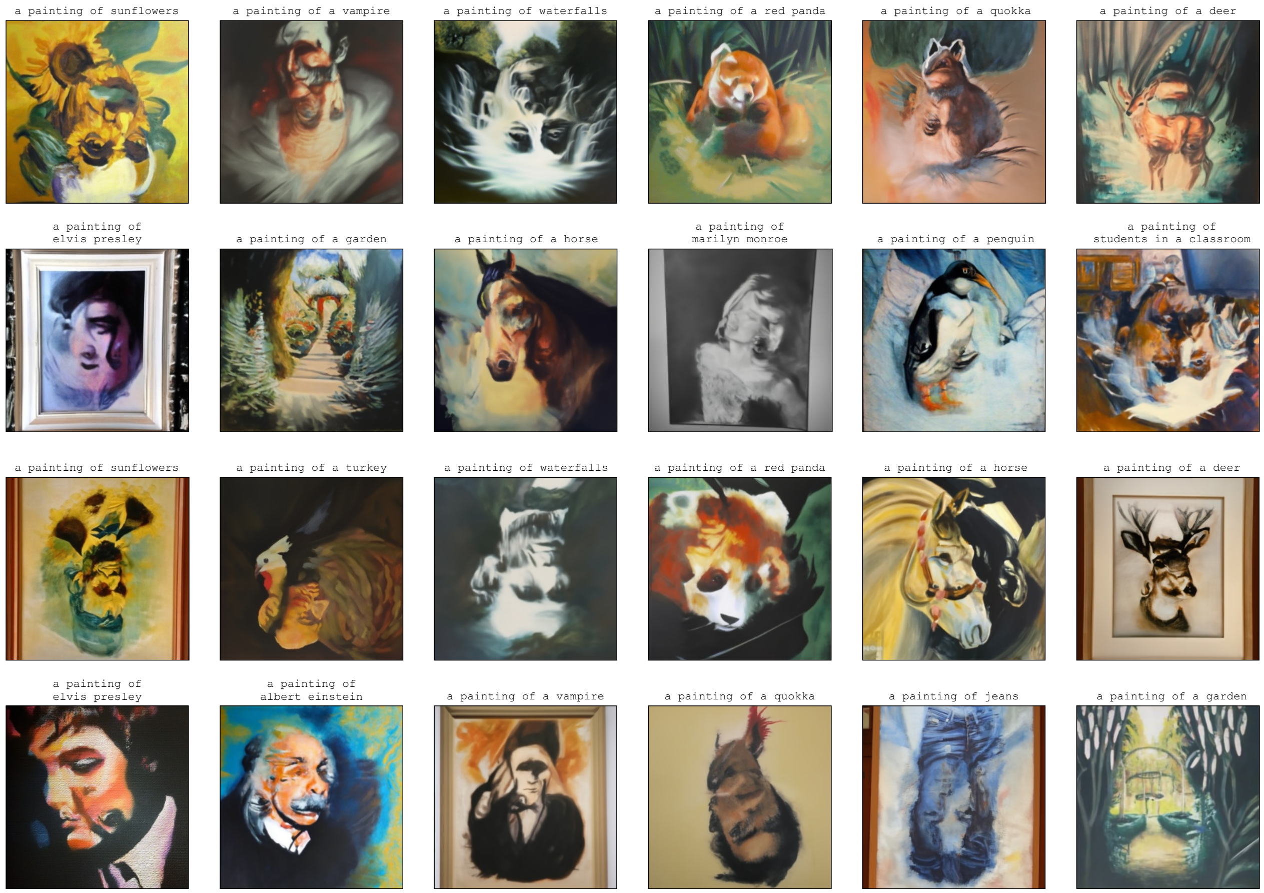


Figure 13. **Flip illusions.** For each row, the prompt of the flipped image is the same. We encourage the reader to guess what the flipped prompt is. For an answer and flipped illusions, please see Fig. 14.

H. Linearity of Views

As discussed in Sec. 3.3, when a view v is a linear transformation, it satisfies:

$$v(\mathbf{x}_t) = v(w_t^{\text{signal}}\mathbf{x}_0 + w_t^{\text{noise}}\epsilon) \tag{10}$$

$$= w_t^{\text{signal}}v(\mathbf{x}_0) + w_t^{\text{noise}}v(\epsilon). \tag{11}$$

This is convenient because applying v to the noisy image $\mathbf{x}_t$ is equivalent to applying v to the signal, $\mathbf{x}_0$, and the noise, ϵ, independently. In addition, the result is a linear combination of transformed signal and transformed noise, and is weighted as the diffusion model expects for timestep t.

However, there may be other conditions that work. For example, we could enforce

$$v(\mathbf{x}_t) = v(w_t^{\text{signal}}\mathbf{x}_0 + w_t^{\text{noise}}\epsilon) \tag{12}$$

$$= w_t^{\text{signal}}v_1(\mathbf{x}_0) + w_t^{\text{noise}}v_2(\epsilon), \tag{13}$$

with the interpretation being that v somehow acts on the signal and noise in different ways, through v_1 and v_2, and combines them with the correct weightings. We leave this for future work.

I. Statistical Consistency

We provide a proof that for $\epsilon \sim \mathcal{N}(0, I)$ and square matrix $\mathbf{A}$, $\mathbf{A}\epsilon \sim \mathcal{N}(0, I)$ if and only if $\mathbf{A}$ is orthogonal, stated in Sec. 3.3. By properties of Gaussians, $\mathbf{A}\epsilon$ is also Gaussian, so we need only compute mean and covariances. The mean is given by

$$\mathbb{E}[\mathbf{A}\epsilon] = \mathbf{A}\mathbb{E}[\epsilon] = 0. \tag{14}$$

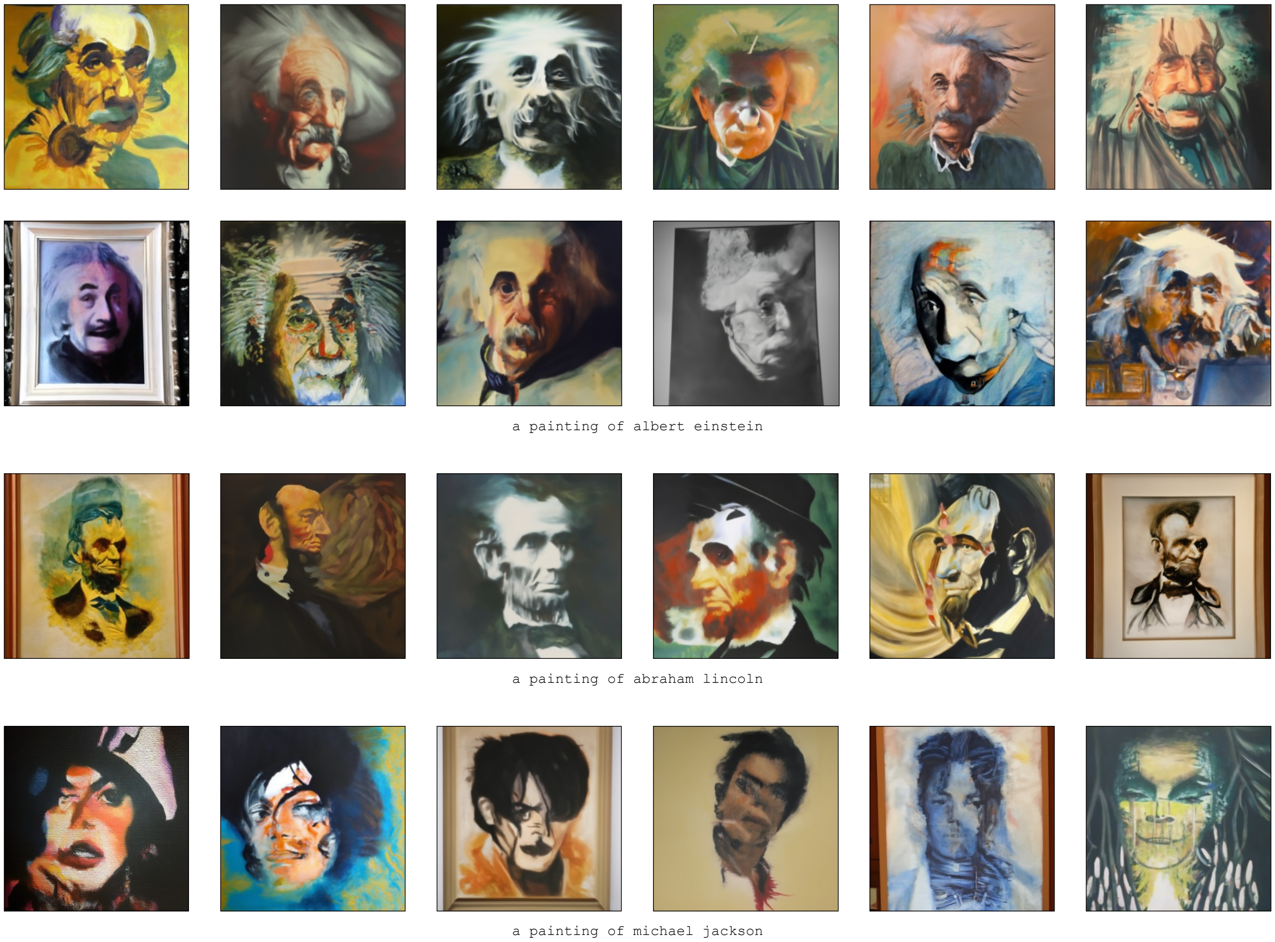
Because the mean is 0, the covariance is given by

$$\text{Cov}(\mathbf{A}\epsilon) = \mathbb{E}[(\mathbf{A}\epsilon)(\mathbf{A}\epsilon)^\intercal] \tag{15}$$

$$= \mathbf{A}\mathbb{E}[\epsilon\epsilon^\intercal]\mathbf{A}^\intercal \tag{16}$$

$$= \mathbf{A}\mathbf{A}^\intercal \tag{17}$$

So if $\mathbf{A}\epsilon \sim \mathcal{N}(0, I)$, then we must have $\text{Cov}(\mathbf{A}\epsilon) = \mathbf{A}\mathbf{A}^\intercal = I$, or equivalently $\mathbf{A}$ must be orthogonal. And if $\mathbf{A}$ is orthogonal, then $\mathbf{A}\mathbf{A}^\intercal = I$ and $\mathbf{A}\epsilon \sim \mathcal{N}(0, I)$.



a painting of albert einstein

a painting of abraham lincoln

a painting of michael jackson

Figure 14. **Flip illusions.** Flipped illusions from Fig. 13, revealing the flipped prompt. Please refer to Fig. 13 for the unflipped images.

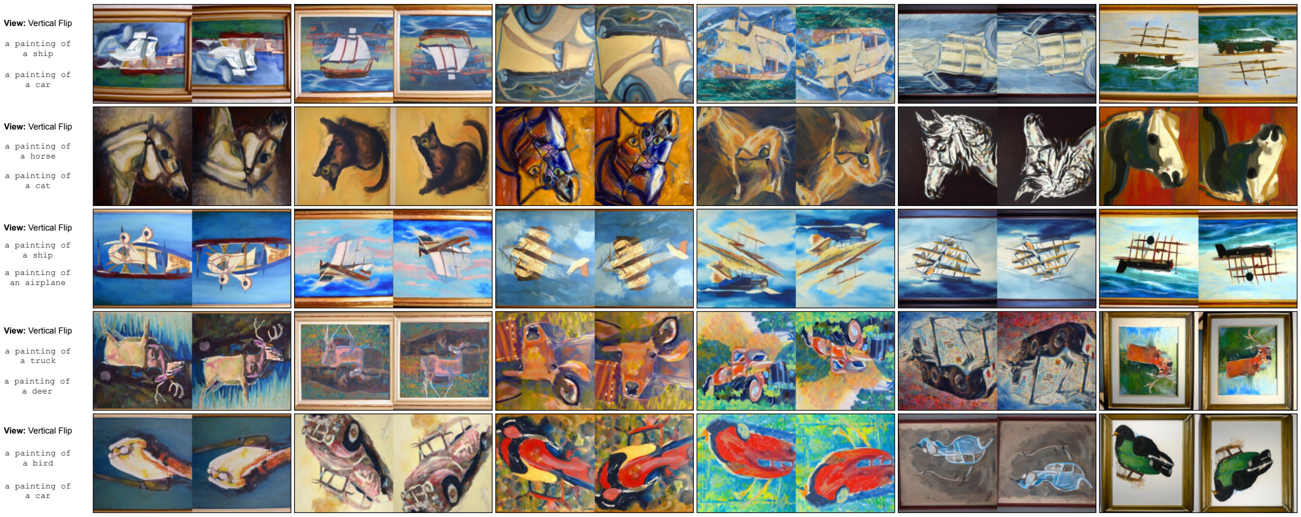


Figure 15. **Random Samples.** We provide random samples for vertical flips using prompts from the CIFAR dataset. We show both views of the illusions side-by-side.

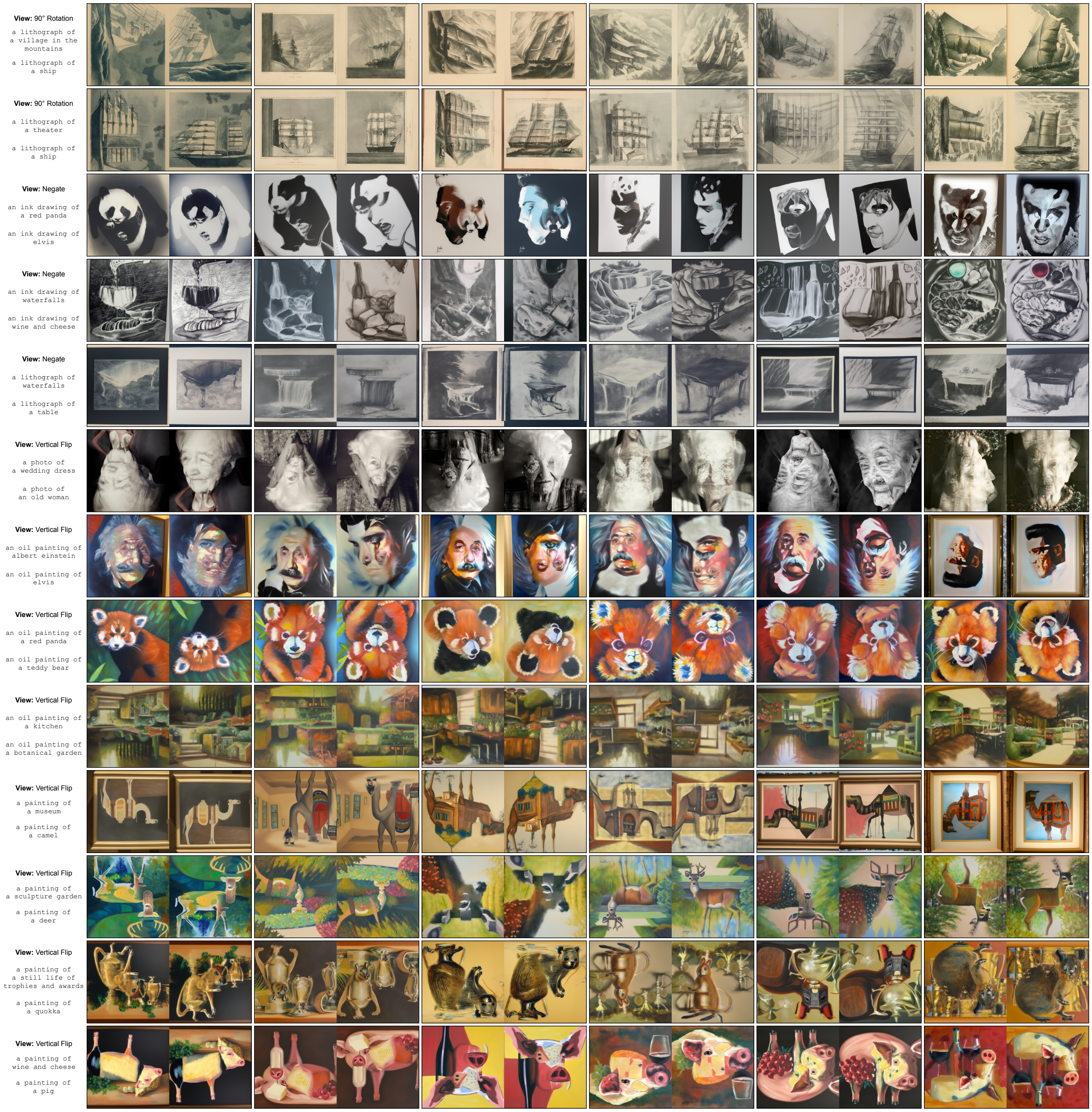

View: 90° Rotation
a lithograph of a village in the mountains
a lithograph of a ship
View: 90° Rotation
a lithograph of a theater
a lithograph of a ship
View: Negate
an ink drawing of a red panda
an ink drawing of elvis
View: Negate
an ink drawing of waterfalls
an ink drawing of wine and cheese
View: Negate
a lithograph of waterfalls
a lithograph of a table
View: Vertical Flip
a photo of a wedding dress
a photo of an old woman
View: Vertical Flip
an oil painting of albert einstein
an oil painting of elvis
View: Vertical Flip
an oil painting of a red panda
an oil painting of a teddy bear
View: Vertical Flip
an oil painting of a kitchen
an oil painting of a botanical garden
View: Vertical Flip
a painting of a museum
a painting of a camel
View: Vertical Flip
a painting of a sculpture garden
a painting of a deer
View: Vertical Flip
a painting of a still life of trophies and awards
a painting of a quokka
View: Vertical Flip
a painting of wine and cheese
a painting of a pig

Figure 16. **Random Samples.** We provide random samples for rotations, negations, and vertical flip views. We show both views of the illusions side-by-side.

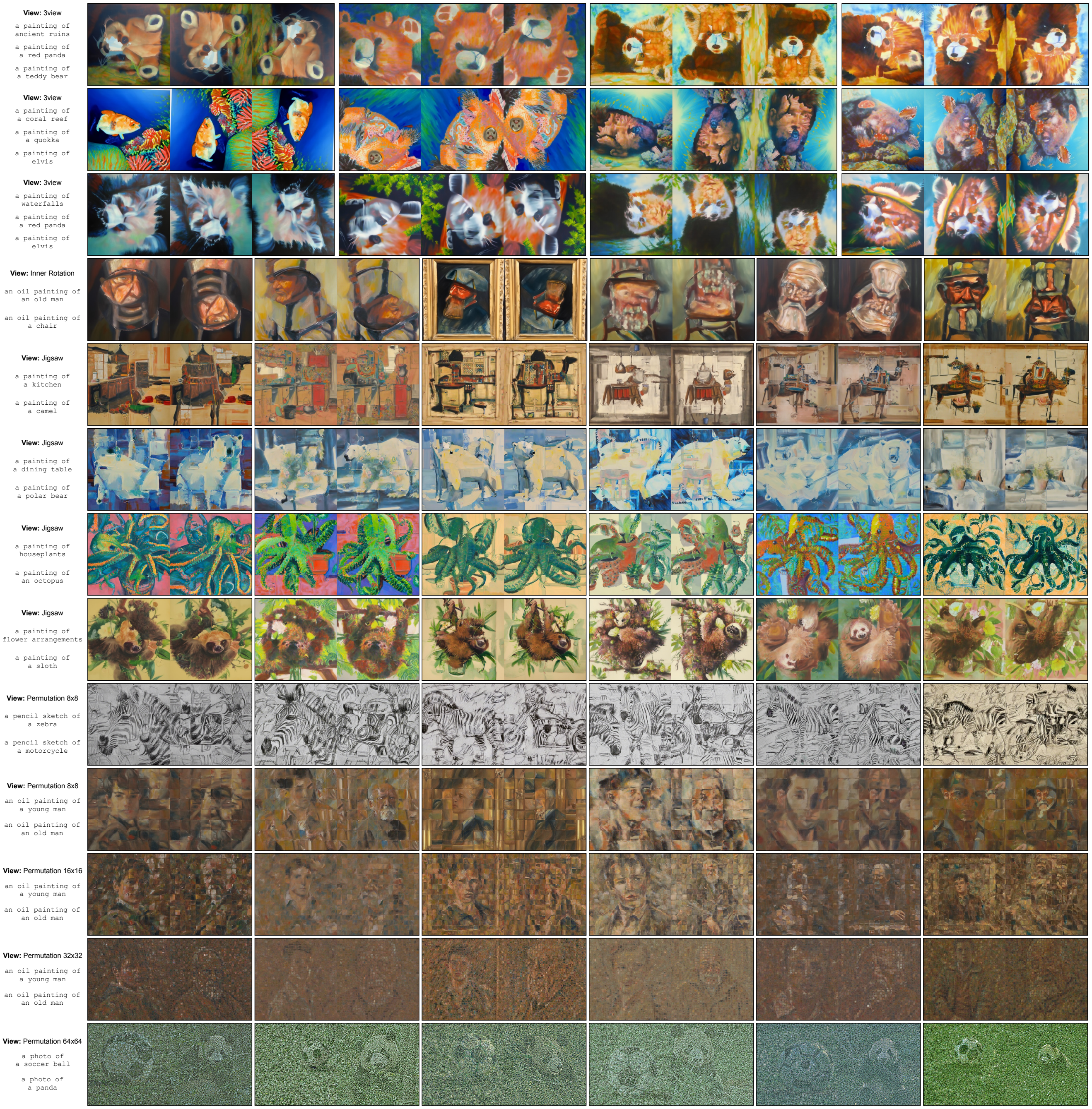

View: 3view
a painting of ancient ruins
a painting of a red panda
a painting of a teddy bear
View: 3view
a painting of a coral reef
a painting of a quokka
a painting of elvis
View: 3view
a painting of waterfalls
a painting of a red panda
a painting of elvis
View: Inner Rotation
an oil painting of an old man
an oil painting of a chair
View: Jigsaw
a painting of a kitchen
a painting of a camel
View: Jigsaw
a painting of a dining table
a painting of a polar bear
View: Jigsaw
a painting of houseplants
a painting of an octopus
View: Jigsaw
a painting of flower arrangements
a painting of a sloth
View: Permutation 8x8
a pencil sketch of a zebra
a pencil sketch of a motorcycle
View: Permutation 8x8
an oil painting of a young man
an oil painting of an old man
View: Permutation 16x16
an oil painting of a young man
an oil painting of an old man
View: Permutation 32x32
an oil painting of a young man
an oil painting of an old man
View: Permutation 64x64
a photo of a soccer ball
a photo of a panda

Figure 17. **Random Samples.** We provide random samples for 3-view, inner rotation, jigsaw puzzle, and patch and pixel permutation views. We show all views of the illusions side-by-side.

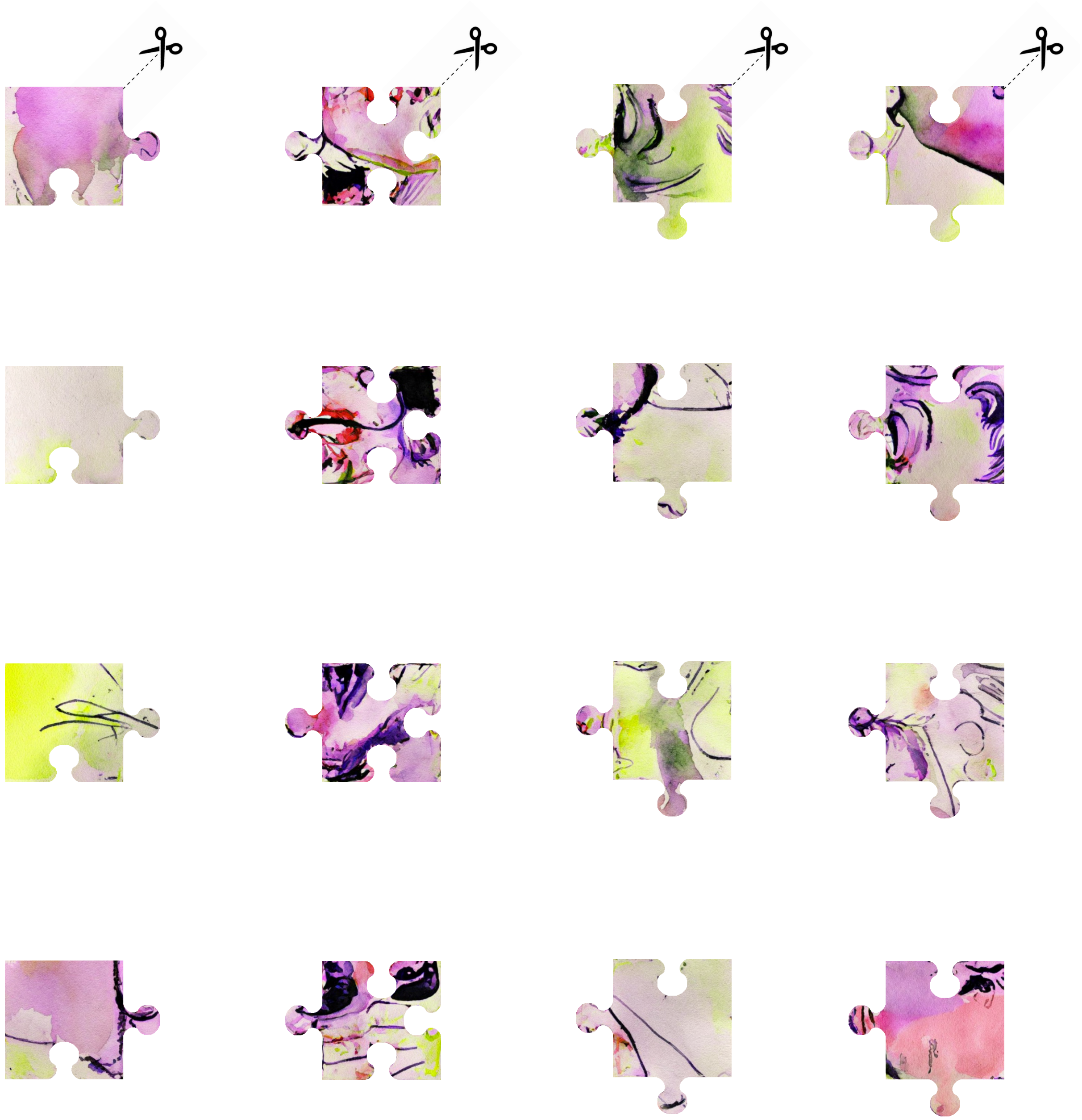


Figure 18. **Cut Your Own Polymorphic Jigsaw!** We invite the reader to cut out their own polymorphic jigsaw puzzle, and try to discover both solutions.