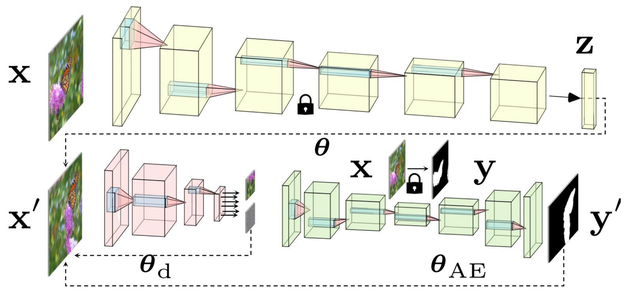
MAGIC: Mask-Guided Image Synthesis by Inverting a Quasi-Robust Classifier

Mozhdeh Rouhsedaghat, Masoud Monajatipoor, C.-C. Jay Kuo, Iacopo Masi

Motivation

With advances of deep learning techniques and the availability of large annotated datasets, image synthesis methods could achieve promising results. However, for synthesizing and manipulating rare or "long tail" images which their data distribution is not effectively learned, such methods tend to perform poorly. **One-shot image synthesis** is about using a single image as the training data for the image synthesis task which not only addresses the mentioned challenge but also obviates the need for large annotated datasets.

MAGIC Architecture



$$\mathbf{x}' = \arg\min_{\mathbf{x}'} \underbrace{\ell(\boldsymbol{\theta}(\mathbf{x}'), c)}_{\substack{\text{semantics via}\\ \text{quasi robust}\\ \text{inversion}}} + \eta \underbrace{\rho_{\boldsymbol{\theta}_\text{d}}(\mathbf{x}', \mathbf{x})}_{\substack{\text{align large}\\ \text{patch distr.}}} + \gamma \underbrace{\rho_{\boldsymbol{\theta}_\text{AE}}(\mathbf{x}', \mathbf{x}, \mathbf{y}')}_{\substack{\text{manipulation}\\ \text{control}}} + \kappa \underbrace{\rho(\mathbf{x}')}_{\substack{\text{classic}\\ \text{image}\\ \text{reg. [23]}}} + \nu \underbrace{\rho_{\boldsymbol{\theta}}(\mathbf{x}', \mathbf{x})}_{\substack{\text{feat. map.}\\ \text{distr. [42]}}}$$

For image synthesis using our proposed model, MAGIC, we feed the training image along with an image initialized with noise, denoted by x′, to the model and then using the loss function start to modify x′ to obtain the result.

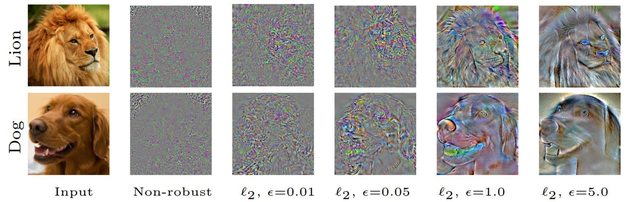
MAGIC Objective Function

The first term in the objective function is the quasi-robust model inversion loss. Model inversion is synthesizing an image through minimizes the loss of a frozen network with respect to some class label. In MAGIC, the target label of model inversion is the label of the training image. Intuitively, model inversion provides the knowledge regarding the image semantics to synthesize more realistic results. We show that inverting a quasi-robust model leads to higher quality results compared with inverting a non-robust or a strongly-robust model. Furthermore, a strongly-robust model has a poor classification accuracy while MAGIC relies on the predicted label of the training image for model inversion.

The second term is the PatchGAN loss which ensure patch consistency between the training and synthesized images. We show that the receptive field of PatchGAN should not very small, as it is in the baseline, for synthesizing more realistic results.

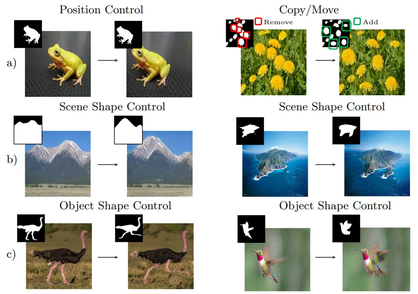
Finally, the patch-based auto-encoder is key in controlling the manipulation and enforcing desired deformations.

Quasi-robust Model Inversion

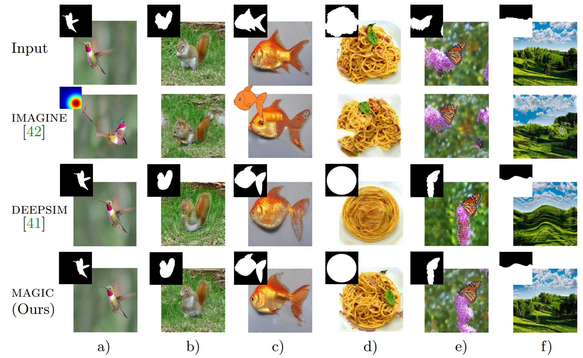


Input gradients are noisy for the non-robust model but for the $\ell 2$ robust models, they start to be aligned with edges as soon as ϵ slightly increases from zero, i.e., the model starts to pay more attention to edges in the input image. For larger ϵ, the model becomes more robust yet gradients are more aligned with course edges so image synthesis using strongly robust models is prone to neglect fine edges and details of the object.

Performance Evaluation



Qualitative results of MAGIC for different tasks.



Qualitative comparison with SOTA.

Summary and Conclusion

MAGIC is an effective method for one-shot mask-guided images synthesis that can find ample applications in advanced image manipulation programs and perform a diverse set of image synthesis tasks using a single training image, its binary segmentation mask, and a guide mask. To the best of our knowledge, this is the first work that demonstrates the advantage of a quasi-robust model inversion for image synthesis.

rouhseda@usc.edu

github.com/mozhdehrouhsedaghat/magic