Large Scale Image Completion via Co-Modulated Generative Adversarial Networks

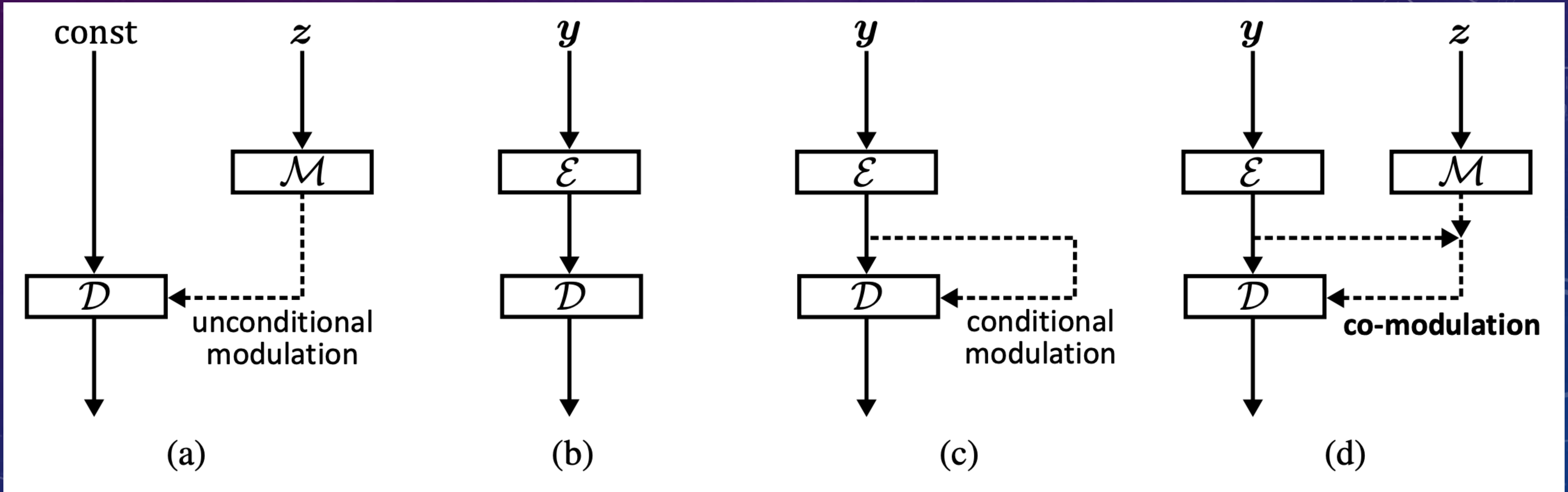
Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I-Chao Chang, and Yan Xu

Unconditional vs. Image-Conditioned generation

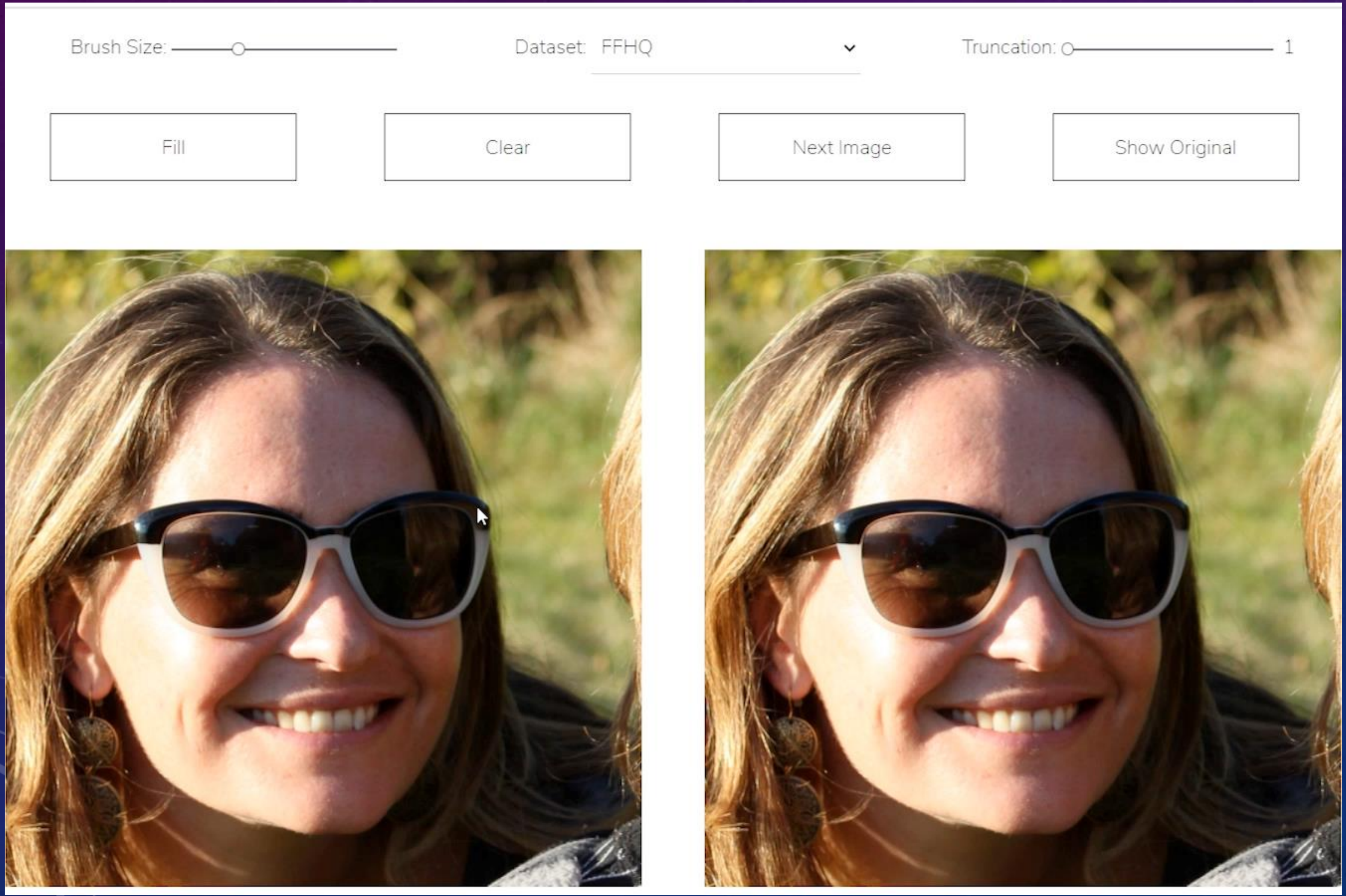
- Image completion aims at ***completing*** images of which certain parts have been deleted/corrupted.
- Previous methods have utilized U-Net-like architectures and the robustness of GANs to attempt the task. However, the lack — or even the absolute absence — of stochasticity led to inadequate capacity to generate ***semantically plausible*** results.

Yu, Jiahui, et al. "Free-form image inpainting with gated convolution." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

Methodology



Qualitative Results: CoModGAN

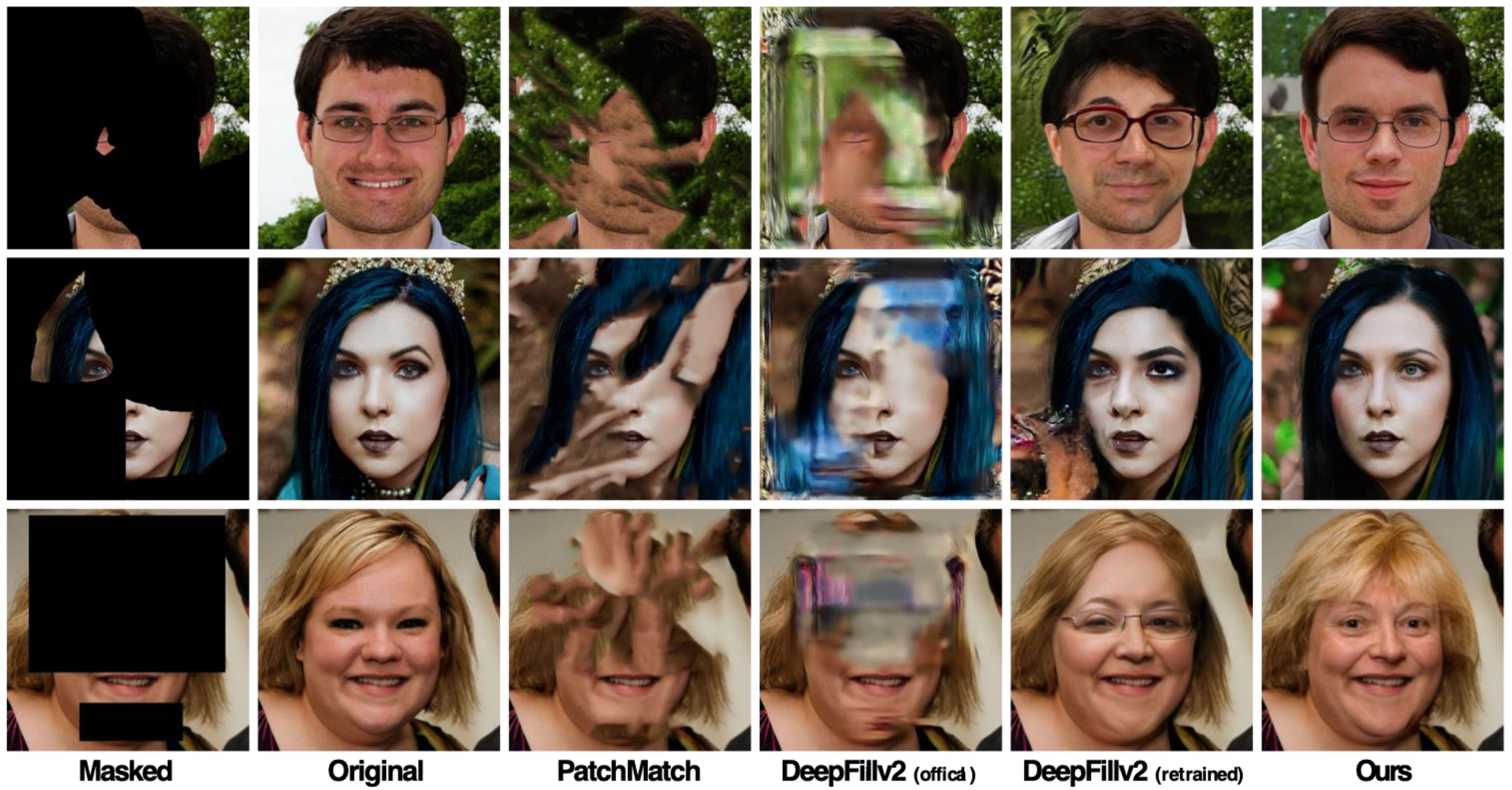


Try out our interactive demo at http://comodgan.ml/!

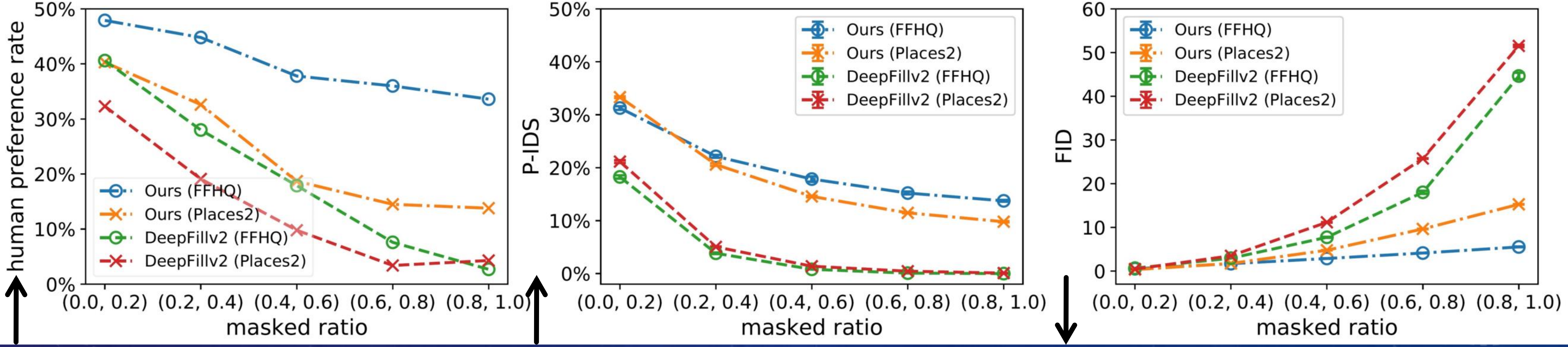
(Currently supports desktops only)

Qualitative Results: CoModGAN

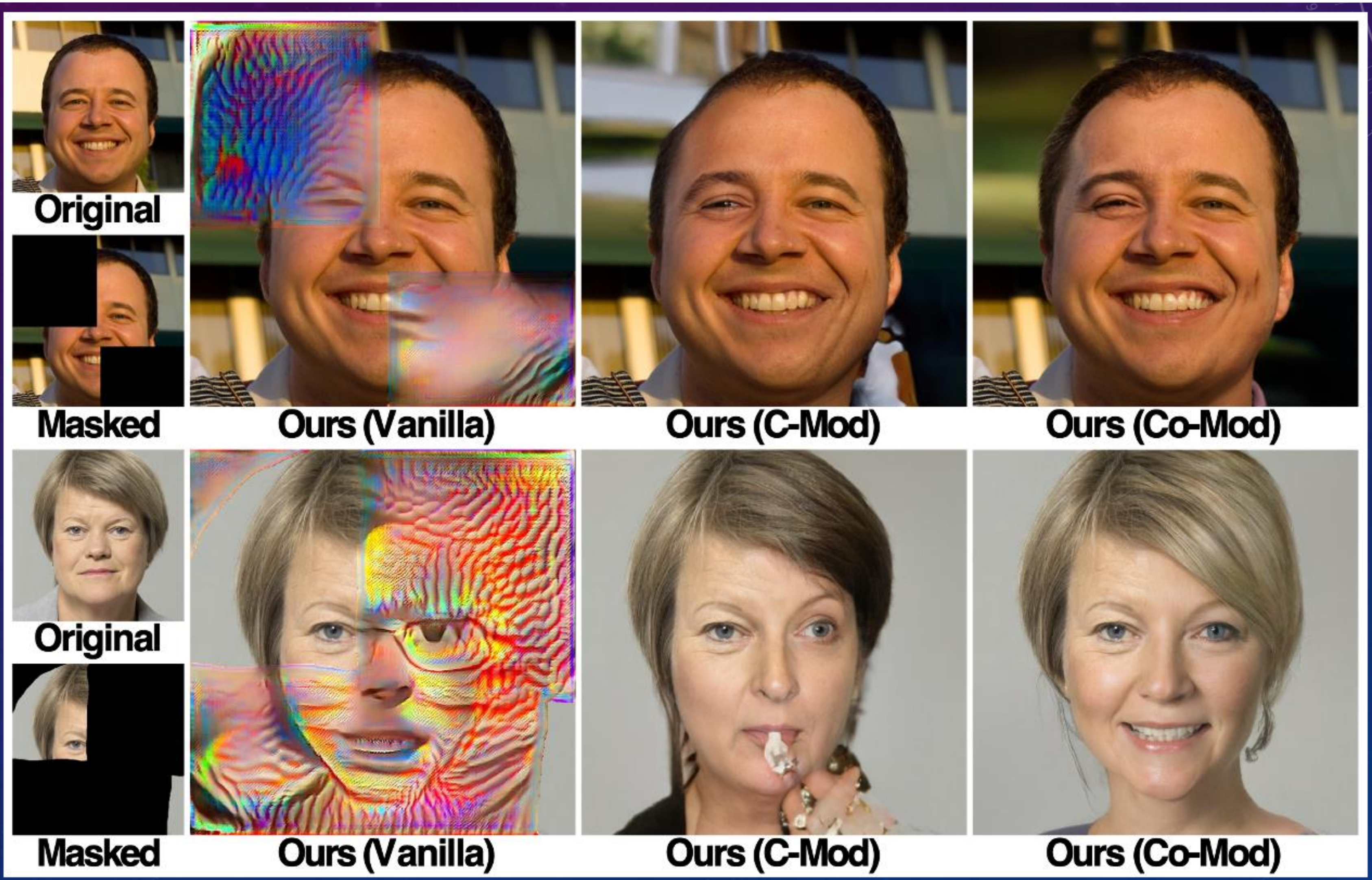
Qualitative Results: CoModGAN–Others



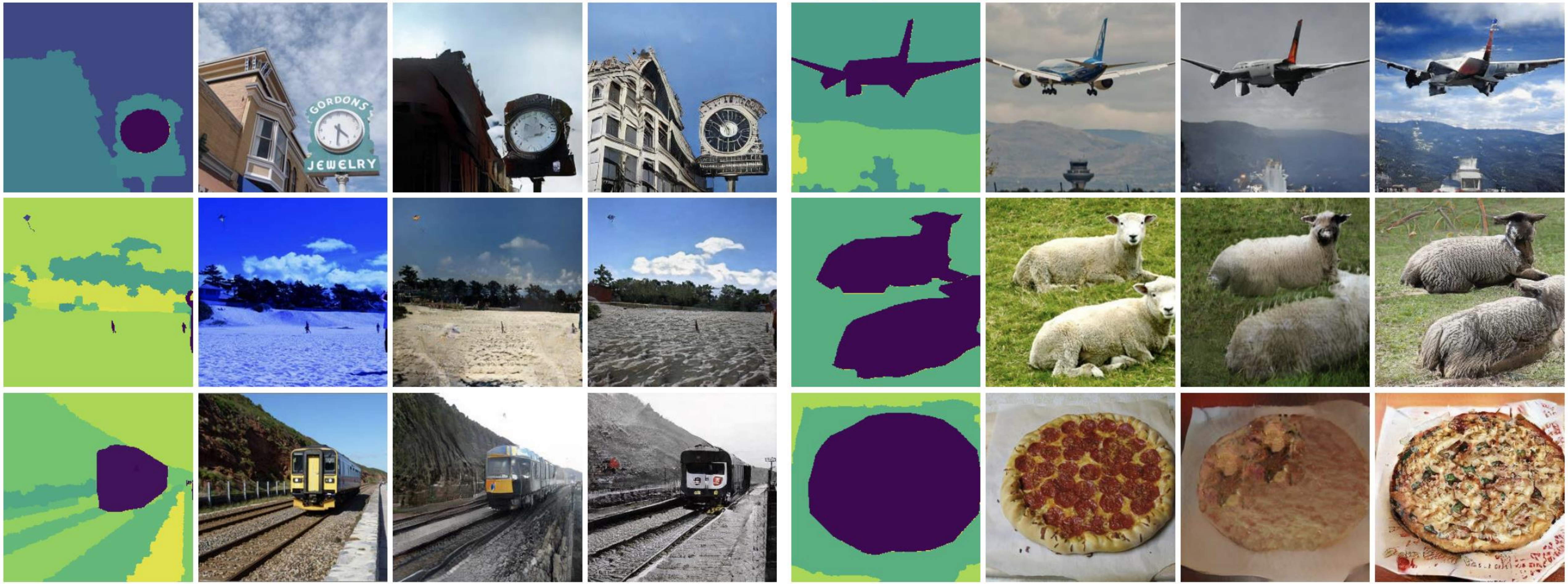
Quantitative Results: CoModGAN–State of the Art



QUALITATIVE RESULTS: ABLATION STUDIES



Generalization to Image-to-Image Translation



Input Real SPADE Ours Input Real SPADE Ours

Metrics for Image Completion: Are They Reasonable?

Unpaired metrics (e.g. FID):

- Unpaired metrics are not ***strict*** enough as they solely measures the difference in the distribution and completely omits the paired nature of the task.

Pixelwise discrepancies (e.g. l_1/l_2, PSNR, SSIM):

- While pixelwise discrepancies intuitively measures the distance between the pair of true–fake images in the pixel space, it inherently favors blurry images and cannot measure the semantic difference.

Metrics for Image Completion: P-IDS & U-IDS

Inspired by the common practice of adopting **user study** as the *gold standard*, we propose Paired Inception Discriminator Score (P-IDS) and Unpaired Inception Discriminator Score (U-IDS), which measure the linear separability of the true–fake image pair in the Inception v3 feature space. Specifically,

- P-IDS reflects the probability that the trained binary SVM classifier deems the fake image more realistic than the true image in a pair:

$$\Pr_{(\boldsymbol{x},\boldsymbol{x}')\in\boldsymbol{X}} \left\{ f(\mathcal{I}(\boldsymbol{x})) > f(\mathcal{I}(\boldsymbol{x}')) \right\}.$$

- U-IDS gives the misclassification rate of the trained bilinear SVM on the features:

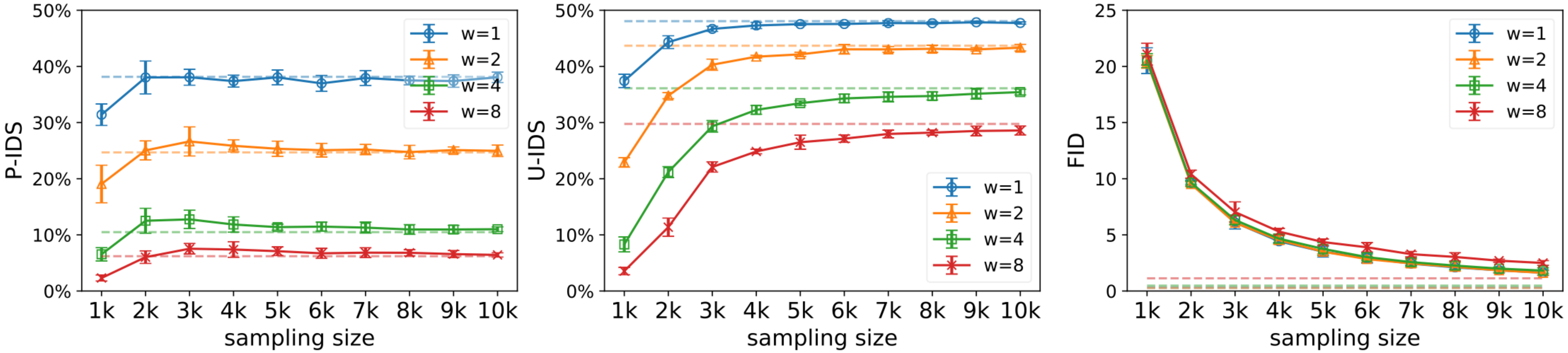
$$\frac{1}{2}\Pr_{\boldsymbol{x}\in X}\{f(\mathcal{I}(\boldsymbol{x}))<0\}+\frac{1}{2}\Pr_{\boldsymbol{x}'\in X'}\{f(\mathcal{I}(\boldsymbol{x}'))>0\}.$$

Metrics for Image Completion: P-IDS & U-IDS

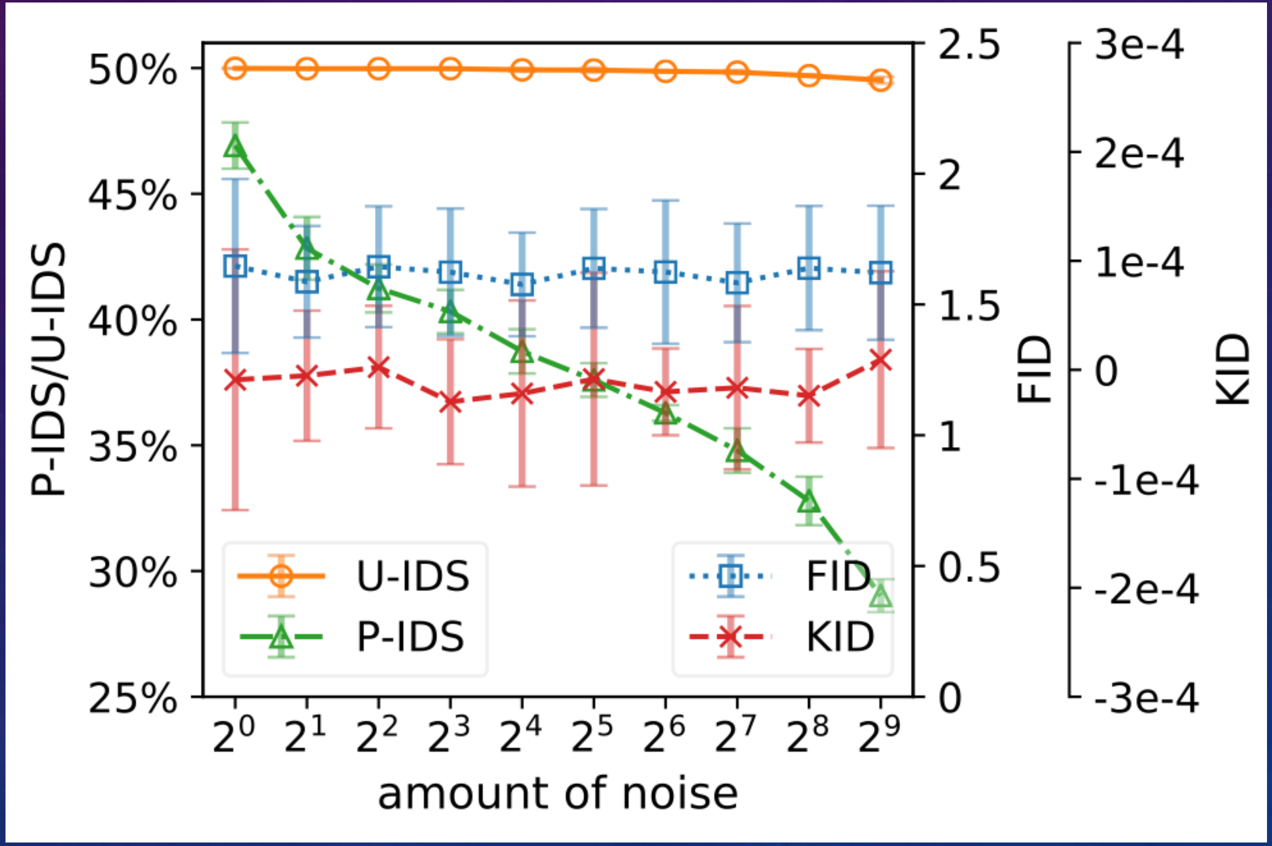
The new metrics are:

- Robust to sampling size,
- Effective of capturing subtle differences, and
- Correlated to human preferences.

Metrics for Image Completion: Sampling Size



Metrics for Image Completion: Subtle Differences



Metrics for Image Completion: Correlation w/ Humans

We evaluate Pearson's correlation coefficient between the metrics and user study results. P-IDS gives $r = 0.870$, which is more preferable in comparison with $r = -0.765$ of FID.

SUMMARY

- We propose CoModGAN, a generic approach the embeds both image-conditional information and stochastic style representations.
- We propose P-IDS & U-IDS for robustly assessing the perceptual fidelity of GANs.
- Experiments demonstrate superior performance in terms of both quality and diversity in free-form image completion and easy generalization to image-to-image translation.

Thank You for Listening!