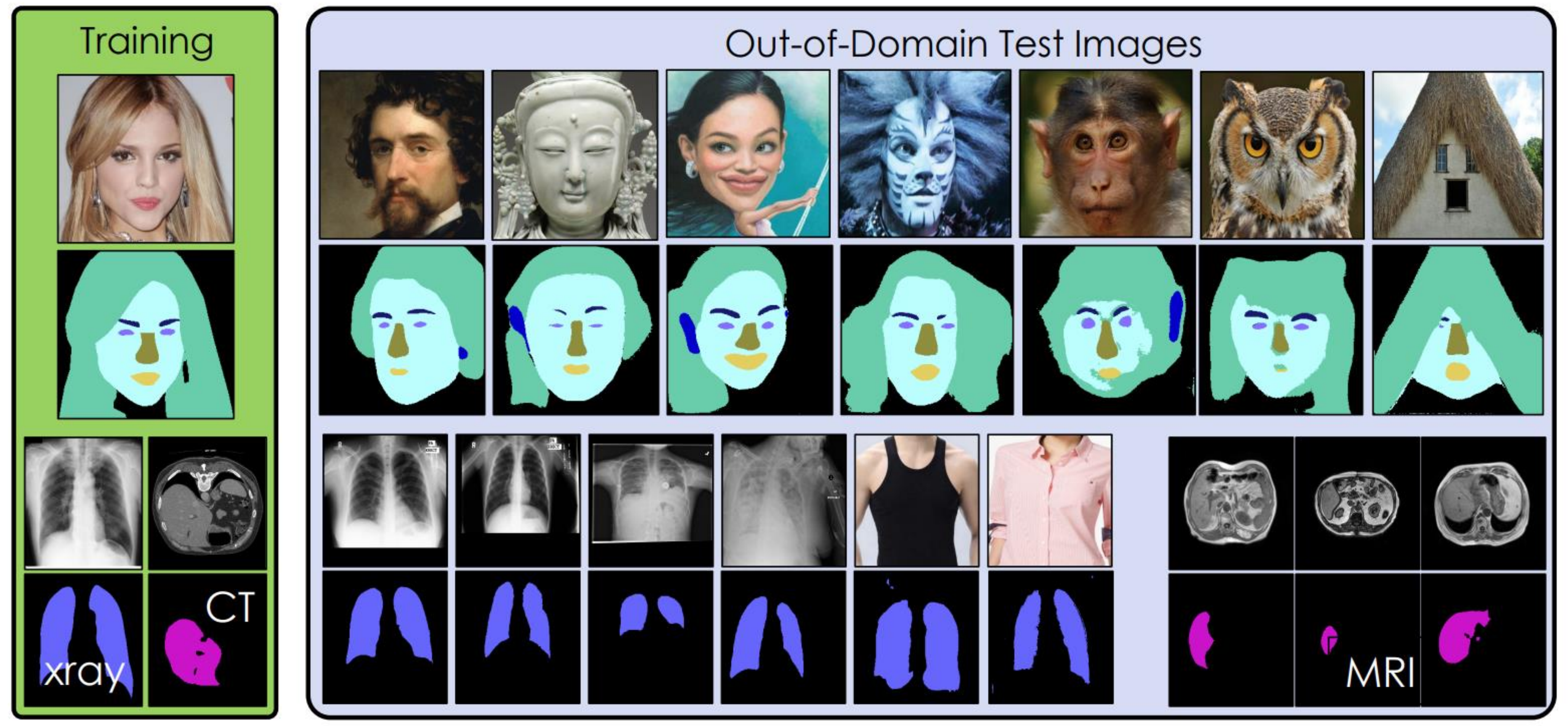
Semantic Segmentation with Generative Models:
Semi-Supervised Learning and Strong Out-of-Domain Generalization

Daiqing Li, Junlin Yang, Karsten Kreis, Antonio Torralba, Sanja Fidler

CVPR VIRTUAL JUNE 19-25

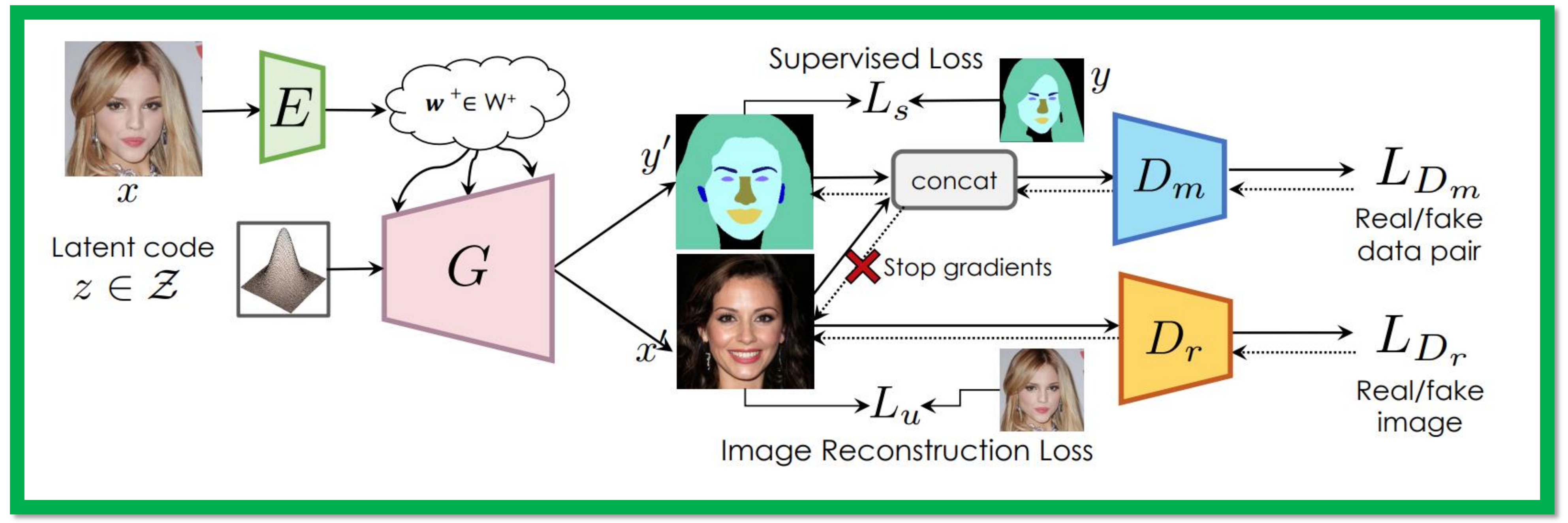
Motivation:



- Pixel-wise labels are expansive to get, some domains require expert annotators (e.g. in medical domain)
- Most methods do not generalize well to unseen domains

Contribution:

- We propose to use the joint image-label generator for pixel-wise inference tasks. We show this to achieve great out of domain generalization
- We demonstrate great label-efficiency by sharing image GAN's feature with the segmentation label branch
- We propose using test-time optimization technique to do pixel-wise task at inference time and show strong generalization capabilities on out-of-domain segmentation tasks



SemanticGAN:

- Augment Stylegan with an additional pixel-wise segmentation branch
- Use our SemanticGAN for pixel-wise segmentation tasks at test time, embed input images into the GAN's latent space

Training SemanticGAN:

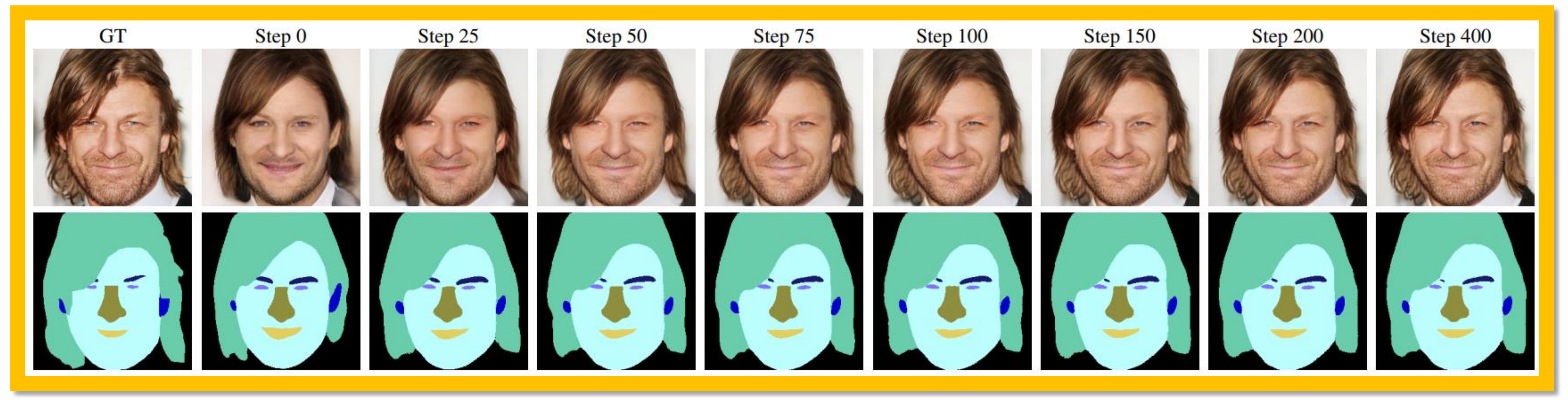
- Jointly train image and segmentation synthesis branches using only adversarial losses. D_r takes images as input to penalize unrealism, D_m takes both images and pixel-wise labels as input to enforce consistency

$$\mathcal{L}_G = \mathop{\mathbb{E}}_{(x_f,\cdot)=G(z), z\sim p(z)} [\log(1 - D_r(x_f))] + \mathop{\mathbb{E}}_{(x_f,y_f)=G(z), z\sim p(z)} [\log(1 - D_m(x_f, y_f))]$$

Inference:

- At inference time, test-time optimization is performed iteratively to find the optimal latent code for the input image
- We train an Encoder to get an initialization of the latent code

$$w^{+*} = \arg\min_{w^+ \in \mathcal{W}^+} [\mathcal{L}_{\text{reconst}}(x^*, G_x(w^+)) + \lambda_2 ||w^+ - E(G(w^+))||_2^2]$$



Experimental Results:

For our-of-domain experiments, we need **50x** fewer labels than baselines

Method	Trained with **40** labeled data samples: ISIC	PH2	IS	Quest	Trained with **200** labeled data samples: ISIC	PH2	IS	Quest	Trained with **2000** labeled data samples: ISIC	PH2	IS	Quest
U-Net	0.4935	0.4973	0.3321	0.0921	0.6041	0.7082	0.4922	0.1916	0.6469	0.6761	0.5497	0.3278
DeepLab	0.5846	0.6794	0.5136	0.1816	0.6962	0.7617	0.6565	0.4664	0.7845	0.8080	0.7222	0.6457
MT	0.5200	0.5813	0.4283	0.1307	0.7052	0.7922	0.6330	0.4149	0.7741	0.8156	0.6611	0.5816
AdvSSL	0.5016	0.5275	0.5575	0.1741	0.6657	0.7492	0.6087	0.3281	0.7388	0.7351	0.6821	0.6178
GCT	0.4759	0.4781	0.5436	0.1611	0.6814	0.7536	0.6586	0.3109	0.7887	0.8248	0.7104	0.5681
Ours-NO	0.6987	0.7565	0.7083	0.5060	0.7517	**0.8160**	0.7150	0.6493	0.7855	0.8087	0.6876	0.6350
Ours	**0.7144**	**0.7950**	**0.7350**	**0.5658**	**0.7555**	0.8154	**0.7388**	**0.6958**	**0.7890**	**0.8329**	**0.7436**	**0.6819**

- Skin Lesion Segmentation Task. Numbers are IoU

Labeled vs Unlabeled data

Labeled	Unlabeled: 3K	10K	28K
30	0.6786	0.6845	0.6902
150	0.7046	0.7438	0.7600
1500	0.7566	0.7710	0.7810

(a) CelebA-Mask (In-Domain)

Labeled	Unlabeled: 3K	10K	28K
30	0.5410	0.5799	0.5883
150	0.5871	0.6152	0.6336
1500	0.6011	0.6204	0.6633

(b) MetFaces-40 (Out-Domain)

- Reducing label cost without performance drop by adding more unlabeled data
- Unlabeled data is more valuable than labeled data in out-of-domain (OOD) task