Improving Discriminative Models

Generative vs Discriminative

From C. Bishop:

Approaches that **explicitly or implicitly model the distribution of inputs as well as outputs are known** as generative models, because by sampling from them it is possible to generate synthetic data points in the input space.

Sampling	Style Transfer	Classification	Evaluating Likelihood
$x \sim p(x)$	$x_k \sim p(x\|C_k)$	$C_k \sim p(C_k\|x)$	$log\ p(x)$

Approaches that model the posterior probabilities directly are called discriminative models.

Modeled with NN

not NN

Classification

$$p(C_k|x) = f(x)$$

Improving Discriminative with Generative

- Improving detection
- Improving alignment
- Improving identification
- Improving segmentation

Improving Detection

Bai, Yancheng, et al. "Finding tiny faces in the wild with generative adversarial network." CVPR'2018

Improving Detection

Bai, Yancheng, et al. "Finding tiny faces in the wild with generative adversarial network." CVPR'2018

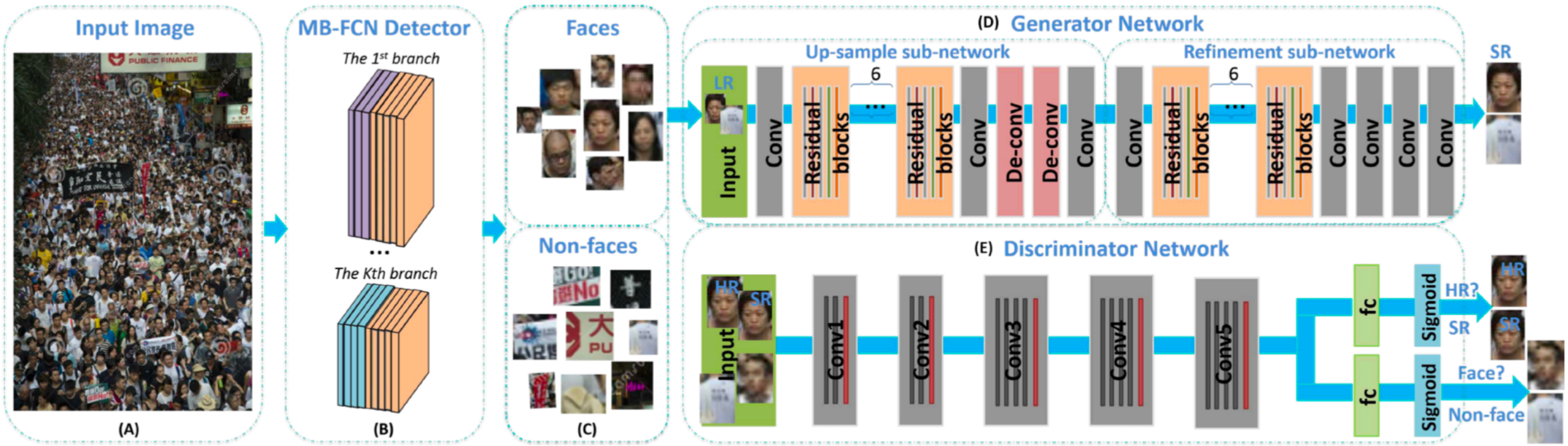
Improving Detection

Super-resolving face images for better detection

Bai, Yancheng, et al. "Finding tiny faces in the wild with generative adversarial network." CVPR'2018

Improving Detection

Super-resolving face images for better detection



- Multi-branch detector is used to predict region proposals
- Generator 4x upsamples images
- Discriminator learn to predict face/non-face and whether an image was super-resolved

Bai, Yancheng, and Bernard Ghanem. "Multi-branch fully convolutional network for face detection."

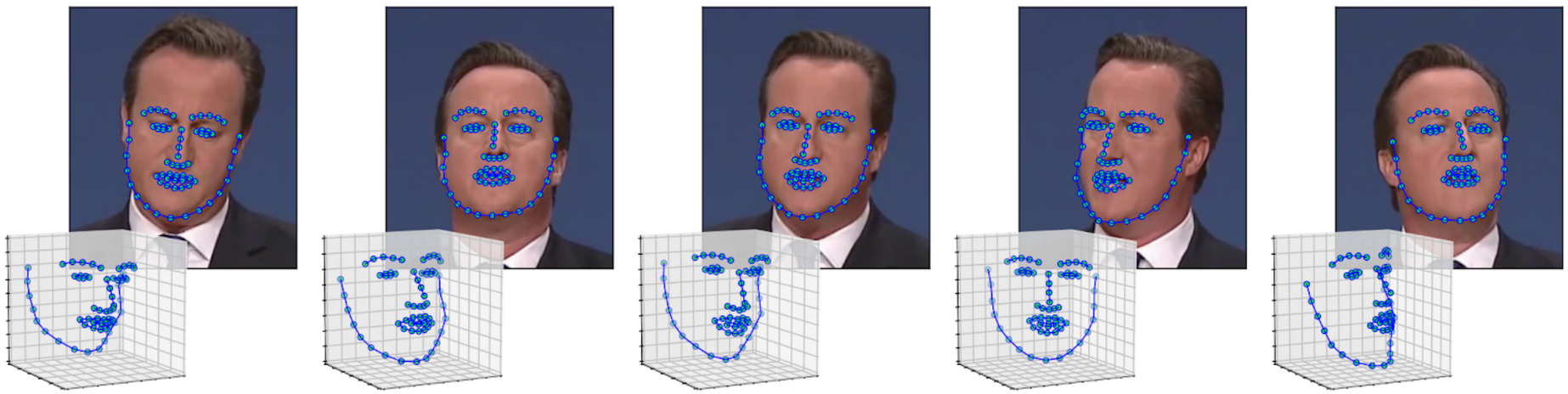
Improving Detection

We improve the quality by a generative network making the job of detector easier

Bai, Yancheng, et al. "Finding tiny faces in the wild with generative adversarial network." CVPR'2018

Improving Alignment

Face Alignment—detecting points on the face corresponding to facial features



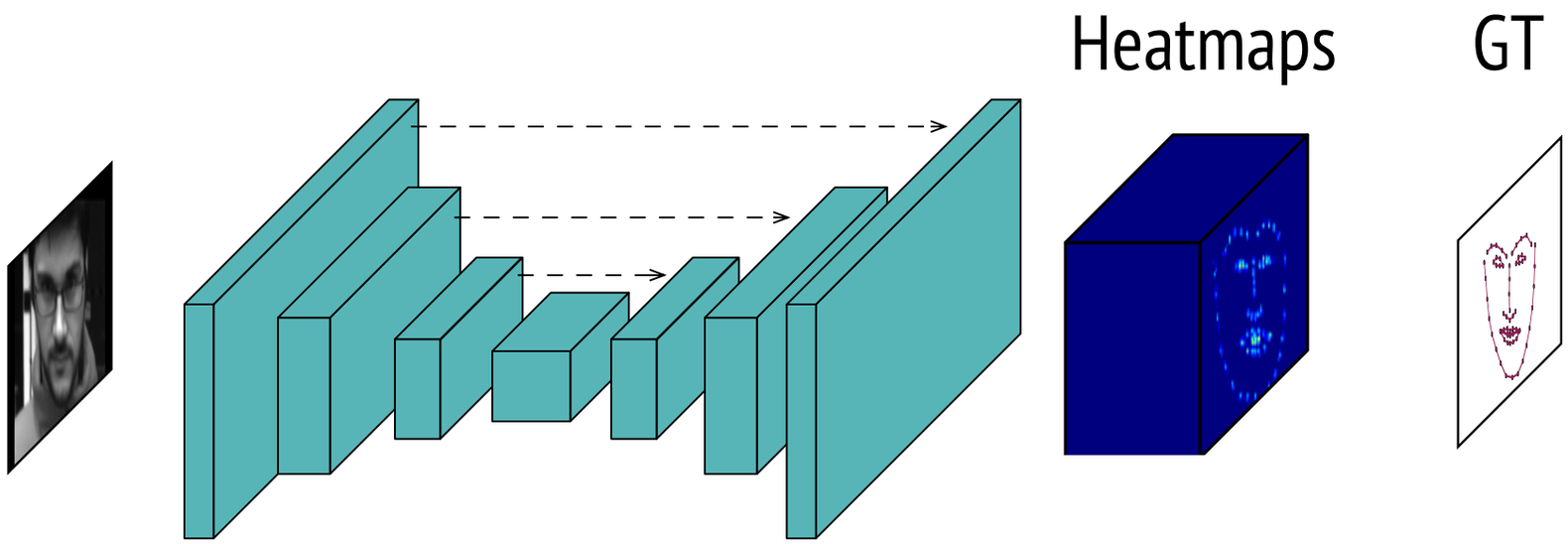
Our goals:

- Improve confidence of alignment algorithms
- Leverage unlabelled data

Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

How to Backprop through argmax?

You have an encoder-decoder network predicting heatmaps



Derivation of SoftArgMax

$$\text{softargmax}(\beta \mathbf{h}) = \sum_x \text{softmax}(\beta \mathbf{h}_x) \cdot x = \sum_x \frac{e^{\beta \mathbf{h}_x}}{\sum_j e^{\beta \mathbf{h}_j}} \cdot x = \sum_x p(x) \cdot x = \mathbb{E}_{\mathbf{h}}[x]$$

Honari, Sina, et al. "Improving landmark localization with semi-supervised learning." CVPR'2018.

How to Backprop through argmax?

You have an encoder-decoder network predicting heatmaps

$$\text{softargmax}(\beta \mathbf{h}) = \sum_x \text{softmax}(\beta \mathbf{h}_x) \cdot x = \sum_x \frac{e^{\beta \mathbf{h}_x}}{\sum_j e^{\beta \mathbf{h}_j}} \cdot x = \sum_x p(x) \cdot x = \mathbb{E}_{\mathbf{h}}[x]$$

We predict the mean only. What about variance?

Gaussian:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\sigma^2 = \mathbb{E}_{\mathbf{h}}[(\mathbf{x} - \mathbb{E}_{\mathbf{h}}[\mathbf{x}])^2]$$

Laplace:

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

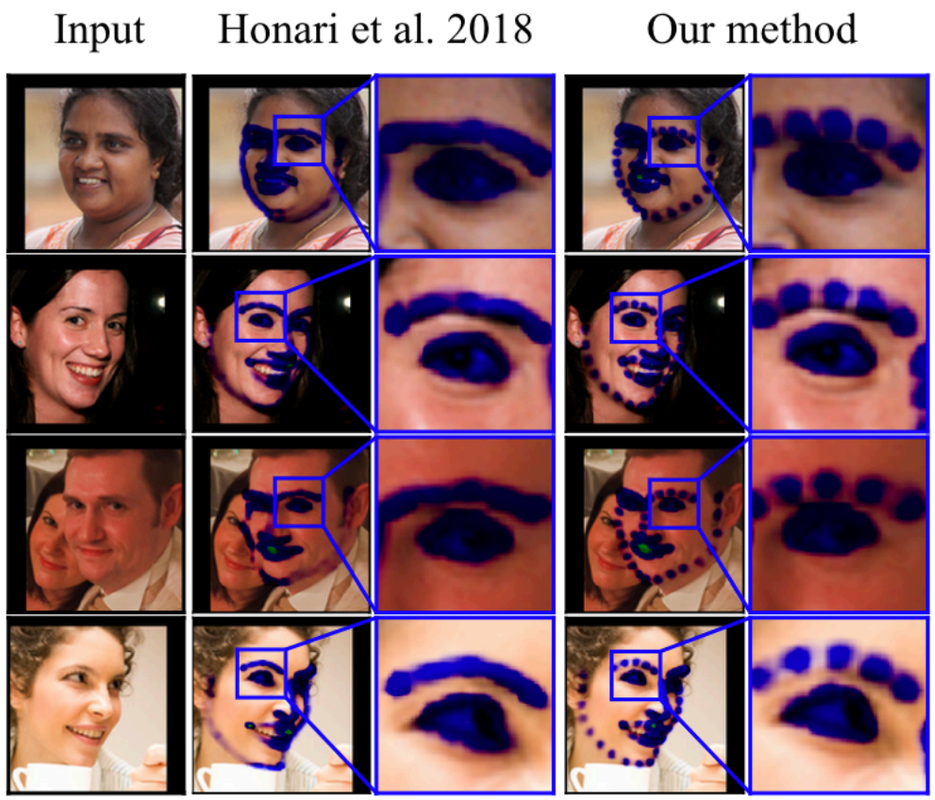
$$b = \mathbb{E}_{\mathbf{h}}[|\mathbf{x} - \mathbb{E}_{\mathbf{h}}[\mathbf{x}]|]$$

Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Considering Variance

Loss function

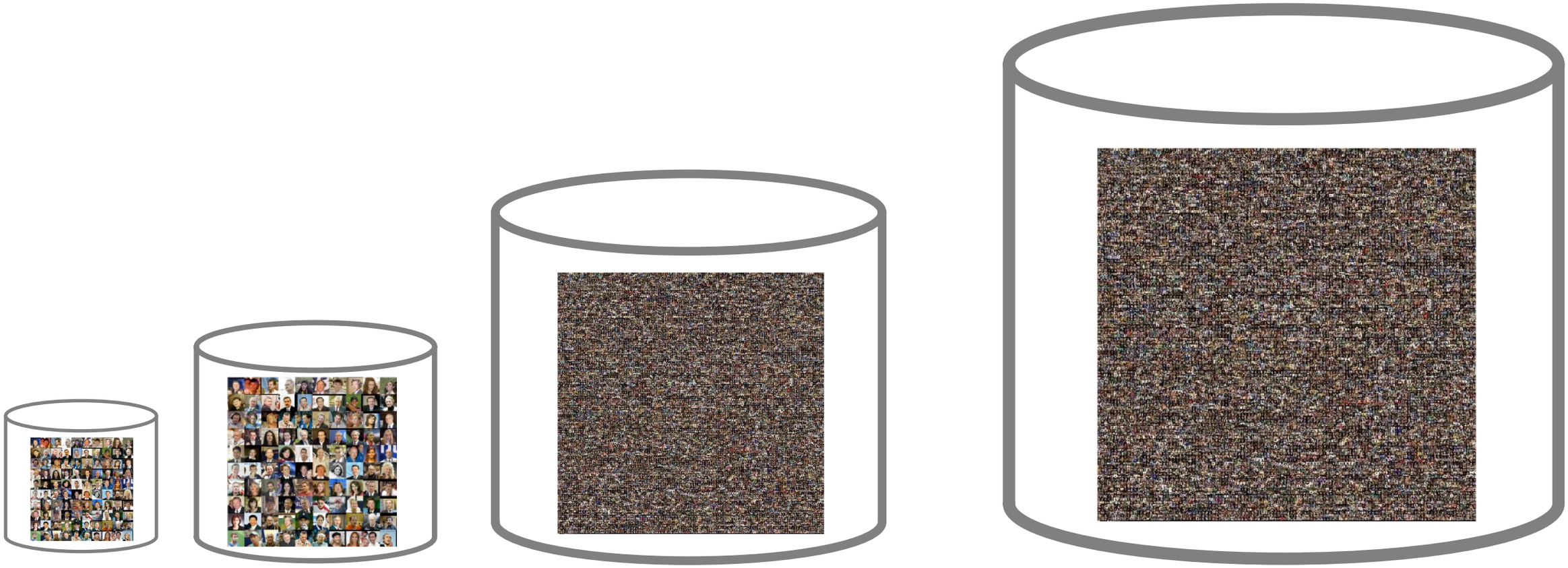
$$\mathcal{L}_{\mathrm{KL}} = \mathbb{E}_{(\mathbf{d},\mathbf{s})\sim p(\mathbf{d},\mathbf{s})}\Big[\mathrm{D}_{KL}(q(\mathbf{s}|\mathbf{d})||p(\mathbf{s}|\mathbf{d}))\Big]$$



Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

What about Unlabeled Data?

Labeling data is expensive. It takes around a minute to label a face image



Labelled videos

Labelled images

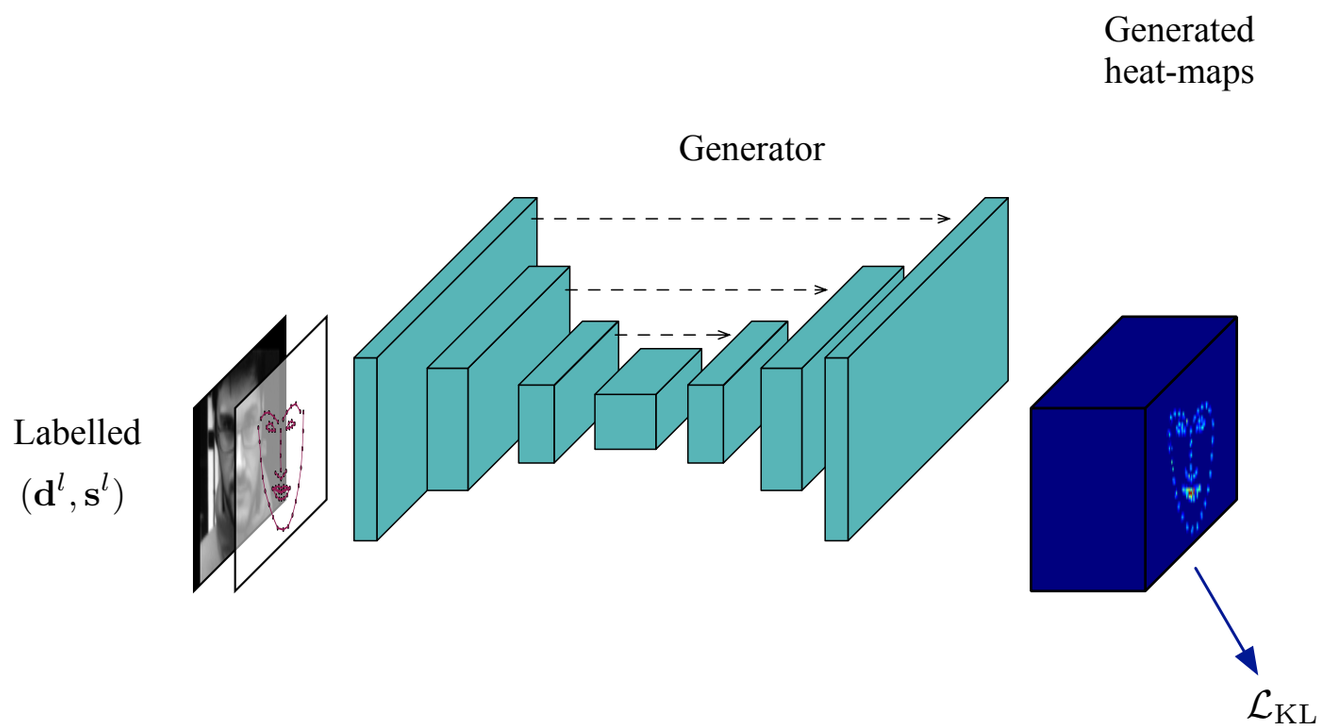
Unlabeled images

Unlabeled videos

Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Using Adversarial Framework for Alignment

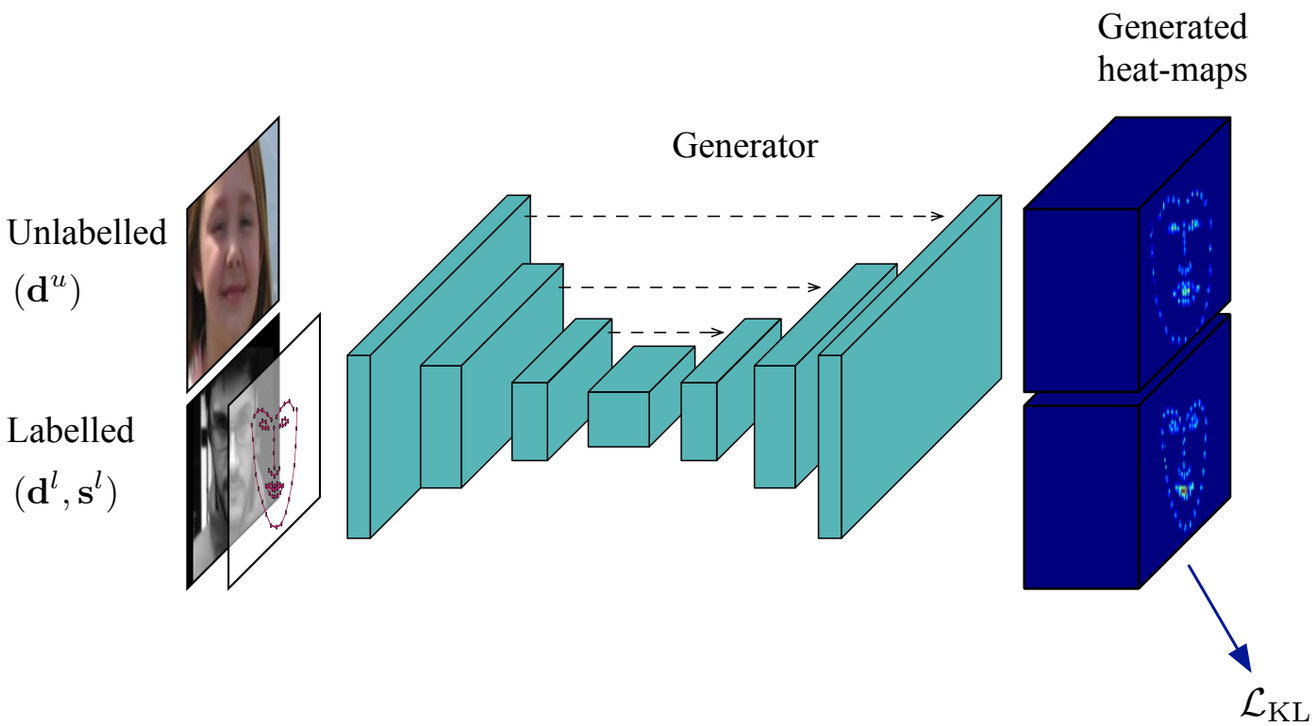
Labeling data is expensive. It takes around a minute to label a face image



Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Using Adversarial Framework for Alignment

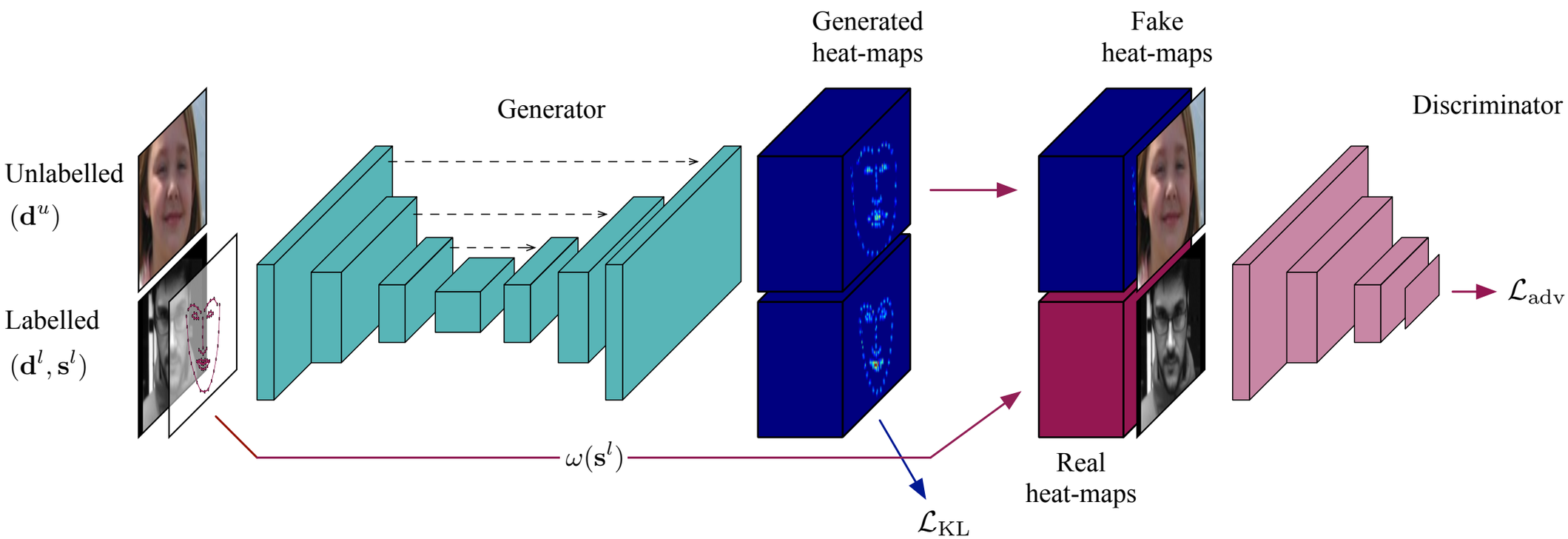
Labeling data is expensive. It takes around a minute to label a face image



Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

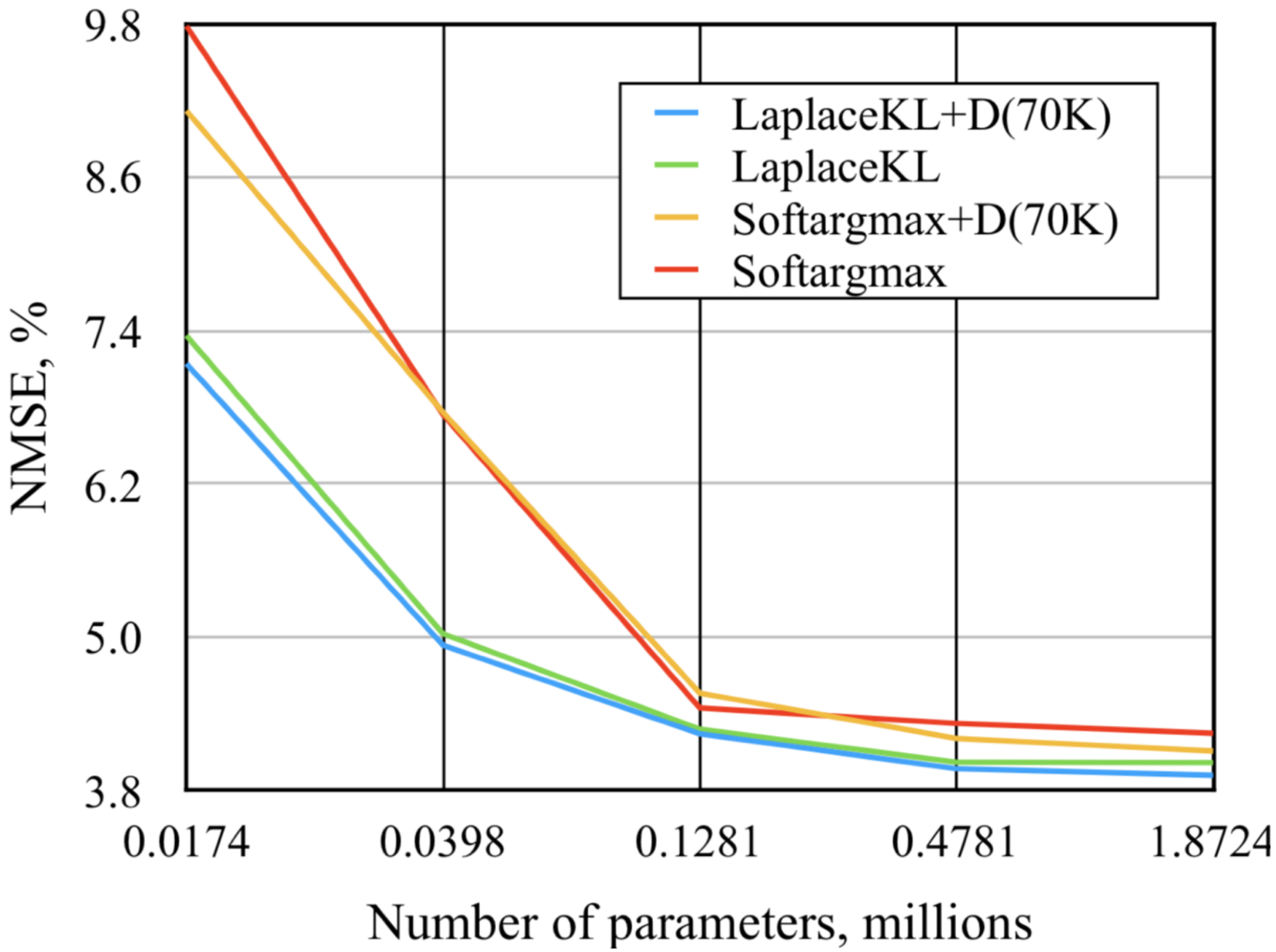
Using Adversarial Framework for Alignment

Labeling data is expensive. It takes around a minute to label a face image



Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Results



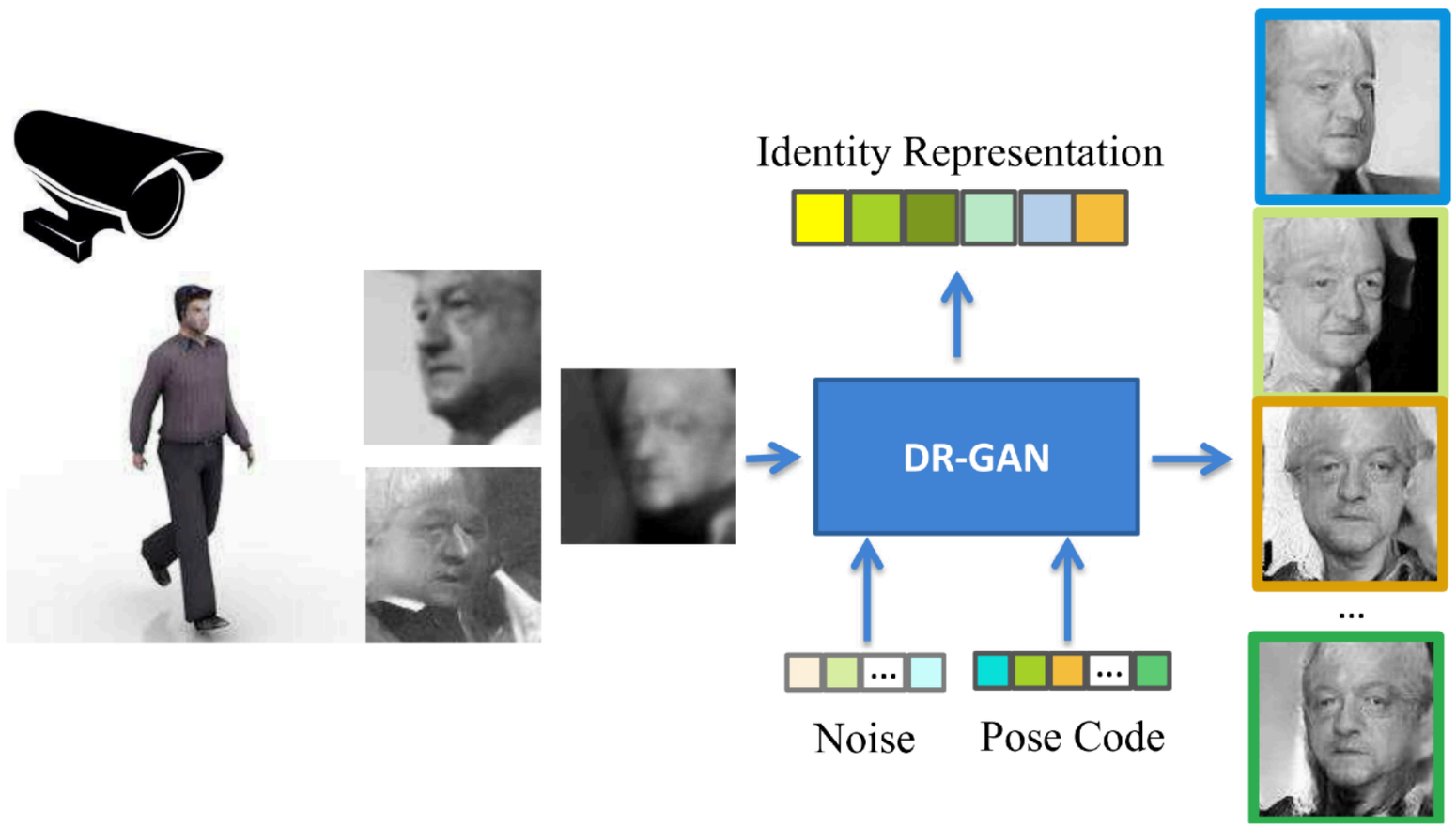
Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Results

Robinson, Joseph P., et al. "Laplace landmark localization." ICCV'2019

Improving Identification

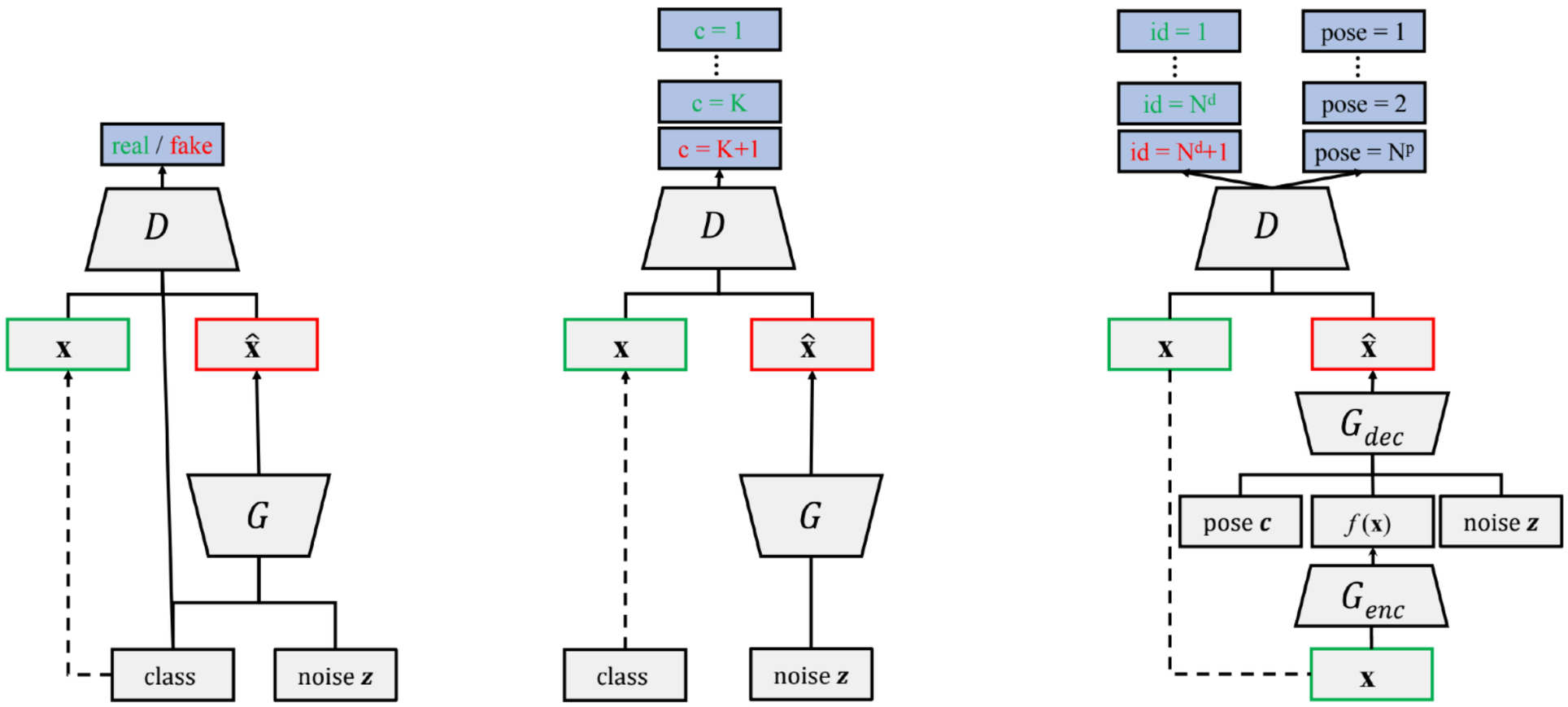
DR-GAN uses several images of an input person to generate different head poses and identify



Tran, Luan, Xi Yin, and Xiaoming Liu. "Disentangled representation learning gan for pose-invariant face recognition." CVPR'2017

DR-GAN

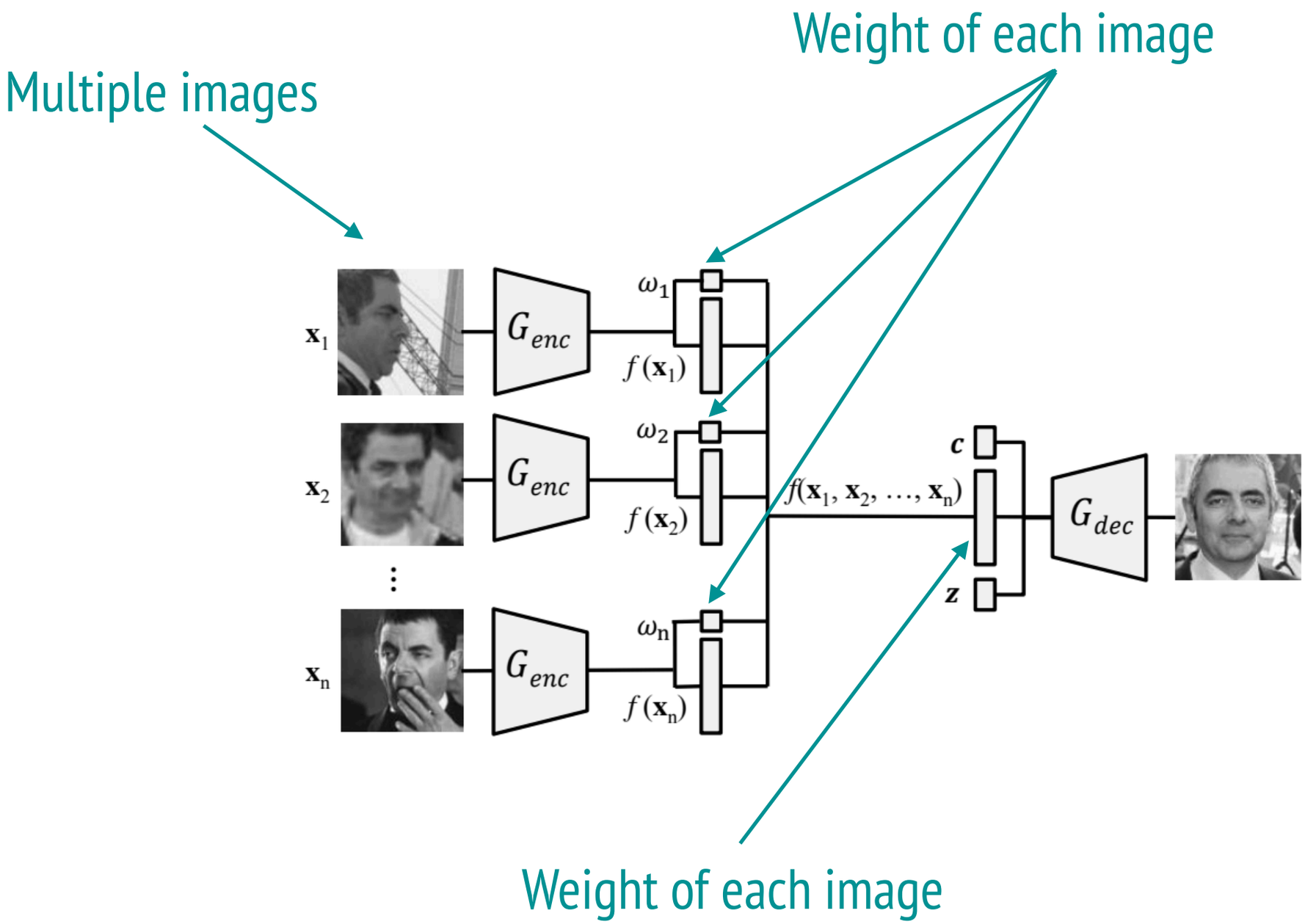
Discriminator predicts identity, head pose and true/fake



(a) Conditional GAN (b) Semi-Supervised GAN (d) DR-GAN

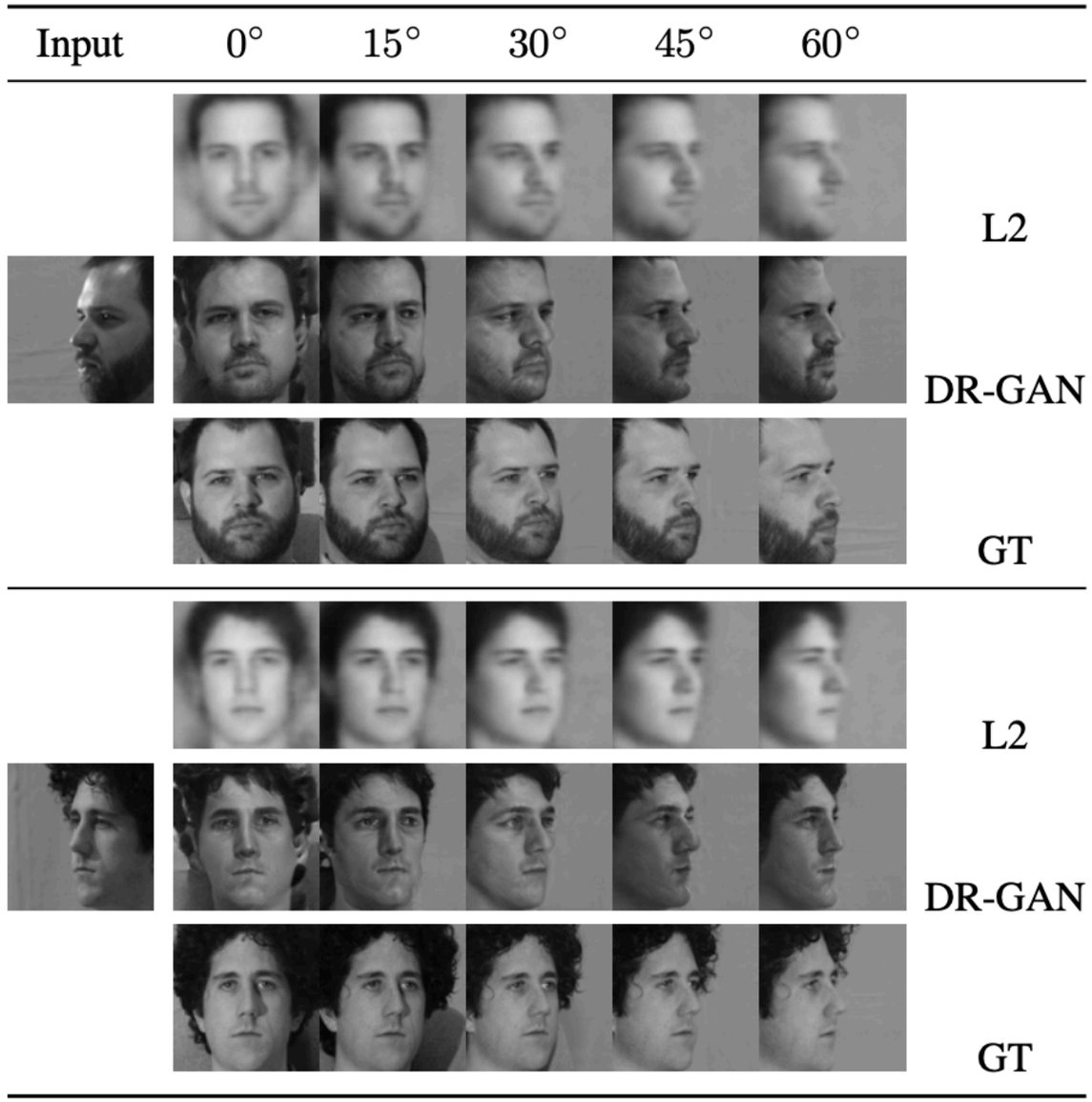
Tran, Luan, Xi Yin, and Xiaoming Liu. "Disentangled representation learning gan for pose-invariant face recognition." CVPR'2017

Extension to Multi Image Input



Tran, Luan, Xi Yin, and Xiaoming Liu. "Disentangled representation learning gan for pose-invariant face recognition." CVPR'2017

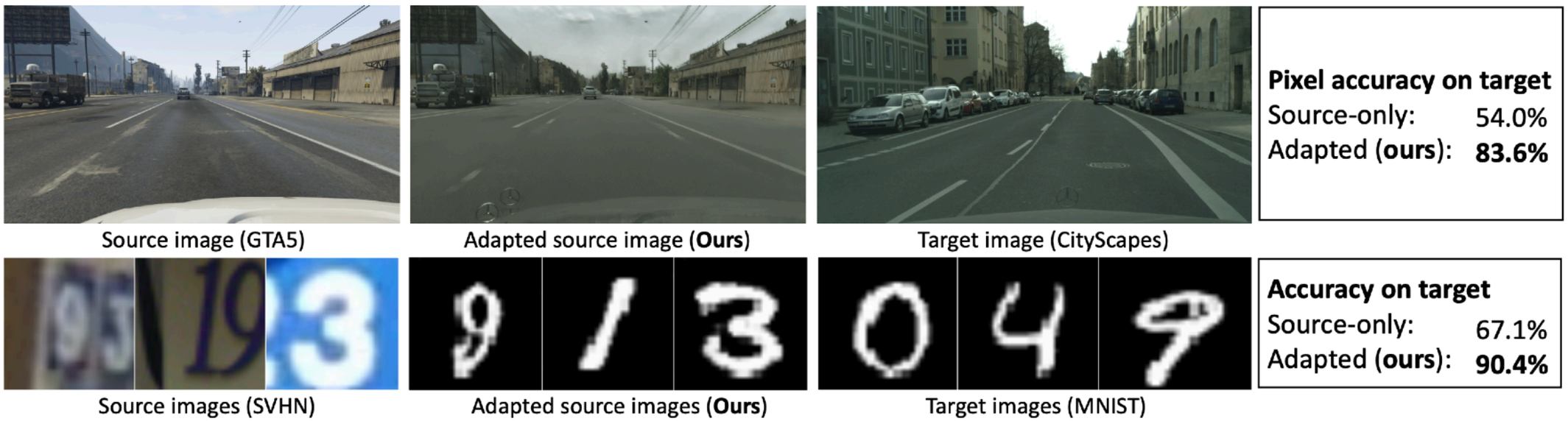
Single Image Generation Results



Tran, Luan, Xi Yin, and Xiaoming Liu. "Disentangled representation learning gan for pose-invariant face recognition." CVPR'2017

Improving Segmentation

Getting data is expensive. What if we can use synthetic data



Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." ICML'2017

CyCADA

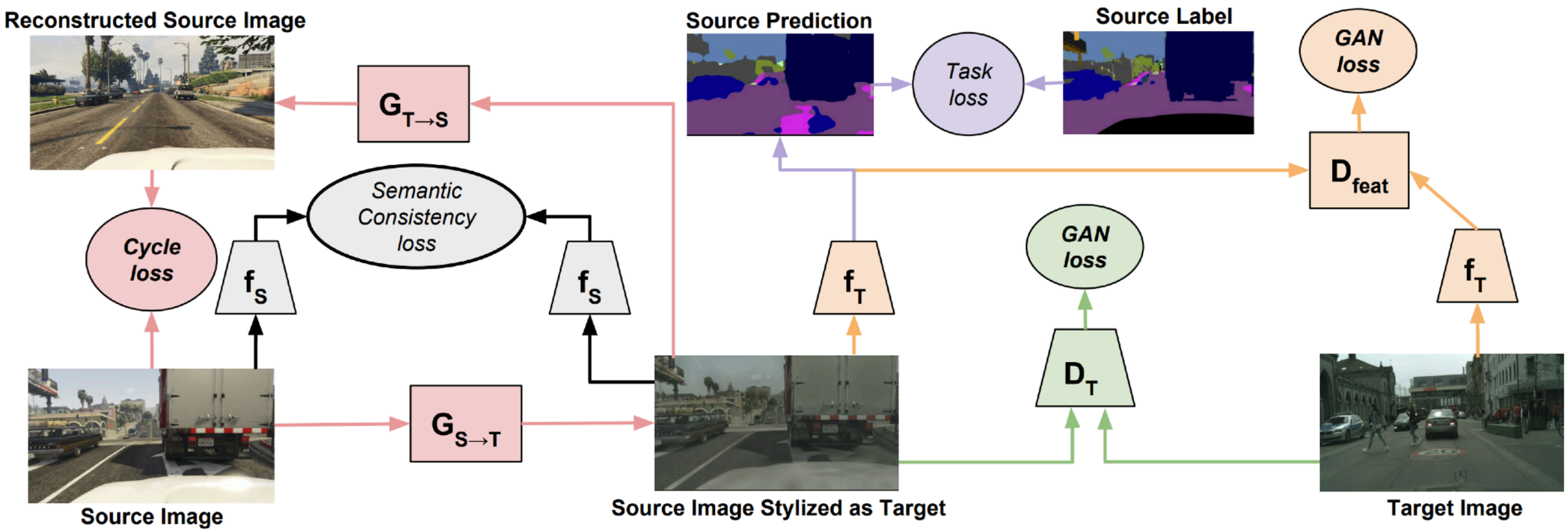


Image-level GAN loss

Feature-level GAN loss

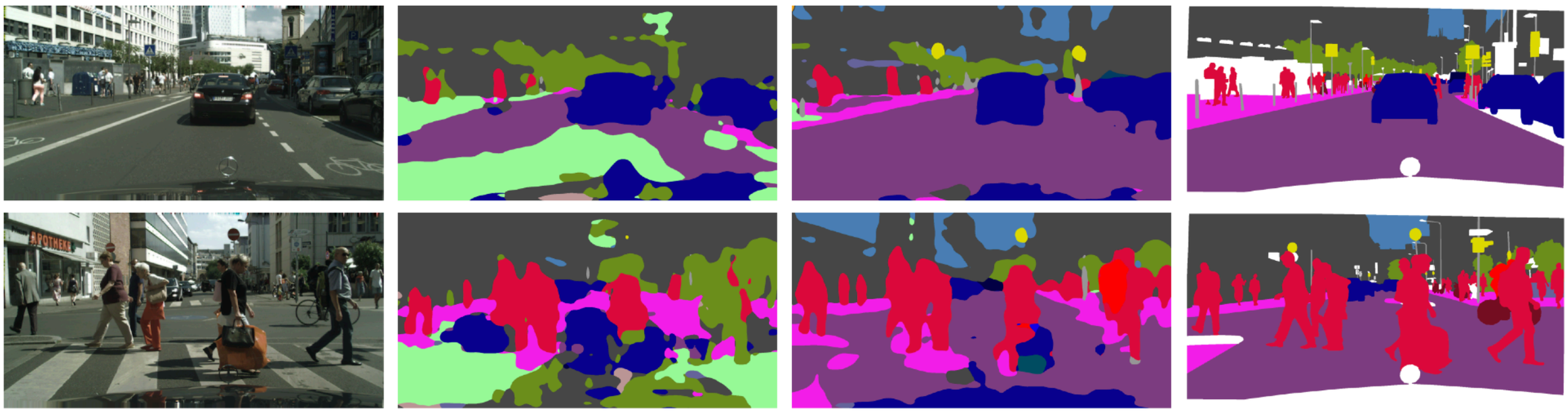
Source cycle loss

Source task loss

Semantic consistency loss

Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." ICML'2017

CyCADA: Results



(a) Test Image (b) Source Prediction (c) CyCADA Prediction (d) Ground Truth

Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." ICML'2017